

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÕES E ARTES
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

DENISE GOMES SILVA MORAIS CAVALCANTE

**Modelagem semântica de sistemas de organização do conhecimento para categorias de
representação de imagens em movimento no ambiente web**

São Paulo
2023

DENISE GOMES SILVA MORAIS CAVALCANTE

Modelagem semântica de sistemas de organização do conhecimento para categorias de representação de imagens em movimento no ambiente web

Versão Corrigida (versão original disponível na Biblioteca da ECA/USP)

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Escola de Comunicações e Artes da Universidade de São Paulo para obtenção do título de Doutorado.

Área de concentração: Cultura e Informação

Orientadora: Dra. Cibele A. C. Marques dos Santos

São Paulo
2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Escola de Comunicações e Artes da Universidade de São Paulo
Dados inseridos pelo(a) autor(a)

Cavalcante, Denise Gomes Silva Morais
Modelagem semântica de sistemas de organização do conhecimento para categorias de representação de imagens em movimento no ambiente web. / Denise Gomes Silva Morais Cavalcante; orientadora, Cibele Araújo Camargo Marques dos Santos. - São Paulo, 2023.
267 p. : il.

Tese (Doutorado) - Programa de Pós-Graduação em Ciência da Informação / Escola de Comunicações e Artes / Universidade de São Paulo.
Bibliografia
Versão corrigida

1. Imagem em movimento. 2. Web semântica. 3. SKOS. 4. OWL. 5. Anotação semântica. I. Araújo Camargo Marques dos Santos, Cibele. II. Título.

CDD 21.ed. - 020

Elaborado por Alessandra Vieira Canholi Maldonado - CRB-8/6194

Nome: CAVALCANTE, Denise Gomes Silva Moraes.

Título: Modelagem semântica de sistemas de organização do conhecimento para categorias de representação de imagens em movimento no ambiente web.

Tese apresentada à Escola de Comunicações e Artes da Universidade de São Paulo para obtenção do título de Doutor em Ciência da Informação.

Aprovado em:

Banca examinadora

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

AGRADECIMENTOS

À Profa.Dra. Cibele, pela confiança, disposição e apoio ao longo do meu crescimento científico.

À Profa.Dra. Vânia e Prof.Dr. Fabiano, agradeço pelo tempo e pela contribuição no exame de qualificação.

Aos professores e funcionários do PPGCI ECA/USP, pela disponibilidade ao longo desta pesquisa.

À minha família e amigos, pelo apoio incondicional e encorajamento durante toda a jornada.

À CAPES: “O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001”

A todos que fazem parte desta jornada, meu profundo agradecimento pelo apoio e contribuições.

Por fim, meu agradecimento especial aos meus pais, Eloiza e Edinilton, e ao meu companheiro, Marcos.

Pela primeira vez no processo de reprodução da imagem, a mão foi liberada das responsabilidades artísticas mais importantes, que agora cabiam unicamente ao olho. Como o olho apreende mais depressa do que a mão desenha, o processo de reprodução das imagens experimentou tal aceleração que começou a situar-se no mesmo nível que a palavra oral. Se o jornal ilustrado estava contido virtualmente na litografia, o cinema falado estava contido virtualmente na fotografia.

Walter Benjamin

RESUMO

CAVALCANTE, D. G. S. M. **Modelagem semântica de sistemas de organização do conhecimento para categorias de representação de imagens em movimento no ambiente web.** 2023. 260 F. Tese (Doutorado em Ciência da Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2023.

O contexto da pesquisa compreendeu a organização e a representação de imagens em movimento no ambiente da web, visando auxiliar a recuperação da informação e, consequentemente, promover o acesso e a difusão desses materiais. A representação temática e a criação de sistemas de organização do conhecimento em cinematecas e arquivos filmicos têm muitos desafios na realização da representação dos assuntos contidos nos documentos, principalmente quando comparados ao estado avançado dos processos de catalogação e à aplicação de metadados descritivos. A questão de pesquisa abordou as dificuldades de representar o conteúdo desses documentos no ambiente web, ao mesmo tempo, em que novas soluções surgem para os problemas de organização do conhecimento, relacionados à imagem em movimento. De modo que, despontaram iniciativas que aplicam a anotação semântica, a partir de tecnologias da Web Semântica, como o *Resource Description Framework*, o *Resource Description Framework Schema*, o *Simple Knowledge Organization System* e a *Ontology Web Language*, para representar conceitos no interior das imagens. O objetivo abrangeu a pesquisa bibliográfica e exploratória sobre os sistemas de organização do conhecimento e os processos de indexação de imagens em movimento no ambiente web. Bem como a proposta de elaboração de microtesauros conceituais no modelo de dados *Simple Knowledge Organization System* e a modelagem semântica de ontologias das categorias de representação temática da imagem em movimento (QUEM, QUANDO, ONDE e O QUE) usadas no processo de indexação. A metodologia iniciou-se na revisão bibliográfica e exploratória, que embasou o referencial teórico e metodológico da tese. O referencial teórico tratou de pesquisas teóricas e aplicadas, relacionadas aos sistemas de organização do conhecimento no ambiente web e à imagem em movimento. Igualmente, foi estruturado a parte aplicada da modelagem de dados para a criação de microtesauros conceituais no ambiente web. Além disso, definiu os requisitos da modelagem da ontologia no domínio da indexação de imagens em movimento. Os resultados incluíram o desenvolvimento de tesaurus conceituais que podem ser utilizados como “vocabulários de valor” para anotações semânticas. Da mesma forma, foi realizada a modelagem semântica da ontologia denominada “Categorias de Representação da Imagem em Movimento”, composta pelas classes Categorias, De Genérico e De Específico. Dessa forma, conclui-se que os sistemas de organização do conhecimento e a Web Semântica são fortes aliados no desenvolvimento de soluções e inovações para desafios e problemas no campo da organização e representação da imagem em movimento. O *Simple Knowledge Organization System* é um modelo que, ao diferenciar termos e conceitos, se adapta ao ambiente dinâmico da web e seus conceitos podem ser usados como valor para metadados no nível da imagem, ou seja, anotações em regiões específicas. A ontologia conceituou o processo de indexação de imagens, permitindo a criação de classes para as categorias e os níveis de indexação, bem como a criação de propriedades de objeto para designar a estrutura hierárquica dos conceitos genéricos e específicos.

Palavras-chaves: imagem em movimento; web semântica; SKOS; OWL; anotação semântica.

ABSTRACT

CAVALCANTE, D. G. S. M. **Modelagem semântica de sistemas de organização do conhecimento para categorias de representação de imagens em movimento no ambiente web.** 2023. 260 F. Tese (Doutorado em Ciência da Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2023.

The context of the research included the organization and representation of moving images in the web environment, with the aim of aiding information retrieval and, consequently, promoting access to and dissemination of these materials. Thematic representation and the creation of knowledge organization systems in film libraries and film archives face many challenges when it comes to representing the subjects contained in documents, especially when compared to the advanced state of cataloguing processes and the application of descriptive metadata. The research question addressed the difficulties of representing the content of these documents in the web environment, at the same time as new solutions emerge for the problems of organizing knowledge related to the moving image. As a result, initiatives have emerged that apply semantic annotation, based on semantic web technologies such as the Resource Description Framework, the Resource Description Framework Schema, the Simple Knowledge Organization System and the Ontology Web Language, to represent concepts within images. The objective included bibliographic and exploratory research into knowledge organization systems and moving image indexing processes in the web environment, as well as a proposal to create conceptual micro treasures using the Simple Knowledge Organization System data model and the semantic modeling of ontologies for moving image thematic representation categories (WHO, WHEN, WHERE and WHAT) used in the indexing process. The methodology began with a bibliographical and exploratory review, which provided the theoretical and methodological framework for the thesis. The theoretical framework dealt with theoretical and applied research related to knowledge organization systems in the web environment and the moving image. It also structured the applied part of data modeling for the creation of conceptual micro thesauruses in the web environment. It also defined the requirements for ontology modeling in the field of moving image indexing. The results included the development of conceptual thesauruses that can be used as "value vocabularies" for semantic annotations. Similarly, the semantic modeling of the ontology called "Moving Image Representation Categories" was carried out, consisting of the Categories, Generic From and Specific From classes. The conclusion is that knowledge organization systems and the semantic web are strong allies in the development of solutions and innovations for challenges and problems in the field of moving image organization and representation. The Simple Knowledge Organization System is a model that, by differentiating terms and concepts, adapts to the dynamic environment of the web and its concepts used as value for metadata at the image level, i.e. annotations in specific regions. In addition, the ontology conceptualized the process of indexing images, allowing the creation of classes for categories and levels of indexing, as well as the creation of object properties to designate the hierarchical structure of generic and specific concepts.

Keywords: moving image; semantic web; SKOS; OWL; semantic annotation.

LISTA DE FIGURAS

Figura 1 - Principais níveis de abstração na base de dados de um museu de arte.....	37
Figura 2 - Estrutura de níveis, tipos e valores visualizados no LodLive.....	59
Figura 3 - Captura de tela do Protégé do arquivo ada_ontology.owl.....	60
Figura 4 - Captura de tela do Protégé do arquivo ada_ontology.owl.....	60
Figura 5 - Captura de tela do Protégé do arquivo ada_ontology.owl.....	61
Figura 6 - Captura de tela do Protégé do arquivo ada_ontology.owl.....	62
Figura 7 - Captura de tela do Protégé do arquivo ada_ontology.owl.....	62
Figura 8 - Captura de tela do Protégé do arquivo ada_ontology.owl.....	63
Figura 9 - Filmontologia AdA na ferramenta OntoViz.....	66
Figura 10 - Valor Acoustics na ferramenta OntoViz.....	67
Figura 11 - Definição de um valor de anotação como grafo RDF	68
Figura 12 - Metadados do corpus AdA codificados como dados RDF	69
Figura 13 - Captura de tela do arquivo metadata_corpus.ttl baixado no github.....	70
Figura 14 - Plataforma online para exploração de anotação da AdA.....	71
Figura 15 - Paradigma semântico	84
Figura 16 - Campo semântico 2.....	84
Figura 17 - Um conjunto provisório de tipos de KOS.....	90
Figura 18 - Modelo para construção de conceitos de Dahlbe.....	94
Figura 19 - Triângulo do conceito	95
Figura 20 - Referentes conceituais	96
Figura 21 - Hierarquia de características na forma de escada	96
Figura 22 - Resultado de busca baseada em palavra-chave.....	107
Figura 23 - Desambiguação e mapeamento de entidade	108
Figura 24 - Interface do buscador.....	109
Figura 25 - Desambiguação de Terra	110
Figura 26 - Mapeamento e desambiguação de terra	110
Figura 27 - Propriedades compartilhadas pelas entidades da classe	111
Figura 28 - Condições verdade para a sentença do círculo está dentro do quadrado	114
Figura 29 - Triângulo do significado (a) para sentenças e (b) usando termos lógicos.....	115
Figura 30 - Imagem clássica do hipertexto.....	119
Figura 31 - Slides de Berners-Lee na primeira Conferência Internacional da WWW	120
Figura 32 - Triplo RDF.....	132
Figura 33 - Rótulo lexical preferencial.....	133
Figura 34 - Rótulo lexical alternativo SKOS.....	134
Figura 35 - Rótulo lexical alternativo usado para abreviaturas	135
Figura 36 - Relações semânticas SKOS	136
Figura 37 - Conceito e rótulo lexical skos:altLabel.....	147
Figura 38 - Conceito e propriedades de relação semântica entre conceitos	148
Figura 39 - Conceito LD service Library of Congress e propriedades de mapeamento	149
Figura 40 - As 3 bases fundamentais da arquitetura da web	162
Figura 41 - Caracterização do RDF	162
Figura 42 - Triplo RDF.....	163

Figura 43 - Fase de construção da metodologia 101	173
Figura 44 - Web Annotation Data Mode	182
Figura 45 - Classe FragmentSelector.....	183
Figura 46 - Web Annotation Vocabulary	184
Figura 47 - Macroestrutura do vocabulário USP.....	198
Figura 48 - KOS 1: Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento	201
Figura 49 - Propriedade skos:hasTopConcept e skos:TopConceptOf.....	202
Figura 50 - Propriedade skos:inScheme	202
Figura 51 - Grafo com a propriedade skos:inScheme	203
Figura 52 - Tesouro conceitual	203
Figura 53 - Índice alfabético renderizado na ferramenta SKOS Play!	205
Figura 54 - Índice alfabético expandido renderizado na ferramenta SKOS Play!	205
Figura 55 - Árvore hierárquica estática renderizada na ferramenta SKOS Play!	206
Figura 56 - Índice KWIC renderizado na ferramenta SKOS Play!	206
Figura 57 - Mapeamento entre conceitos renderizado na ferramenta SKOS Play!.....	207
Figura 58 - Planilha de conversão de Excel para RDF na estrutura do SKOS Play! Convert.....	207
Figura 59 - Visualização do grafo do esquema de conceito “quem:”	208
Figura 60 - Propriedade skos:hasTopConcept.....	209
Figura 61 - Tesouro conceitual KOS 2	210
Figura 62 - KOS 2: Esquema de conceitos da categoria locais (ONDE) de representação da imagem em movimento	210
Figura 63 - A visualização do esquema de conceito “onde:”, KOS 2, através de um grafo ..	211
Figura 64 - Conceito SKOS onde:brasil.....	211
Figura 65 - Relações semânticas SKOS	212
Figura 66 - KOS 3: Esquema de conceitos das categorias de representação da imagem em movimento QUEM /QUANDO/ONDE/OQUE	215
Figura 67 - Alinhamento entre conceitos dos esquemas KOS 2 e KOS 3.....	216
Figura 68 - Grafo do alinhamento (Mapeamento hierárquico e associativo).....	217
Figura 69 - KOS 4: Esquema de conceitos dos países da américa latina	218
Figura 70 - Rede semântica entre KOS 3 e KOS 4 (reutilização do conceito países:brasil)..	218
Figura 71 - Conceito categorias:saopaulo renderizado no grafo	219
Figura 72 - KOS na forma de planilha	219
Figura 73 - Coleção de conceitos da categoria ONDE.....	220
Figura 74 - Renderização da coleção 1 no serviço online de grafo.....	221
Figura 75 - Renderização da coleção 2 no serviço online de grafo.....	221
Figura 76 - Representação da estrutura de facetas da ontologia CRIM	222
Figura 77 - Diagrama de Classe da ontologia CRIM	224
Figura 78 - Grafo de representação da ontologia CRIM	226
Figura 79 - Classes da ontologia CRIM no editor Protégé.....	227
Figura 80 - Classe Categorias da ontologia CRIM no editor Protégé	227
Figura 81 - Classe De Genérico da ontologia CRIM no editor Protégé.....	228
Figura 82 - Classe De Específico da ontologia CRIM no editor Protégé.....	228
Figura 83 - Propriedade temDeGenerico da ontologia CRIM no editor Protégé	229

Figura 84 - Propriedade temDeEspecifico da ontologia CRIM no editor Protégé.....	229
Figura 85 - Instância QUEM da classe Categorias da ontologia CRIM no editor Protégé	230
Figura 86 - Instância Ponte da classe De Genérico da ontologia CRIM no editor Protégé....	230
Figura 87 - Propriedade PertenceCategoria da ontologia CRIM no editor Protégé	231
Figura 88 - Propriedade valorDeEspecifico da ontologia CRIM no editor Protégé.....	231
Figura 89 - Instância Ponte das Bandeiras da classe De Específico da ontologia CRIM no editor Protégé.....	232
Figura 90 - Ontologia CRIM modelada em Turtle	233
Figura 91 - Prefixos e URIS das classes OWL e indivíduos SKOS.....	241
Figura 92 - Criação da ontologia CRIM.....	241
Figura 93 - Classes OWL e subclasses SKOS.....	242
Figura 94 - Classe :Categoria serializado em turtle e no editor de ontologia Protégé	242
Figura 95 - Propriedades serializadas em turtle e no editor de ontologia Protégé	243
Figura 96 - Instâncias serializadas em turtle e no editor de ontologia Protégé	243
Figura 97 - Prefixo e URIs	244
Figura 98- Imagem anotada no nível da dimensão espacial.....	244
Figura 99 - Esquema de conceitos SKOS.....	243
Figura 100 - Metadados do corpus de imagem em movimento em turtle	247
Figura 101 - Dimensões da representação temática	248
Figura 102 - Imagem anotada no nível da dimensão espacial.....	248
Figura 103 - Propriedade DE específico.....	249
Figura 104 - Anotação de recurso de vídeo com a ontologia CRIM e o Web Annotation Data Model.....	250
Figura 105 - Fotograma do filme "O canto da saudade"	251
Figura 106 - Anotação de recurso com a ontologia CRIM e o Web Annotation Data Model.....	251

LISTA DE QUADROS

Quadro 1 - Resumo da proposta de Shatford, relacionando-a à teoria de Panofsky	33
Quadro 2 - Categorias para representação da imagem	34
Quadro 3 - Proposta de representação do conteúdo das imagens.....	34
Quadro 4 - Identificação de objetos na análise documentária	40
Quadro 5 - Representação da imagem em movimento e planos cinematográficos	41
Quadro 6 - Termos não convencionais para indexação de filmes	42
Quadro 7 - Organização de imagem em movimento com categorias diegéticas.....	44
Quadro 8 - Indexação de filmes a partir de categorias diegéticas	45
Quadro 9 - Dados provenientes de bibliotecas, arquivos e museus.....	47
Quadro 10 - Metadados e níveis de anotação	56
Quadro 11 - Classes AdA ontologia	58
Quadro 12 - Entidades da ontologia AdA	59
Quadro 13 - Exemplo do valor Acoustics	64
Quadro 14 - Exemplo do valor Music	65
Quadro 15 - Exemplo do valor Crescendo	65
Quadro 16 - Marcadores semânticos	85
Quadro 17 - Categorias dos KOS	88
Quadro 18 - Classificação de Hodge de KOSs.....	89
Quadro 19 - Três tipos gerais de predicções e três tipos de características	97
Quadro 20 - Características que constituem um conceito	97
Quadro 23 - Cinco primitivas semânticas LN, LD e operadores lógicos	104
Quadro 24 - Lógicas descritivas (DL)	105
Quadro 25 - Quadro sinótico das abordagens semânticas em Linguística	113
Quadro 27 - Classificação para a semântica de Uschold.....	115
Quadro 28 - Categorias para a semântica de Sheth, Ramakrishnan; Thomas (2005)	116
Quadro 29 - Novos padrões de elaboração de tesouros	128
Quadro 31 - Rótulo lexical preferencial SKOS	133
Quadro 32 - Rótulo lexical alternativo SKOS	134
Quadro 33 - Rótulo lexical alternativo usado para abreviaturas	134
Quadro 34 - Linked open data 5 star	150
Quadro 35 - Produtores de conjuntos de dados LOD usando vocabulários KOS LOD.....	151
Quadro 36 - Produtores de vocabulários (pvs) envolvidos no enriquecimento de kos	152
Quadro 37 - Modelo DERA (Domínio, Entidade, Relações, Atributos).....	158
Quadro 38 - Definição de classes e propriedades RDFS.....	163
Quadro 39 - Propriedades OWL.....	166
Quadro 41 - Busca estratégica nas bases da ciência da informação	189
Quadro 42 - Palavras-chave inseridas na estratégia	190
Quadro 43 – Lista e nuvem das 20 palavras-chave mais utilizadas nas publicações	191
Quadro 44 - Quadro das categorias de representação da imagem fotográfica	197
Quadro 45 - Propriedade skos:scopeNote	213
Quadro 46 - Propriedade skos:definition.....	213
Quadro 47 - Excel do esquema.....	214

Quadro 48 - Classes e propriedades da ontologia CRIM	225
Quadro 49 - Planilha com as entidades da ontologia CRIM	226

SUMÁRIO

1	INTRODUÇÃO.....	14
2	OBJETIVOS	21
3	ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO DE IMAGENS EM MOVIMENTO	22
3.1	ANÁLISE DOCUMENTÁRIA E RECUPERAÇÃO DA INFORMAÇÃO DA IMAGEM EM MOVIMENTO.....	38
3.2	SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO DA IMAGEM EM MOVIMENTO NA WEB.....	45
3.2.1	<i>Linked Open Data na FIAF.....</i>	<i>49</i>
3.2.2	<i>Anotação semântica da imagem em movimento.....</i>	<i>52</i>
3.2.3	<i>AdA ontologia</i>	<i>57</i>
3.2.4	<i>Plataforma I-Media-Cities.....</i>	<i>71</i>
4	SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO NA WEB	76
4.1	LINGUAGEM NATURAL E ARTIFICIAL NA ORGANIZAÇÃO DO CONHECIMENTO	78
4.2	SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO.....	85
4.2.1	<i>Teoria do conceito e elementos da lógica</i>	<i>91</i>
4.3	SEMÂNTICA FORMAL	107
4.4	WEB SEMÂNTICA	119
4.5	ONTOLOGIA LEVE (<i>LIGHTWEIGHT ONTOLOGY</i>)	127
4.5.1	<i>KOS LOD</i>	<i>144</i>
4.5.2	<i>Enriquecimento semântico de dados estruturados e humanidades digitais.....</i>	<i>156</i>
4.6	ONTOLOGIA FORMAL (<i>HEAVYWEIGHT ONTOLOGY</i>).....	158
4.6.1	<i>SPARQL 1.1.....</i>	<i>175</i>
4.7	SKOS E OWL.....	178
4.8	WEB ANNOTATION DATA MODEL.....	181
5	ABORDAGEM METODOLÓGICA: MODELO SEMÂNTICO.....	185
6	RESULTADOS	201
6.1	ESQUEMA DE CONCEITO “QUEM”	201
6.2	ISO25964 E SKOS/SKOS-XL	208
6.3	MAPEAMENTO HIERÁRQUICO E ASSOCIATIVO	214
6.4	REDE SEMÂNTICA.....	217
6.5	COLEÇÃO DE CONCEITOS.....	220
6.6	MODELAGEM SEMÂNTICA DA ONTOLOGIA CRIM	222
6.6.1	<i>Consulta SPARQL na ontologia CRIM.....</i>	<i>234</i>
6.7	CONCEITOS SKOS E CLASSES OWL	237
6.8	METADADOS DO CORPUS	245
6.9	ANOTAÇÃO SEMÂNTICA DA IMAGEM	247

7	CONCLUSÕES.....	253
	REFERÊNCIAS.....	258

1 INTRODUÇÃO

Nos últimos anos, a transformação digital, gerada pelo World Wide Web (WWW) e pela internet, nos ciclos de dados e informação, atingiu diversas esferas da sociedade, provocando mudanças nos processos e sistemas de organização do conhecimento (SOC). De tal maneira, o contexto de informação e dados digitais no ambiente web, ampliou a atuação dos SOC e dos processos de organização do conhecimento, pois aumentou a demanda da classificação e da categorização de banco de dados, repositórios digitais, bases de conhecimento, aplicações web, sistemas computacionais, modelos de representação do conhecimento para inteligência artificial (IA), entre outros. O SOC no ambiente web age na organização do conhecimento e na recuperação da informação, além de que, atua na criação de bases de conhecimento e na interoperabilidade de dados.

Essas expansões das aplicações de SOC, também abrangeram os arquivos, bibliotecas e museus, resultando no movimento de digitalização de documentos e publicação de metadados. O que gerou uma crescente demanda por informação, dados e metadados, relacionados ao patrimônio histórico e cultural, tanto estruturados quanto não estruturados, em canais na web. Nesse panorama, o ambiente da web foi empregado para democratizar o acesso aos bens culturais através da publicação de acervos e catálogos digitalizados e bases de dados online. O surgimento de websites e plataformas online, de organizações e instituições públicas e privadas de artes, cultura e patrimônio, possibilitou aproximar o público das atividades promovidas nesses espaços, assim como, compartilhar a missão, objetivos e as ocupações de cada lugar.

Nesse sentido, diferentes instituições, como museus, arquivos e bibliotecas do Brasil, publicaram na web parte de seus acervos na forma de objetos digitais, bem como, publicaram suas bases de dados, metadados dos itens e demais informações dos catálogos de suas coleções. Como, por exemplo, a Biblioteca Nacional¹, o MASP², o MAM³, a Pinacoteca de São Paulo⁴, o IPHAN⁵, o Arquivo Público do Estado de São Paulo⁶ e o IBRAM⁷, que concentra os acervos online em diferentes projetos, como o *Google Arts & Culture* e o Tainacan, que faz a publicação e a difusão de coleções de mais de 1000 instituições. Essas são algumas das instituições de

¹ <https://antigo.bn.gov.br/explore>

² <https://masp.org.br/acervo>

³ <https://mam.org.br/acervo>

⁴ <https://acervo.pinacoteca.org.br/online/>

⁵ <http://acervodigital.iphan.gov.br/xmlui/>

⁶ <http://www.arquivoestado.sp.gov.br/web/acervo/digitalizados>

⁷ <https://www.gov.br/museus/pt-br/assuntos/os-museus/acervos-online>

grande porte, sendo que instituições menores e com menos recursos também investiram na democratização de acesso aos seus acervos, coleções e catálogos.

No cenário brasileiro de arquivos audiovisuais, também ocorreram iniciativas para democratizar o acesso aos bens culturais e ao patrimônio cinematográfico, usando o ambiente web para publicar bases de dados de suas filmografias e alguns casos, também foram criadas bases com os metadados da documentação e da bibliografia. Além disso, há exemplos da difusão de conteúdo online na forma de vídeo digital, provenientes de imagens em movimento digitalizadas, apesar de haver menos projetos brasileiros com esse tipo de difusão. Alguns exemplos de iniciativas do patrimônio audiovisual são: a Bases de dados da Cinemateca Brasileira⁸, o Banco de Conteúdos Culturais da Cinemateca Brasileira⁹, a Bases de dados da Cinemateca Capitólio¹⁰, a Base de dados da Cinemateca Pernambucana¹¹, o Acervo online do MIS/SP¹², a Base de dados do acervo de filmes do LUPA¹³, a Base de dados da Coleção Esdras Baptista¹⁴ do LUPA, o Repositório digital de fotografias e filmes ‘Niterói em imagens’¹⁵ do LUPA, a Base de dados online do CTAV¹⁶ e a Hemeroteca de cinema: Minas Gerais¹⁷.

No cenário da Europa, a Comissão europeia elaborou uma recomendação que incluía uma série de iniciativas para promover o patrimônio cinematográfico e o poder público apoiou essas iniciativas, comprometendo-se a executá-las para a divulgação do patrimônio, resultando em diferentes iniciativas.

Portanto, devido ao processo de digitalização de acervos, da criação de metadados e informação, de bibliotecas, arquivos e museus, foi possível criar iniciativas para aprimorar a organização de metadados, para melhorar a descoberta e recuperação de informação, assim como, possibilitou a modelagem e exploração dos dados de instituições com propósito de realizar iniciativas científicas, educacionais e culturais. Em consequência disso, o patrimônio cultural tornou-se a área mais ativa de aplicação das tecnologias de *linked data* e **Web Semântica** (Hyvönen, 2020).

O advento da Web Semântica trouxe novas abordagens para a organização do conhecimento, bem como, a representação e a recuperação da informação, nos ecossistemas de

⁸ <https://www.cinemateca.org.br/bases-de-dados/>.

⁹ <http://www.bcc.org.br/>

¹⁰ <http://capitolio.org.br/portal>

¹¹ <http://cinematecapernambucana.com.br/acervo/>

¹² <https://www.mis-sp.org.br/acervo/online>

¹³ <https://lupa.uff.br/base-de-dados-do-acervo-de-filmes-do-lupa/>

¹⁴ https://esdras.lupa.uff.br/?page_id=52

¹⁵ <https://niteroiemimagens.uff.br/>

¹⁶ <http://antigo.ctav.gov.br/basededados/>

¹⁷ <https://www.hemerotecacinema.org/>

informação do WWW, que apresentam uma grande variedade de fontes de dados e uso de metadados. A Web Semântica constituiu a possibilidade de representar os conteúdos tratados no interior das páginas web, ao permitir identificar recursos informacionais como sendo documentos, pessoas, objetos concretos e conceitos abstratos, por meio de seus metadados.

A pilha de tecnologias semânticas de organização do conhecimento e recuperação da informação, no contexto da Web Semântica, surgem integradas ao desenvolvimento de SOC, a fim de desenvolver soluções e suprir necessidades informacionais presentes no ambiente web contemporâneo, como a classificação e categorização de dados, a modelagem semântica de dados, a conceituação formal de um domínio e a busca semântica de termos e conceitos.

A Web Semântica se desenvolveu no sentido das recomendações e especificações acerca de declarações *Resource Description Framework* (RDF), do conjunto de práticas do *Linked Data* e das ontologias leves, como o *Simple Knowledge Organization System* (SKOS). A documentação do W3C, também engloba recomendações e normativas para o desenvolvimento de ontologias para web, como *Ontology Web Language* (OWL), das linguagens de busca de triplos (Query/SPARQL) e dos conjuntos de dados abertos e vinculados da nuvem *Linked Open Data* (LOD). No modelo de dados RDF, por meio da sintaxe de suas declarações, são atribuídos identificadores globais, para recursos capazes de identificar qualquer “objeto do mundo real” ou uma coisa (*thing*), sendo que as recomendações não determinam o escopo de significação de um recurso informacional.

Nesse contexto, o SKOS se apresenta como um modelo para publicar dados na web e diminuir os problemas da aplicação de tesouros lexicais em um ambiente dinâmico como a web, pois embora os tesouros tenham se adaptado à sua publicação e consulta em ambientes digitais, uma abordagem centrada em conceitos fornece resultados mais adequados aos processos de indexação (Pastor-Sanchez; Mendez; Muñoz, 2009).

O contexto da Web Semântica pode auxiliar na representação de recursos não baseados em texto, como a imagem e a imagem em movimento, mediante processos de anotação semântica usando KOS LOD como vocabulário de valor, para identificar conceitos e extrair termos de regiões específicas do recurso. Dessa forma, é possível anotar recursos com imagem em movimento, utilizando SOC desenvolvidos a partir das recomendações e especificações do W3C, como KOS LOD e ontologias OWL. Um vocabulário KOS LOD significa que ele segue os princípios do *Linked Data*, possui acesso aberto e é usado como **vocabulários de valor**, os exemplos incluem: tesouros, listas de códigos, listas de termos, esquemas de classificação, listas de títulos de assuntos, arquivos de autoridade, dicionários geográficos digitais, esquemas de

conceito, entre outros, sendo exemplos vocabulários como o EuroVoc, serviços como a *Library of Congress Linked Data Services* e o Getty Vocabularies LOD.

Devido às dificuldades e impossibilidades de realizar uma transcodificação e tradução de um signo imagético para um textual, existem inúmeros desafios no desenvolvimento de instrumentos terminológicos, como tesouros e vocabulários controlados, para representar imagens empregando texto (Lima, 2016). Desta forma, a literatura demonstra que em comparação a catalogação da imagem, os processos de representação temática e os SOC, para imagens em movimento são pouco aplicados no ambiente do patrimônio cultural da imagem em movimento (Domínguez-Delgado; López-Hernández, 2017), assim como, a imagem fotográfica (Smit, 1996; Lancaster, 2004). Devido a isso, buscou problemas apontados na pesquisa bibliográfica e exploratória, como as dificuldades para a representação de imagem em movimento aplicando palavras e textos, junto as dificuldades na análise, síntese e representação da imagem movimento, devido à polissemia da imagem, produzindo diferentes perspectivas para a representação da imagem.

Visto que, são inúmeras as mudanças que ocorreram no universo da descrição de conteúdo, pois quando foi publicado o *FIAF Cataloguing Rules for Film Archives* (1991), o termo metadado era virtualmente desconhecido (Van Malssen, 2017). Ainda segundo a autora, o manual mais recente de catalogação da FIAF, 2016 *FIAF Moving Image Cataloguing Manual*, traz a descrição de imagens em movimento para a era digital, orientando a criação de um conjunto abrangente de metadados para materiais de imagens em movimento. Porém, apesar do extenso e importante trabalho desenvolvidos pela Federação de Arquivos Fílmicos *International Federation of Film Archives* (FIAF) e pelo *European Committee for Standardization* (CEN) na publicação de regras e catalogação de filmes, “para além dos aspectos técnicos ou formais dos filmes contemplados nestas regras, existem alguns aspectos importantes sobre a análise do conteúdo do filme que não foram tidos em consideração” (Domínguez-Delgado; López-Hernández, 2017a, p. 655, tradução nossa). Desse modo, a representação temática da imagem em movimento e o desenvolvimento de SOC, que são os vocabulários controlados e tesouros, para criação de descritores, é uma atividade de menor frequência em relação às atividades que envolvem as regras de catalogação, já consolidadas de instituições internacionais.

O cenário digital potencializou os instrumentos e os métodos referentes a busca semântica dentro das imagens, com objetivo de recuperação da informação e organização de conhecimento. A indexação de imagens com anotação semântica opera para categorizar e classificar recursos por meio de metadados no nível da anotação, ou seja, são anotados pontos

de interesse específicos ou regiões de interesse em partes da imagem. E são identificados conceitos que são retirados de KOS LOD, que podem ser estruturados como instrumentos terminológicos, como listas, índices de assuntos, tesouros, taxonomias e vocabulários controlados, entre outros.

Por conseguinte, surgiram projetos-pilotos internacionais que aplicaram tecnologias semânticas em suas iniciativas, como a grupo de trabalho da FIAF para investigar o LOD e o desenvolvimento de uma ontologia FIAF de catalogação; também a plataforma I-Media-Cities de anotação semântica sobre as imagens em movimento do patrimônio urbano europeu e por fim, a AdA Filmontology, ontologia para anotação semântica de corpórea audiovisual com expressões multimodais cinematográficas de imagens de movimento. Todos esses projetos são abordagens que estão trazendo novos enfoques e experimentando metodologias e teorias na esfera digital na web para auxiliar a representação do conhecimento com a indexação de itens de referência na forma de imagens em movimento.

Destarte, a representação da informação e a organização do conhecimento da imagem em movimento tem importância nos contextos cultural, histórico, científico, artístico, entretenimento, entre outras conjunturas. Neste sentido, as instituições, com bibliotecas, arquivos, museus e cinematecas, possuem acervos que demandam a construção de SOC e a indexação de imagens em movimento, visando melhorar o tratamento dos metadados temáticos e promover a interoperabilidade semântica.

A enorme herança cultural digital que está sendo acumulada na web, requer novas formas de representação do conhecimento, para que os usuários sejam capazes de realizar buscas e recuperar as informações que não estão disponíveis, com o emprego de terminologias de domínio específicas. O que pode resultar em novas formas de acesso e difusão do patrimônio cultural, sendo processos que apanhem a democratização ao acesso de bens culturais digitais e digitalizados.

Portanto, a imagem em movimento no ambiente web, interage com diferentes personas e possui inúmeros cenários e arranjos, como: a) a representação temática e descritiva para a recuperação da informação em bases de conhecimento, banco de dados e repositórios digitais de coleções e acervos, b) a preservação, proveniência e arquivamento de objetos digitais em plataformas web e c) a curadoria digital e os ciclos de vida dos dados.

Esta pesquisa, no formato de tese acadêmica, se intitula “Modelagem semântica de sistemas de organização do conhecimento para anotação semântica de categorias de representação de imagens em movimento no ambiente web”, tem como objeto e questão central

o desenvolvimento de KOS LOD para realizar a anotação semântica de imagens em movimento no ambiente web.

Nesse sentido, foi feita uma releitura dos instrumentos desenvolvidos por Smit (1996), para os problemas de representação temática do conteúdo e análise de dimensões da imagem fotográfica, que descreve a imagem a partir das perguntas: QUEM, QUANDO, ONDE, O QUE (Bléry, 1976), assim como, foi realizada uma transposição desses instrumentos de indexação e métodos, para o ambiente da Web Semântica. Ademais, também trata as dimensões de representação da imagem e os diferentes níveis de análise, como, por exemplo, os níveis pré-iconográfico, iconográfico, iconológico (Panofsky, 1979) e a diferenciação entre os aspectos factuais da imagem e os expressivos (Shatford, 1986).

Devido a isso, como parte dos resultados da pesquisa, foi experimentado a modelagem de diferentes SOC para a organização do conhecimento e recuperação da informação da imagem em movimento, utilizando o SKOS, modelo de dados que serve para publicar SOC no ambiente web. A modelagem dos dados considerou diferentes tipos de SOC, como o tesouro, o campo semântico, a lista de conceitos e a coleção ordenada de conceitos, que podem ser usados como vocabulários de valor para KOS LOD e para a indexação de metadados no nível da anotação. Dessa forma, foi experimentando um sistema híbrido formal e semiformal, vinculando os conceitos dos SOC modelados em SKOS na ontologia OWL.

Portanto, os resultados também compreenderam a modelagem de uma ontologia de domínio (OWL) da representação temática da imagem em movimento, na metodologia a conceituação foi definida com base na estrutura analítica-metodológica do processo de indexação, partindo de metodologias para a indexação de imagem selecionadas na literatura, condensando as estratégias epistemológicas nas classes e propriedades da ontologia de domínio para a web.

Em tais termos, realizou-se a estruturação de metadados de um corpus de imagem em movimento (obras cinematográficas), que foram codificados como dados RDF usando vocabulários existentes como o Linked Movie DataBase, Schema.org e DBpedia Ontology.

Além disso, foi desenvolvido um padrão de anotação que considerou o modelo de dados para anotação na web (*Web Annotation Data Model*), para que os itens dos arquivos de vídeo (objetos digitais) também sejam anotados usando uma ferramenta de anotação de vídeo que permite criar anotações conforme ontologia e indexar com diferentes níveis o corpus de filmes (itens).

Diante do exposto, a pesquisa tem como questão central: como realizar a modelagem semântica de KOS LOD para a anotação semântica de imagens em movimento e a modelagem

de ontologia de domínio para representar processos de organização do conhecimento, como a indexação no ambiente web?

Considerando as questões expostas, a hipótese de pesquisa conjectura que a modelagem semântica de ontologias, com os KOS LOD, potencializa e auxilia os processos de organização do conhecimento, indexação da imagem em movimento e recuperação da informação no ambiente da web, possibilitando a exploração de novas metodologias na área da organização e representação do conhecimento e da informação.

2 OBJETIVOS

Objetivo geral:

Realizar uma pesquisa bibliográfica e exploratória sobre os aspectos dos SOC e os processos de indexação de imagens em movimento, no ambiente web; estabelecendo uma abordagem aplicada sobre os métodos e instrumentos para o desenvolvimento e modelagem de SOC no ambiente web, para anotação de imagens em movimento.

Objetivos específicos:

- Propor a elaboração de SOC no modelo de dados SKOS, usando diferentes instrumentos de modelagem.
- Propor um modelo de conhecimento (*knowledge model*) de domínio baseado em ontologia, das categorias de representação temática da imagem em movimento QUEM, QUANDO, ONDE e O QUE.

3 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO DE IMAGENS EM MOVIMENTO

A seção trata a representação e organização do conhecimento da imagem em movimento, a partir de teorias da ciência da informação, na perspectiva do desenvolvimento de SOC e dos processos de organização do conhecimento, mais especificamente a indexação.

Inicialmente, a pesquisa faz uma breve introdução ao período conhecido como primeiro cinema, a fim de identificar os aparatos técnicos e os sistemas de representação da realidade que utilizavam imagens fotográficas e em movimento. Esses aparatos foram responsáveis pela criação de diversas máquinas e ambientes, tanto individuais quanto coletivos, comercializados como grandes inovações e invenções. Desse modo, reflexiona sobre os efeitos da “gênese automática” no conceito de realismo e como fotografia foi tratada no domínio da semiótica (índice e ícone) e no campo da organização do conhecimento.

Além do que, versa-se sobre as teorias no campo da Ciência da Informação, que refletem sobre as categorias de representação da imagem fotográfica, como os instrumentos tratados por Smit (1996), que referenciam o teórico do cinema Dubois (1993) e sua teoria sobre a relação entre a fotografia analógica e o referente real. Nesse sentido, são estabelecidas conexões com as teorias de Panofsky (1979) para identificar os aspectos iconográficos da imagem e a metodologia de Shatford (1986) para indexação de fotografias, utilizando a intensão e extensão dos conceitos (genérico e específico).

Realizou-se pesquisa bibliográfica, descritiva e exploratória em fontes nacionais e internacionais de periódicos da área de Ciência da Informação, bem como em bibliografia pertencente a área dos estudos de cinema. Foram abordadas as contribuições teórico-metodológicas provenientes da literatura científica em Ciência da Informação e sua interdisciplinaridade com a área de estudos cinematográficos. Nessa proposição, as categorias para a indexação de termos (QUEM, QUANDO, ONDE, O QUE) e construção de SOC são apresentadas com ênfase nos seus aspectos de conceituação e princípios norteadores diante da literatura das áreas de Organização e Representação do Conhecimento em Ciência da Informação.

A organização e representação do conhecimento da imagem em movimento, a qual é um tipo de documento não textual, envolvem procedimentos como a indexação de termos e a construção de SOC. Diante da multiplicidade de sentidos que podem ser atribuídos durante a análise de fotografias e imagens em movimento, surgem desafios relacionados ao significado das imagens na estruturação de conceitos para instrumentos terminológicos. (Lima, 2016;

Domínguez-Delgado; López-Hernández, 2017). Por conseguinte, a representação do conhecimento da imagem em movimento apresenta muitos desafios para as instituições que possuem esse tipo de documento devido principalmente a polissemia da imagem em movimento.

Dessa forma, apesar do trabalho realizado no campo dos arquivos cinematográficos ao longo dos quase 120 anos de história do cinema, envolvendo colecionadores, cinematecas, FIAF, bibliotecários, arquivistas e UNESCO, e tendo sido relativamente resolvidos muitos problemas técnicos de conservação e catalogação de documentos cinematográficos, as etapas do tratamento documentário foram negligenciadas ou esquecidas (Domínguez-Delgado; López-Hernández, 2017). Contudo, para a recuperação pelos usuários, destaca-se a análise de conteúdo documental, como “tarefa de vital importância em qualquer sistema de informação audiovisual, uma vez que nela os conteúdos dos documentos são identificados e descritos” (Domínguez-Delgado; López-Hernández, 2017, p. 74, tradução nossa).

De acordo com Gracy (2018), são inúmeras as informações valiosas sobre criadores, locais, eventos, tópicos, características de objetos e ações institucionais, que não são suficientemente representadas, além disso, as conexões possíveis ficam presas nos registros dos recursos e documentos audiovisuais, muitas vezes essas informações são encontradas apenas por descoberta acidental.

Segundo Lima (2016) uma das principais dificuldades encontradas pelos catalogadores e indexadores para recuperação da informação é a justaposição de informações, pois todos os documentos apresentam informações sobre o seu suporte e sobre o seu conteúdo. Além de que, devido à polissemia das imagens, a representação temática (classificação e indexação) de um “documento visual, sonoro ou visual e sonoro, como as imagens em movimento, envolvem não só o que vemos ou ouvimos, mas também o que a imagem e o som nos sugerem. A informação audiovisual pode ser criada com determinado objetivo e utilizada para outro objetivo” (Lima, 2016, p. 90).

No campo da teoria cinematográfica, as especificidades do documento audiovisual geraram diversas abordagens teóricas-conceituais, a partir de comparações com teorias da linguística (signo, significado e significante), analisando os possíveis (ou não) papéis dos eixos paradigmáticos e sintagmáticos (Saussure) na formação do sintagma cinematográfico. Essas abordagens foram muito presentes nos estudos de cinema, como em Cris Metz, Jacques Aumont, Deleuze, entre outros.

Do ponto de vista histórico, a fotografia, que precedeu a imagem em movimento, esteve ligada a diferentes paradigmas acerca do realismo na representação, devido aos processos técnicos de formação da fotografia. Os métodos fotoquímicos da fotografia foram a base do surgimento do cinema e da captura do movimento. No surgimento da fotografia foram ressaltados os aspectos **icônicos** da representação fotográfica e os teóricos se pautavam em considerá-la como um espelho do real (Dubois, 1993). Visto que, a popularização da fotografia foi acompanhada de discursos que a enalteciam como um registro neutro e objetivo da realidade; afirmações que perduraram décadas. Como aparece no texto canônico de Bazin “Ontologia da imagem fotográfica” publicado pela primeira vez em 1945, segundo o autor a imagem fotográfica libertou as artes plásticas de representar o real, ao afirmar que a máquina substituía o artista no processo de criação e que o cinema reproduzia a duração real das coisas (Bazin, 1991).

Porém, o desenvolvimento de disciplinas que analisavam a fotografia a partir de outras perspectivas, como a comunicação, sociologia e estudos culturais, o paradigma **simbólico** da imagem fotográfica ganhou força em relação ao icônico, pois mobilizava os códigos e as características sociais na leitura da imagem.

Em contrapartida, a fotografia como **índice** considera o referente como traço do real, ao analisar os processos de formação da imagem-índice como uma gênese automática com implicações semânticas que possibilitam analisar a condição de imagem fotográfica no universo dos signos.

Partindo da relação do signo com seu referente, as abordagens semióticas auxiliam os instrumentos terminológicos, na definição dos descritores e dos conceitos baseados em controle de vocabulário, ao tangenciarem a abordagem de Peirce para signos não-textuais, como os ícones e índices. Na área da ciência da informação, as teorias da semiótica foram usadas para construir metodologias e teorias para signos não textuais, incluindo as imagens fotográficas. A semiologia de Peirce (2005) para signos não textuais, orientou métodos de análise documentária de documentos imagéticos nas diversas camadas de leitura e nos problemas de polissemia da imagem, ao enunciar a conexão física do **índice** com o objeto representado e as conexões mentais do **ícone** com o referente real. Sendo assim, a teoria semiótica de Peirce (2005) é referência para diversas outras teorias no campo dos estudos de informação e imagem, com relevância para a fotografia que tem como atributo principal a “impressão luminosa fotoquímica”, diretamente arrolada as impressões físicas do índice.

No campo do cinema, as disciplinas de teoria do cinema e análise fílmica, conduziram a semiologia do filme nas publicações de Dubois (1993), Metz (1972) e Aumont (1993), visto que, trataram a linguagem, os signos e as significações da imagem em movimento. Esses autores refletiram sobre os diferentes sistemas cinematográficos, ao demonstrar que estes são formados pelas aproximações e afastamentos da referência verbal. Dessa forma, esses autores se criticaram as proposições que associavam a linguagem cinematográfica a linguística, a partir de um tronco comum de códigos acumulados ao longo do tempo. Assim, a semiótica também passou a ser objeto de estudo de outras disciplinas cinematográficas, como o documentário, montagem, estudos culturais, linguagem cinematográfica, entre outros.

Dessa forma, é possível encontrar nas teorias de análise fílmica e nos métodos dos SOC, a influência dos códigos icônicos e não icônicos, na análise da imagem no domínio da linguagem, enfatizando as diferenças e as similaridades entre a linguagem nas imagens e a linguagem verbal. Essas questões acerca dos aspectos icônicos da imagem em movimento cinematográfica, refletem os procedimentos da representação do conhecimento da imagem em movimento, sobre principalmente a indexação e a construção de SOC.

A representação do conhecimento a partir de SOC, no contexto da imagem em movimento, tem como ênfase a relevância entre a palavra e a imagem nos processos de análise e representação dos assuntos do conteúdo da imagem. Todavia, devido à polissemia da imagem, não é possível transpor os processos da análise documentária de documentos textuais para imagens e imagens em movimento (Smit, 1996).

De tal modo, existem desafios para indexar imagens em movimento usando um vocabulário controlado, podendo ser necessário a construção de instrumentos específicos para a indexação. Visto que, na maioria das vezes, não é possível usar instrumentos elaborados para indexar documento textual, pois os termos não abrangem a cobertura necessária para representar os assuntos de uma imagem. Além de que, existe a impossibilidade de categorizar os termos em uma hierarquia pré-existente, no caso de uma instituição com um vocabulário comum para todas as áreas do conhecimento (Macambyra, 2017).

O conceito de documento audiovisual que melhor define os materiais estudados na pesquisa é o de Edmonson (2017):

Documentos audiovisuais são obras que contêm imagens e/ou sons reprodutíveis reunidos em um suporte e que:

- em geral, exigem um dispositivo tecnológico para serem registrados, transmitidos, percebidos e compreendidos;
- o conteúdo visual e/ou sonoro tem duração linear; e

- o objetivo é a comunicação desse conteúdo e não a utilização da tecnologia para outros fins (Edmonson, 2017, p. 27)

Como o espectro de materiais abrangidos pelo termo “documento audiovisual” é extenso, a metodologia e coleta de dados do trabalho concentra-se nas imagens em movimento, assim, excluindo os documentos que são apenas sonoros e as imagens estáticas. A definição de imagem em movimento auxilia a compreensão dos conceitos que envolvem esses tipos de materiais audiovisuais. O Manual de Catalogação de Imagens em Movimento da FIAF 2016 (*The FIAF Moving Image Cataloguing Manual*) especifica que:

Imagens em movimento incluem uma série de materiais sobre os quais sequências de imagens visuais foram gravadas ou registradas e que criam a ilusão de movimento quando projetadas, difundidas ou reproduzidas (por meio de um projetor, aparelho de televisão, computador, software ou dispositivos equivalentes). Tais imagens podem ou não ser acompanhadas de som. A definição inclui imagens em movimento de todos os tipos (FIAF, 2016, p. 4, tradução nossa).

As imagens em movimento e os sons são subjetivos por natureza, não têm essência objetiva, são criadas no cérebro quando uma sequência de imagens fixas é sujeita a uma sucessão rápida, essa sensação ocorre quando uma informação é transmitida para nossos olhos e ouvidos a partir de uma fonte registrada num suporte e reproduzido por uma tecnologia adequada (Edmonson, 2017).

Dessa forma, a reprodutibilidade técnica de imagens e a presença do objeto real na formação da imagem, é o que distinguia a fotografia e a imagem em movimento cinematográfica, das demais reproduções técnicas realizadas anteriormente, como, por exemplo, a xilogravura. Por tais motivos, os conceitos e teorias do campo da fotografia e do cinema, quando inseridos na perspectiva histórico conceitual, expõem questões sobre o referente real na imagem.

O primeiro cinema designa os filmes e práticas surgidos no período entre 1894 e 1908, termo traduzido da expressão inglesa *early cinema*, pois a presença de algumas características constantes no cinema até 1908, são o que fundamentou esse período. Havendo particularidades que são relativas aos modos de produção e exibição dos filmes, à composição e comportamento do público e às formas de representação destes filmes, por estarem sempre em transformação, uma “transformação que é visível na evolução técnica dos aparelhos e na qualidade das películas, na rápida transição de uma atividade artesanal e quase circense para uma estrutura industrial de produção e consumo [...]” (Costa, 2005, p. 35).

O primeiro cinema foi um período que antecedeu a popularização do cinema como exibição e projeção coletiva numa sala escura, sendo uma época em que foram desenvolvidas inúmeras invenções para reproduzir algumas experiências da realidade. Assim, eventos importantes da virada do século XX, como as exposições universais de Paris, contava com diversos ambientes que tentavam reproduzir experiências imersivas de ilusão da realidade, via diferentes sistemas de representação imagética, como os panoramas em movimentos, o cineorama e o mareorama, entre outros panoramas. Essas atrações consistiam em simular viagens no tempo e no espaço, utilizando métodos ilusionistas que misturavam pinturas e fotografias, “eram o que os contemporâneos chamavam de espetáculos ‘totais’ ou ultrarrealistas, muito comuns no início do século XX, entre eles o cinema” (Costa, 2005, p. 25).

O mareorama simulava uma viagem pelo mediterrâneo, entre Marselha e Constantinopla, estava construído num prédio de 40 andares, com capacidade para mil e quinhentas pessoas, uma tela de 15 metros era desenrolada mostrando pinturas com efeitos de luz para as horas do dia, além da cenografia, havia atores e demais artistas, contracenando com os usuários e uma equipe era responsável por simular o movimento das ondas marítimas, incluindo experiência olfativa com algas marinhas (Costa, 2005).

O balão Cineorama foi uma atração que utilizou, ao invés de fotografias, projeções de imagens em movimentos (filmes) para simular uma viagem de balão sobre a Europa, numa projeção de tela de 360 graus com um filme 70 mm de uma viagem real de balão pintada a mão (Costa, 2005). Esse panorama foi patenteado em 1896 por Grimoin-Samson, como ocorria com muitas atrações dessa época que envolviam maquinário específico de exibição (Costa, 2005).

Além de que, o período foi marcado pelo extenso desenvolvimento de objetos ópticos e dispositivos mecânicos, para exibição individual de imagens, que ao invés de projetadas eram exibidas em máquinas, como o quinetoscópio e o mutoscópio.

O quinetoscópio foi uma máquina inventada por Thomas Edison nos Estados Unidos em 1894, onde o usuário podia assistir uma tira de filme de imagens em movimento de lutas, danças, cenas eróticas, animais amestrados, inclusive filmes realizados pela companhia do inventor (Costa, 2005). O mutoscópio foi uma máquina onde também era possível assistir imagens fotográficas com a ilusão de movimento a partir de um visor.

Do mesmo modo, os vaudevilles e feiras assimilaram em suas atrações a fotografia e a projeção de imagens em movimento, como, por exemplo, as fotografias de viagens e os espetáculos de fantasmagoria. Nos Estados Unidos, o vaudeville se tornou a principal forma de atração popular e por questões de licenciamento de exibição, a maioria deles ficaram sob o domínio dos franceses, os irmãos Lumieres.

Posteriormente, no discurso que definia a fotografia como espelho do real, as teorias de Bazin (1991, p. 24) no ensaio “Ontologia da imagem fotográfica”, enaltecem os aspectos de neutralidade, objetividade, realismo e prova da fotografia, conferindo-lhe condição de manifestação da natureza imitando o artista. Segundo o autor, a fotografia libertou as artes plásticas de sua obsessão realista e lhe deu autonomia estética. Assim, sua abordagem fenomenológica com o real, afirma que “pela primeira vez, entre o objeto inicial e a sua representação nada se interpõe, a não ser outro objeto. Pela primeira vez, uma imagem do mundo exterior se forma, automaticamente, sem a intervenção criadora do homem, segundo um rigoroso determinismo” (Bazin, 1991, p. 22). Ainda nessa perspectiva, o instante fotográfico da realidade é tomado pela evolução da imagem em movimento enquanto representação da realidade, em comparação a fotografia, evidenciando os aspectos de realismo no movimento, “a imagem das coisas é também a imagem da duração delas[...]” (Bazin, 1991, p. 24).

No campo da teoria do cinema e da fotografia, os escritos de Walter Benjamin discorreram sobre o aparato da reprodutibilidade técnica, com a possibilidade da objetiva captar as imagens do mundo de forma mais rápida do que os artefatos de reprodução manual. Logo, Benjamin (1992) destacou a reprodutibilidade técnica a partir de um referente real, libertando as artes visuais da obrigação em reproduzir imagens realistas. Deste modo, “pela primeira vez, com a fotografia, a mão liberta-se das mais importantes obrigações artísticas no processo de reprodução de imagens, as quais, a partir de então, passam a caber unicamente ao olho que espreita por uma objetiva” (Benjamin, 1992, p. 76). O autor esclarece que a fotografia e o cinema revolucionaram o cenário de produção de imagens, que era dominado pelas obras de artes plásticas. Não pelo fato de serem reproduzíveis, visto que a reprodução sempre esteve presente na arte, mas porque pela primeira vez a criação da imagem e sua reprodução técnica, ocorriam sem o intermédio da mão humana. Entretanto, Benjamin não tratou a imagem resultante de processos fotográficos como prova do real e não lhe conferiu a condição de espelho do real.

Barthes (1984) no seu último livro a “Câmara clara”, levou a análise da imagem fotográfica a um percurso ensaístico e declarou o “isso foi” da fotografia, condição que refletia a presença do objeto real diante da máquina durante a criação da imagem fotoquímica. Dos modos de interação com a imagem, o autor definiu que no STUDIUM os signos da imagem podem ser nomeados, categorizados, classificados, a partir da conceituação convencional dos assuntos representados na imagem, além disso, carregam os aspectos culturais, antropológicos e sociais. Diferentemente, o PUNCTUM diz sobre a interação com uma imagem cujo efeito não pode ser nomeado por serem da ordem do inominável e de implicação não situável, um

caminho reflexivo que se afasta da teoria semiótica estruturalista acerca do significado e significação da imagem. Sobre o PUNCTUM, “o efeito é seguro, mas não é situável, não encontra seu signo, seu nome; é certo e no entanto aterrissa em uma zona vaga de mim mesmo; é agudo e sufocado, grita em silêncio. Curiosa contradição: é um raio que flutua” (Barthes, 1984, p. 83).

No contexto da semiótica e os estudos da fotografia, Dubois (1993) apresenta uma perspectiva histórica sobre o realismo na fotografia, que foi publicada numa época em que a fotografia era somente analógica e a imagem resultado de processos fotoquímicos. De tal modo, sobre a semiótica e o instantâneo fotoquímico, o autor analisa a relação da fotografia com as dimensões que abordam o real, percorre desde o enfoque da verossimilhança a imagem índice. Desse modo, apresenta três paradigmas que acompanharam os percursos históricos da fotografia e retoma as categorias ícone, índice e símbolo (Peirce, 2005), a partir de um referente real.

Destarte, o paradigma da **fotografia como espelho do real** aborda a percepção que acompanhou o surgimento da fotografia e que vincula a imagem fotográfica com o real devido sua reprodutibilidade técnica e seus procedimentos de acordo as leis da ótica e da química. “A foto é concebida como espelho do mundo, e um ícone no sentido de Ch. S. Peirce” (Dubois, 1993, p. 53). O **ícone** representa o objeto através de sua similaridade, mas não apenas por semelhança aparente ou representação figurativa, como também por afinidades na relação entre o ícone e o objeto. Por tanto, o referente na imagem icônica não precisa ser necessariamente um objeto real ou físico, o foco são as características que remetem ou denotam a um objeto real ou imaginário. A semelhança entre o ícone e seu referente é de ordem mental, como exemplo, a imagem, o diagrama e a metáfora (Peirce, 2005).

Enquanto, no paradigma da **fotografia como transformação do real**, a busca pela representação automática da realidade perdeu o foco, dando espaço analítico para a transformação-interpretação do real e para processos que destacam os códigos presentes na linguagem da fotografia. Desse jeito, a imagem fotográfica se torna um objeto de pesquisa, estudo e práticas das ciências humanas e sociais. “A foto aqui é um conjunto de códigos, um símbolo nos termos peircianos” (Dubois, 1993, p. 53). O **símbolo** também é um signo da ordem mental e a associação entre símbolo e seu referente, ocorre pela convenção e associação de ideias gerais. A associação mental de um símbolo e seu referente é um conjunto de códigos.

Em contrapartida, o paradigma da **fotografia como traço do real**, faz um retorno a questão da relação física existente entre a representação e seu referente. Todavia, o enfoque é o referente como traço do real, determinando o caráter de índice da imagem fotográfica, "a foto

é, em primeiro lugar, índice. Só depois ela pode tomar-se parecida (ícone) e adquirir sentido (símbolo)" (Dubois, 1993, p. 53). O **índice** é um signo onde existe uma conexão física com seu referente. O índice tem como condição que o referente seja real e haja materialidade física, visto que, o signo índice carrega a impressão física, a marca do referente, como um cata-vento que fisicamente mostra a direção do vento, a fumaça que só existe com a combustão da matéria. Para o autor, um exemplo importante de imagem como índice é a fotografia, onde a incidência da luz na emulsão fotográfica, que é sensível à luz, torna a imagem latente no filme fotográfico.

Dubois ainda esclarece sua escolha por Peirce e não Barthes, visto que, o último preconizou a reflexão da imagem fotográfica como um conjunto de códigos e se afastou da obsessão mimética, ao analisar a imagem fotográfica a partir do objeto referencial, com o “isso foi” da fotografia e com o conceito de Punctum. Todavia, o autor considera que Barthes caiu na armadilha não mais da mimese, mas do referencialismo, ao generalizar o princípio da transferência da realidade com uma atitude subjetiva de pretensão ontológica. “Barthes está longe de ter escapado a esse culto a essa loucura da referência pela referência” (Dubois, 1993). Por tais motivos, o autor apresenta análises de teóricos inspirados nos conceitos semióticos de Peirce, na sua noção de índice e em análises que consideram não o produto icônico concluído, mas o processo que leva a “gênese automática” da imagem, tanto anunciada por André Bazin. Desse modo, não enfocou nas implicações ético-estéticas da “gênese automática” da imagem, como a neutralidade, a naturalidade, a objetividade, mas sim nas implicações **lógico-semióticas da noção de índice**. Assim, indo além da fixação com a referência, já que “Peirce vai abrir o caminho, por meio de suas considerações sobre o índice, a uma verdadeira análise da condição da imagem fotográfica” (Dubois, 1993).

Os conceitos apresentados pelo autor são constantemente empregados para referenciar os estudos da imagem fotográfica, principalmente a fotoquímica, visto que, em trabalhos posteriores do autor, é ressaltado que essas teorias da fotografia-índice foram desenvolvidas no contexto da imagem fotoquímica e sua relação direta entre o índice e o traço físico da “impressão luminosa” deixada na emulsão fotoquímica.

Peirce (2005) escreveu sobre a fotografia e a colocou no patamar da linguagem e informação ao identificar as características provenientes dos signos na imagem fotográfica, colocando-a na categoria dos índices, através da semiótica. Sobre a semelhança entre o objeto e a sua representação, o autor afirma que “esta semelhança, porém, deve-se ao fato de terem sido produzidas em circunstâncias tais que foram fisicamente forçadas a corresponder ponto por ponto à natureza. Sob esse aspecto, então, pertencem à segunda classe dos Signos, aqueles que o são por conexão física” (Peirce, 2005, p. 65).

A linguagem cinematográfica, durante o auge da semiologia, sofreu aproximações conceituais com os processos de formação da língua, onde se acreditava ser possível extrair signos uniformizados, como os signos linguísticos e a partir disso, seria viável a identificação de um sintagma da linguagem cinematográfica a partir de unidades correspondentes a diferentes linguagens (Metz, 1972; Aumont, 1993).

A semiologia de Christian Metz apresentará duas fases: num primeiro momento desenvolverá o sentido narrativo do cinema, tendo como influência a linguística estrutural e num segundo momento, “o semiólogo dá uma guinada rumo aos códigos articulando-se com as teorias do texto e da escrita, refazendo vários passos de sua semiologia em prol de uma escrita filmica” (Macedo, 2019, p. 4).

Em tal medida, primeiramente surgiu uma abordagem da linguagem cinematográfica como o conjunto de diferentes códigos cinematográficos que sucederam no decorrer da história do cinema, agrupados num sistema único, tendo como premissa vincular as semelhanças dos diversos códigos num tronco comum, porém essas abordagens negligenciaram as diferenças entre os códigos cinematográficos, pois a linguagem cinematográfica designa um conjunto de sistemas cinematográficos que variam conforme as escolas cinematográficas, gêneros, países e épocas (Metz, 1972).

Posteriormente, Metz (1980) criticou a noção de signo cinematográfico e de unidade mínima na linguagem cinematográfica ao apresentar sua teoria de sistemas e códigos cinematográficos gerais e específicos, mostrando as fragilidades da comparação sem precedente entre o conjunto de sistemas cinematográficos e as análises semióticas linguísticas. Como, por exemplo, equipar um movimento de câmera a um signo linguístico, no sentido de unidade mínima da língua na linguística, visto que, a época semiólogos tentavam conectar o fotograma e os códigos cinematográficos (movimentos de câmera, planos, etc.) a uma unidade mínima cinematográfica, postulando a existência no cinema de um tipo único de signo comparável ao fonema da língua ou as palavras. O autor conclui que a noção de signo não deve ter o papel mais importante na semiologia cinematográfica e filmica, não precisa ser o enfoque principal das análises, pois outras noções importantes foram descobertas, como a criação ou transformação, sintagma e paradigma, sistema e código, expressão/conteúdo, forma/matéria/substância, etc. Um sistema de significação “não é apenas um sistema de signos; as unidades maiores ou menores que o signo, desempenham um papel considerável; o "grau de signo" não deve ser separado dos outros” (Metz, 1980, p. 246).

Smit (1996), afirma que não é possível uma transposição automática dos procedimentos de Análise Documentária para realizar a representação da imagem, por duas razões primordiais:

[...] o estatuto da imagem fotográfica distingue-a do texto; a utilização da imagem fotográfica (e da imagem em geral) não se baliza unicamente por seu conteúdo informacional, mas também por sua expressão fotográfica (Smit, 1996, p. 29).

A autora ainda define os três níveis estabelecidos por Panofsky (1979, p. 47 *apud* Smit, 1996) para a análise das imagens são:

Nível pré-iconográfico: nele são descritos, genericamente, os objetos e ações representados pela imagem;
 Nível iconográfico: estabelece o assunto secundário ou convencional ilustrado pela imagem. Trata-se, em suma, da determinação do significado mítico, abstrato ou simbólico da imagem, sintetizado a partir de seus elementos componentes, detectados pela análise pré-iconográfica;
 Nível iconológico: propõe uma interpretação do significado intrínseco do conteúdo da imagem. A análise iconológica constrói-se a partir das anteriores, mas recebe fortes influências do conhecimento do analista sobre o ambiente cultural, artístico e social no qual a imagem foi gerada

Por isso, define que as camadas **pré-iconográfica** e **iconográfica** dialogam com as análises documentárias, ao estarem pautadas respectivamente nas características genéricas da imagem e pelo assunto convencional tratado na imagem. Diversamente, a dimensão **iconológica** aborda análises que estão fora da imagem, sendo objeto da História, crítica de arte, antropologia, sociologia, “portanto, não pertinente, para o universo documentário” (Smit, 1996, p. 31). A autora cita a “entidade tripartida da imagem para fins documentários” composta pelo **suporte, expressão fotográfica e conteúdo informacional**.

Desse modo, a autora apresenta Shatford (1986), a partir de Panofsky, fazendo uma diferenciação para analisar a representação da imagem, “onde o significado fatural corresponde à pergunta: A IMAGEM É DE QUÊ? E o significado expressivo pergunta: A IMAGEM É SOBRE O QUÊ?” (Shatford, 1986, p. 43 *apud* Smit, 1996). Nesse sentido, Smit (1996) elaborou o quadro 1 para resumir a proposta de Shatford, relacionando-a à teoria de Panofsky, onde é possível observar os desdobramentos das três dimensões para representação da imagem fotográfica a partir dos dois autores citados:

Quadro 1- Resumo da proposta de Shatford, relacionando-a à teoria de Panofsky

PANOFSKY	Exemplo	SHATFORD	Exemplo
Nível pré-iconográfico, significado fatural	Homem levanta o chapéu.	DE genérico	Ponte
Nível iconográfico, significado fatural	Sr. Andrade levanta o chapéu.	DE específico	Ponte das bandeiras
Nível pré-iconográfico + Iconográfico, significado expressivo	Ato de cortesia, demonstração de educação, etc	SOBRE	Transporte urbano, São Paulo. Rio Tietê. Arquitetura, urbanização.

Fonte: Smit (1996, p. 32).

Destarte, no nível do conteúdo informacional, a análise de toda imagem fotográfica ocorre tanto nível pré-iconográfico (genérico) quanto iconográfico (específico). Pois a representação da imagem de uma ponte é constituída tanto pela categoria genérica, que está incluso o conceito “ponte”, como também pela categoria específica, como a “Ponte das Bandeiras” (Smit, 1996), “nessa perspectiva a imagem é ao mesmo tempo, específica e genérica” (Shatford, 1986, p. 47). Assim, a autora aborda a teoria de Shatford (1986), a partir de Panofsky (1979), onde o nível pré-iconográfico da imagem tem relação direta com a condição de semelhança e a identificação do que se vê nela. E o caráter iconográfico se relaciona com o significado atribuído ao conjunto de componentes que formam a imagem, identificados no nível pré-iconográfico.

De tal modo, Smit (1996) elaborou um quadro para resumir a proposta de Shatford (1986), relacionando-a à teoria de Panofsky (1979), onde é possível observar os desdobramentos das três dimensões para representação da imagem fotográfica. Assim, o nível pré-iconográfico apresenta a determinação “a imagem é de que” genérica e seu significado fatural, assim como, no nível iconográfico, efetua a delimitação “a imagem é de que” específica e seu significado expressivo. Por fim, o quadro apresenta o “sobre o que” da imagem que junta os níveis pré-iconográfico e iconográfico.

As categorias QUEM, ONDE, QUANDO, COMO e O QUE, são utilizadas para a análise geral de textos, incluindo a análise documentária (Smit, 1996), ademais, são categorias também indicadas para a indexação de obras de ficção e de imagens. Ainda Smit (1996) atualiza o quadro de representação do conteúdo da imagem acrescentando essas categorias a partir de Bléry (1991), que desenvolveu uma proposta de representação do conteúdo das imagens, Shatford (1986) usa as mesmas categorias. Dessa forma, a autora agrega os conceitos que estabelecem semelhança e conexão física com o referente real e os que identificam o assunto convencional, numa matriz de categorias das determinações De genérico, De específico e SOBRE e desenvolveu uma proposta de representação do conteúdo das imagens no quadro 2:

Quadro 2 - Categorias para representação da imagem

Categoria	Definição geral	De genérico	De específico	SOBRE
QUEM	Animados e inanimados, objetos e seres concretos.	Esta imagem é de quem? De que objetos? De que seres?	De quem, especificamente, se trata?	Os seres ou objetos funcionam como símbolos de outros seres ou objetos? Representam a manifestação de uma abstração?
	Exemplo	Ponte	Ponte das Bandeiras	Urbanização
	Exemplo			Arquitetura dos anos 40
ONDE	Onde está a imagem no espaço?	Tipos de lugares geográficos, arquitetônicos ou cosmográficos	Nomes de lugares geográficos, arquitetônicos ou cosmográficos	O lugar simboliza um lugar diferente ou mítico? O lugar representa a manifestação de uma pensamento astrato?
	Exemplo	Selva	Amazonas	Paraíso (supõe um contexto que permite esta interpretação)
	Exemplo	Perfil de cidade	Paris	Monte Olimpo (como o exemplo anterior)
QUANDO	Tempo linear ou cíclico, datas e períodos específicos, tempo recorrentes.	Tempo cíclico	Tempo linear	Raramente utilizado, representa o tempo a manifestação de uma ideia abstrata ou símbolo?
		Primavera	1996	Esperança, fertilidade e juventude
O QUE	O que os objetos e seres estão fazendo? Ações, eventos, emoções.	Ações, eventos	Eventos individualmente nomeados	Que ideias abstratas (ou emoções) estas ações podem simbolizar?
	Exemplo	Morte	Pietá	Dor (Emoção)
	Exemplo	Jogo de futebol (ação)	Copa do mundo 1994	Esporte

Fonte: Smit (1996).

Os quadros teóricos-metodológicos de Smit (1996) são muito referenciados em publicações que discutem a representação temática da fotografia (imagem estática) e as categorias do quadro 3:

Quadro 3 - Proposta de representação do conteúdo das imagens

CATEGORIAS	REPRESENTAÇÃO DO CONTEÚDO DAS IMAGENS
QUEM	Identificação do “objeto enfocado”: seres vivos, artefatos, construções, acidentes naturais, etc
ONDE	Localização da imagem no “espaço”: espaço geográfico ou espaço da imagem (p. ex São Paulo ou interior de danceteria)
QUANDO	Localização da imagem no “tempo”: tempo cronológico ou momento da imagem (p. ex:1996, noite, verão)
COMO/O QUE	Descrição de “atitudes” ou “detalhes” relacionados ao “objeto enfocado”, quando este é um ser vivo (p. ex. cavalo correndo, criança trajando roupa do século XVIII).

Fonte: Smit (1996, p. 32).

Porém, os teóricos tratados na perspectiva histórica sobre o realismo na fotografia, também pertencem aos estudos do realismo no cinema, como Dubois (1993) e Barthes (1984). Visto que, a imagem em movimento cinematográfica apresenta os níveis de índice e de ícone na representação do conteúdo, logo considerar o movimento é fundamental na indexação de termos e na criação de SOC.

O artigo de Lacerda (1993) que foi usado como referência por Smit (1996) para cunhar o termo *expressão fotográfica*, é um conceito que aborda as três dimensões do registro fotográfico (conteúdo, expressão e forma). E reflete a reprodutibilidade técnica computadorizada da imagem, que havia sido recém popularizada pelo uso das “novas tecnologias de imagens computadorizadas que utilizam scanners, discos óticos” (Lacerda, 1993, p. 46). E a partir disso, conjectura questões sobre a perda da autenticidade e da aura da imagem fotográfica, bem como, questões sobre a “história da técnica daquele registro ou informações que não passam necessariamente pelo conteúdo da imagem reproduzida são indispensáveis ao enriquecimento de sua potencialidade enquanto fonte documental” (Lacerda, 1993, p. 47). Ou seja, a expressão fotográfica remeteria as **técnicas específicas empregadas no processo de captação da imagem** e que deveriam ser preservadas na transferência dos acervos analógicos para a imagem computadorizada. Embora ainda não se falasse em **metadados**, os aspectos que Lacerda (1993) descreve como *expressão*¹⁸, na contemporaneidade podem ser definidos como metadados técnicos: a angulação, o enquadramento, a luminosidade, o tempo de exposição, entre outros. Nas imagens digitais é comum que alguns parâmetros¹⁹ da captação da imagem fiquem anexos aos documentos digitais automaticamente através dos metadados da câmera, como, por exemplo, o EXIF²⁰. Do mesmo modo, existem outros formatos de metadados com dados editáveis, que podem ser acrescentados manualmente e programas²¹ para visualizar, acrescentar e modificar os metadados de uma imagem digital.

Segundo Lancaster (2004), a indexação da imagem, visa representar os assuntos da imagem através da palavra, nos processos documentários de alto nível, a chamada de indexação

¹⁸ É comum que em fotografias analógicas antigas e/ou que fazem parte de acervos de centros de memória/documentação, essas “expressões” estejam escritas manualmente atrás ou num canto da imagem.

¹⁹ Informações dos metadados: Modelo da câmera, Marca do fabricante, Data e horário, Tempo de exposição, Sensibilidade ISO, Tempo de exposição, Comprimento focal, Abertura do diafragma, Tipos de Lentes, Abertura do diafragma, Controle de ganho, Saturação, Balanço de branco, Taxa de Zoom, Dimensões (Largura x altura), Resolução (DPI), orientação da câmera (rotação), Palavras-chaves (tags), entre outros.

²⁰ O *Exchangeable image file format* (Exif) é uma especificação seguida por fabricantes de câmeras digitais que gravam informações na forma de metadados etiquetados. As informações são obtidas automaticamente a partir da câmera fotográfica e não são editáveis.

²¹ Adobe bridge, Adobe photoshop, Adobe lightroom e visualizadores de fotos (Windows/MACOS), entre outros.

baseada em conceitos (alto nível). Além de palavras, as imagens digitais podem ser indexadas e recuperadas de forma automática por atributos intrínsecos, como cor, forma e textura (baixo nível), informação no nível da informação eletrônica, que apenas o computador consegue processar.

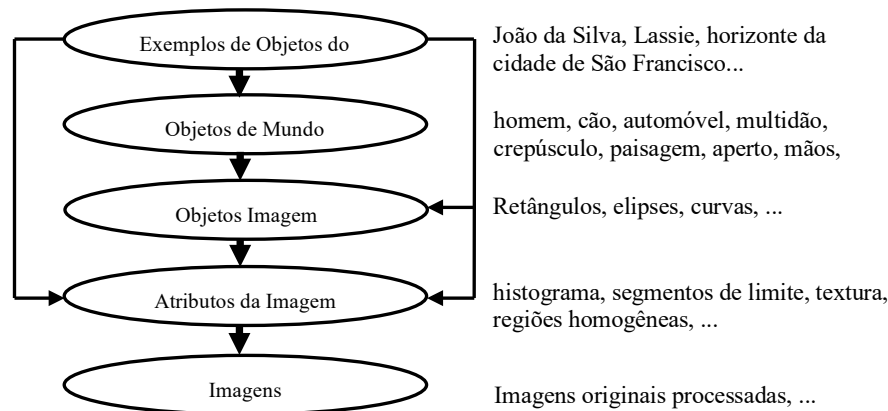
Um ponto muito relevante sobre a análise, síntese e representação temática da imagem em movimento é a relação entre a palavra e a imagem nos processos descritivos. Além de palavras, as imagens digitais podem ser indexadas (automaticamente) e recuperadas por atributos intrínsecos, como cor, forma e textura. Os dois métodos de descrição de imagens:

- a) com palavras, feita por seres humanos, denomina-se em geral indexação baseada em conceitos, b) a indexação de imagens por seus atributos intrínsecos é baseada em conteúdos (Rasmussen, 1997 *apud* Lancaster). As características como cor, forma e textura são amiúde denominadas características de nível baixo. As características de nível alto são descrições da imagem baseadas em palavras (Lancaster, 2004, p. 214).

As abordagens baseadas no conteúdo incluem diversos sistemas que permitem a busca de imagens por meio de características de nível baixo, como forma, cor e textura. “Na maioria dos casos, o computador ²² (possivelmente com ajuda humana) extrai das imagens características úteis de nível baixo e recodifica esses dados numa forma simbólica, mais fácil de ser usada em operações posteriores de indexação e recuperação” (Lancaster, 2004, p. 219). O autor cita que ao tratar de determinada abordagem, uma base de dados conhecida como MUSEUM, Mehrotra (1997), estabeleceu características como níveis variáveis de abstração. Na figura 1, deve-se ler os níveis variáveis de abstração de baixo para cima:

²² Recentemente, os avanços da inteligência artificial possibilitaram a popularização de um terceiro tipo de indexação, a visão computacional, que permite a indexação automática de alto nível, ou seja, a classificação e categorização baseada em conceitos, através da identificação, categorização e segmentação da imagem, realizada pelo computador.

Figura 1 - Principais níveis de abstração na base de dados de um museu de arte



Fonte: Lancaster (2004 *apud* Mehrota, 1991).

- nível 1 (imagens): estão imagens de bases de dados ou imagens-exemplo;
- nível 2 (atributos da imagem): “uma imagem é caracterizada em termos de suas propriedades, como cores de último plano/primeiro plano, cores dominantes, histogramas e propriedades de textura” (Lancaster, 2004, p. 61);
- nível 3 (objetos da imagem): “descrição de imagens em termos de objetos – regiões da imagem, segmentos de limite e contornos - e relações entre eles forma o nível seguinte de abstração” (Lancaster, 2004, p. 61);
- nível 4 (objetos do mundo genérico): “as imagens são descritas em termos de objetos genéricos, relações e conceitos, como homem, cão, carro, multidão, horizonte, crepúsculo, nublado, colorido e sorriso” (Lancaster, 2004, p. 61);
- nível 5 (exemplos de objeto do mundo): “as imagens são descritas em termos de casos específicos de objetos do mundo genérico. Por exemplo, um homem pode ser descrito como João da Silva, uma cadela pode ser descrita como Lassie, uma imagem pode ser descrita como o horizonte da cidade de São Francisco” (Lancaster, 2004, p. 61).

As “descrições de imagens em qualquer um desses níveis de abstração pode ser multinível e ser derivadas das descrições nos níveis inferiores de abstração, ou associadas a elas” (Lancaster, 2004, p. 61). Logo, conclui que é difícil chegar a um acordo sobre a indexação de imagens porque é difícil haver concordância quanto no que uma imagem realmente mostra (Lancaster, 2004). Shatford (1986) faz uma distinção entre “de que é” uma imagem (coisas concretas) e “do que” ela trata (abstrações). Krause (1988 *apud* Lancaster, 2004, p. 218)

“concorda com a distinção entre de e do que trata, mas adota nomes diferentes, a saber, aspectos rígidos (hard) e flexíveis (soft) da imagem”. Orbach (1990 *apud* Lancaster, 2004) apontou ser necessário indexar uma coleção de imagens do ponto de vista de determinado grupo de usuários.

Para certas exigências, como, por exemplo, recuperação de uma imagem que ilustre uma emoção, a indexação de bases de dados de imagens tem algo em comum com a indexação de obras de ficção como vimos no capítulo precedente (Lancaster, 2004, p. 218).

Ademais, vários autores (Markey, 1984; Shatford, 1986; Svenonius, 1994; Van Der Starre, 1995; Enser, 1995 *apud* Lancaster, 2004), entre outros, referem-se ao trabalho do historiador da arte Panofsky, que propôs a análise da imagem do ponto de vista pré-iconográfico, iconográfico e iconológico. Nesse sentido, Enser (1994 *apud* Lancaster, 2004) fez uma experiência de que participaram 18 pessoas, de antecedentes variados e os resultados demonstraram que uma mesma imagem seria indexada em todos os três níveis. Ele utilizou uma cena da torre Eiffel, que recebeu os termos nos níveis pré-iconográfico (torre, rio, árvore), iconográfico (torre Eiffel, rio Sena) e iconológico (romantismo, férias, emoção).

3.1 ANÁLISE DOCUMENTÁRIA E RECUPERAÇÃO DA INFORMAÇÃO DA IMAGEM EM MOVIMENTO

A análise documentária e a representação do conteúdo da imagem em movimento, compreendem a identificação dos conceitos e das relações entre conceitos, relacionados com imagens visuais oriundas da ilusão de movimento quando projetadas ou reproduzidas. Visto que, o conceito de **imagem em movimento** tem relação direta com a sequência de imagens reproduzidas em sucessão rápida e o material sempre será disposto em sequências de imagens. Assim, diferente de uma imagem estática, nesse tipo de material o conteúdo será apresentado numa linha do tempo e em constante alternância de imagens. A imagem em movimento cinematográfica trata de uma série de imagens em sequência, sustentando a polissemia e a necessidade de relativizar a leitura documentária.

Assim, a experiência de aplicar a representação temática num contexto de indexação de imagens em movimento de uma biblioteca universitária mostrou que as categorias QUEM, QUANDO, ONDE, O QUE puderam ser mantidas e alinhadas com as determinações De Genérico e específico, entretanto nem todas as categorias ou subdivisões apresentam essa divisão genérico e específico (Macambyra, 2020). Dessa forma, um quadro foi desenvolvido

pela autora para indexação de imagem em movimento com as categorias Locais de ação (reais ou imaginários), Épocas de ação, Personagens (reais ou imaginários), assuntos tópicos.

O desenvolvimento de SOC para imagens em movimento envolve transcodificar dois tipos de signos diferentes, a imagem e o texto. O que traz especificidades a documentação textual, pois “a documentação audiovisual traz consigo especificidades inerentes ao seu processo de criação, suporte e registro que devem ser consideradas quando se discutem as possibilidades de organização e de recuperação da informação aí presente” (Lima, 2016, p. 86).

Representar o tema de um documento, como as imagens em movimento, envolvem também o que a imagem e o som nos sugerem, não só o que é visto ou ouvido, pois o conteúdo da imagem pode ter sido criado com um objetivo e utilizado para outro. Como, por exemplo, uma sequência que mostre o interior de um laboratório farmacêutico, pode ter diferentes denotações e conotações, como doença, medicamentos ou uma pesquisa científica (Lima, 2016).

A representação temática, que se refere ao conteúdo documento, envolve dois processos, a classificação, onde um código, que nos permite localizar o documento, dentro de um conjunto de documentos armazenados em um local físico, será atribuído ao item e a indexação, onde descritores de assuntos serão atribuídos ao documento para que seja possível a sua recuperação em conjunto a outros documentos indexados sob o mesmo tema, ainda que armazenados física ou virtualmente, em locais diferentes (Lima, 2016, p. 89).

A conversão de imagens em movimento para texto, conhecida como transcodificação, envolve a tradução de um código para outro, enfrentando os desafios inerentes a toda tradução: perda de precisão, seleção de informação, possibilidade de erro, visto que, a imagem em movimento contém uma grande variedade de informações físicas e semânticas distintas (Lima, 2019). Devido, a esses fatos, a autora enumera os seguintes problemas para análise da imagem:

- Transparência
- Transcodificação
- Interpretação
- Descrição
- Falta de uma terminologia comum

As imagens podem ser analisadas como manifestações comunicativas, ou seja, pode transmitir informação por mensagens. Ademais, imagens são de “extraordinário valor para o documentalista, que tenta apreendê-las para decodificar analiticamente seu significado para que sua posterior comunicação secundária seja possível” (Agustín-Lacruz, 2010, p. 86, tradução nossa).

Assim, González e Arillo (2003, p. 122) enfatizam que a imagem e a palavra pertencem a sistemas não homogêneos em relação à natureza linguística e semiótica, sendo a imagem um signo icônico e as legendas que acompanham as imagens são signos linguísticos. Os autores afirmam que “não chegaremos a sua total compreensão, se consultarmos os elementos separadamente. Por isso, o registro de identificação da fotografia [...] está obrigado a guardar os textos que lhe dão sentido, permitindo compreender seu significado”.

As discordâncias na definição terminológica também se apresentam como um desafio para chegar a uma solução simples (McCarthy; Targino, 1984, p. 304 *apud* Bethônico, 2006, p. 58).

Não existe consenso nem no tipo de material envolvido e nem mesmo na terminologia. Os termos mídia (ou média) audiovisual, materiais audiovisuais, meios audiovisuais, materiais especiais, materiais não impressos, materiais não-bibliográficos, multimeios, AVM, AV, nonbook media e meios não-gráficos vêm insistentemente sendo utilizados como sinônimos (Bethônico, 2006, p. 58).

Ainda, para identificar os objetos, durante a análise documental e a recuperação da informação audiovisual, as seguintes categorias do quadro 4 (Lima, 2019)

Quadro 4 - Identificação de objetos na análise documental

Quem?	Identificar todos os protagonistas (nome, idade, profissão, função) e/ou objetos (seres vivos, artefatos, construções). Em que situação ou que objetos estão representados? Infraestrutura, eventos sociais e ou políticos
Onde?	Que lugar representa: praça, teatro, campo, praia.
Quando?	Data, estação do ano, época
Como?	Descrever as ações das pessoas ou atitudes em relação aos objetos: máquinas trabalhando, político discursando.

Fonte: Elaborado pela autora.

No registro da fotografia cinematográfica por um dispositivo ótico é possível captar os seres/objetos, o espaço, o tempo e a ilusão do movimento dos referentes que estiveram presentes na frente da câmera. Assim como o índice, o registro e a projeção possuem conexão física entre objeto e sua representação, característica que pertence à gênese da formação da imagem em movimento cinematográfica. No caso da imagem em movimento, a natureza sequencial das imagens participa do conceito de plano, pois um conjunto sequencial de imagens (fotogramas) entre dois cortes estabelecem um plano. O plano determina a escala e quais partes dos seres/objetos irão aparecer no quadro, ainda define o tema numa cena e ajuda a especificar as emoções, sendo a forma e o conteúdo, indissociáveis na linguagem cinematográfica. O modo

de mostraç o (*showing* em ingl s)   um n vel mostrativo que ocorre no plano isolado, ao passo que o n vel narrativo est  na montagem dos planos entre si, sendo que a narra o na sua totalidade s  acontece no percurso de uma leitura cont nua, anulando a autonomia do plano, todavia a narra o e a mostra o s o n veis potenciais de narrativa, o primeiro na imagem  nica e o segundo na sequ ncia de imagens (Aumont, 1993, p. 245 *apud* Gaudreault, 1988).

Assim, existe uma complexidade na perspectiva de indexa o de imagens em movimento, como os termos para descrever uma s rie de a o es e eventos. Partindo dos quadros de Smit (1996) para representa o do cont do de fotografias,   poss vel aplicar a determina o “do que se trata” gen rico (Shatford, 1986) e o n vel pr -iconogr fico (Panofsky, 1979) para indexa o dos planos, cenas e sequ ncias. De modo que, se a categoria “o que” para a o es, eventos e emo o es (Shatford, 1986; Blery, 1989; Smit, 1996) fosse usada para indexar imagens em movimento que tratassem de um jogo de futebol, por exemplo, existiria uma diversidade de planos na composi o das cenas e sequ ncias. A fim de,   poss vel adaptar as categorias gen ricas e espec ficas e os n veis pr -iconogr fico e iconogr fico, usando os principais planos da linguagem cinematogr fica para indexar as imagens, como exemplifica o quadro 5 a seguir:

Quadro 5 - Representa o da imagem em movimento e planos cinematogr ficos

SHATFORD	Exemplo	Close/Detalhe	Plano m�dio (PM)	Plano geral (PG)	Grande plano geral (GPG)
Gen�rico (Semelhan�a e conex�o f�sica com o referente real)	Jogo de futebol	Bola, P�, Rosto		Est�dio, Pessoa, Time de futebol	Panor�mica do Espa�o (est�dio)
Espec�fico (Identifica o assunto convencional)	Copa do mundo	Pel�		Maracan�, Pel�, Sele�o brasileira	Panor�mica do Maracan�

Fonte: Baseado em Smit (1996).

Em tais termos, o uso da linguagem cinematogr fica, junto a uma no o dos principais planos utilizados nas imagens em movimento cinematogr ficas, pode auxiliar na identifica o de conceitos pertencentes a imagem em diferentes n veis de sequ ncias e de enquadramento. No Quadro 1 o Grande plano geral (GPG) registraria uma imagem panor mica do est dio, mas n o identificaria claramente as pessoas e objetos na imagem, contudo, o Plano Geral (PG) seria um plano que identificaria todos os principais elementos da cena: jogadores, torcida e objetos. O Plano M dio (PM) indica a rela o entre partes dos seres/objetos principais em cena e os seres/objetos que aparecem de fundo ou em composi o secund ria, tamb m exalta as mudan as na profundidade de campo da imagem e na mudan a de foco dos seres/objetos. J  o Close e o “Plano Detalhe” operam no n vel da identifica o e do detalhamento da cena. Dessa forma,   poss vel selecionar e delimitar conceitos mais importantes em cada plano e no conjunto

da cena. A imagem em movimento pode ser indexada não apenas pelas cenas e planos, mas também pelas sequências, fotogramas e pelo conteúdo integral do material, sendo que todas essas possibilidades têm com premissa reproduzirem imagens baseadas no tempo e em sequência. Todavia, diferente indexar o item por completo do que planos e sequências, pois estas apresentam maior complexidade, principalmente trechos que mostram ações: alguém ou algo fazendo alguma coisa, pois isso dificulta constituir paridades entre os termos e as ações presentes nos diversos segmentos do filme (Macambyra, 2013).

Nesse sentido, um estudo de usuário desenvolvido num contexto de biblioteca universitária, identificou termos não convencionais específicos para indexação de filmes, sendo um tipo de imagem em movimento, que em comparação ao vocabulário controlado das principais áreas do conhecimento da instituição, nem todos os termos apresentaram caráter acadêmico e científico, descritos no Quadro 6 (Macambyra, 2017):

Quadro 6 - Termos não convencionais para indexação de filmes

Termos não convencionais específicos	Exemplos
Termos relacionados à linguagem, técnica ou narrativa cinematográfica	Não linear, Multi-plot, Plot-twist, Final aberto, Cenários gerados por computador, Bullet-time, Happy-end
Temas clássicos e recorrentes em obras ficcionais	Universo paralelo, Sociedade em crise, Realidade alternativa
Termos descritivos muito específicos	Moça de vestido vermelho, Olhos nas mãos
Termos de conotação moral	Decência, Imbecilidade, Arrogância, Cinismo
Termos não relacionados ao conteúdo	Maior bilheteria nacional, Diretores famosos

Fonte: Baseado em Macambyra (2017).

Uma questão que estabelece desafios é definir os assuntos principais de imagens em movimento cinematográficas. Principalmente documentos que são obras artísticas, que apresentam uma diversidade de símbolos culturais, diferentes tipos de linguagem visual e signos imagéticos. Que possuem significações com maior multiplicidade de sentidos, como o caso dos filmes de ficção e documentário.

Existe uma dificuldade de indexar e criar instrumentos para imagens em movimento cinematográficas como os filmes de **ficção**. Visto que, muitas vezes, os assuntos podem não fazer parte da hierarquia das grandes áreas do conhecimento de vocabulário controlados desenvolvidos para contextos textuais. Principalmente por conterem termos que denotam conceitos com significados fantasioso, ficcional, irreal ou inexistente. Esse problema em criar vocabulários controlados para ficção, não é exclusivamente cinematográfico ou das artes visuais, mas também da literatura, que apresenta a distinção entre ficção e não-ficção entre os documentos. Igualmente, a análise documentária de filmes **documentários** também apresenta desafios na indexação e na representação de assuntos, pelo motivo de ser um gênero que

tangencia a realidade. Dessa forma, as análises, incluindo a documentária, costumam evidenciar o conteúdo através dos ícones, símbolos e códigos pertinentes a diversas áreas, como exemplo, a ciências sociais, história, antropologia, educação, artes, entre outros; mantendo um grande enfoque no conteúdo discursivo do filme em detrimento dos códigos cinematográficos.

A classificação das obras cinematográficas como expressões artísticas, permite a definição de uma metodologia que identifique o gênero da obra em ficção ou documentário, a partir de uma análise que considere o conteúdo e a forma do documento. Através das características cinematográficas intrínsecas de cada gênero e subgênero, os quais são resultados das escolhas do roteiro, planos, enquadramentos, fotografia cinematográfica (iluminação, exposição, contraste, brilho, foco, balanço de cor, calorimetria), movimentos de câmera, *raccords*, áudio, música, montagem, roteiro, tipos de personagem; todos são elementos que formatam a história de modo audiovisual e definem a expressão cinematográfica.

Assim, a ficção cinematográfica apresenta a estrutura do roteiro e escrita de diálogos, a montagem narrativa clássica (ilusão de continuidade), o uso do *raccord* (continuidade temporária ou espacial dos planos), eixo cinematográfico, planos e enquadramentos, locação, estúdio e ambiente sonoro controlados, música incidental, performance de atores cênicos. Além de estruturas narrativas (jornada do herói) e tipologias de gênero e subgêneros com temáticas específicas. Diferentemente, o gênero documentário é uma obra artística que trata elementos da realidade na narrativa ou ainda um gênero que não permite a total roteirização dos eventos tratados no desenvolvimento do filme. Convencionalmente os personagens e ambientes são reais, os diálogos e o som ambiente são diretamente captados na filmagem. E diversamente da ficção que segue um roteiro pré-estabelecido, no documentário a estruturação da narrativa acontece na montagem e no tratamento final, porque não é possível prever tudo ocorrerá no momento da filmagem.

De maneira que, estas características cinematográficas dos gêneros ficção e documentário, fazem parte do conceito de diegese, a qual é a dimensão espaço-temporal da narrativa, é o universo fictício da história tanto para narrativas ficcionais e documentárias. Metz (1972) afirma a importância da diegese na semiologia do filme.

A noção de diegese é tão importante para a filmo-semiologia como a idéia de arte. A palavra provém do grego (*ainynas*) significando "narração" e designava particularmente uma das partes obrigatórias do discurso judiciário, a exposição dos fatos [...]isto é em suma, o conjunto da denotação fílmica: o enredo em si, mas também o tempo e o espaço implicados no e pelo credo, portanto, as personagens, paisagens, acontecimentos e outros elementos narrativos, desde que tomados no seu estado denotado. Como o cinema

significa as sucessões, precessões, hiatos temporais, causalidades, relações adversativas, consequência, proximidade ou afastamento espacial etc.: tantas questões centrais para a semiologia do cinema (Metz, 1972, p. 118).

Além disso, Aumont (1993) discorreu como a representação do espaço e do tempo diegéticos transformam a diegese ou um fragmento de diegese, em imagem, sua construção é determinada principalmente por suas convenções, por códigos e pelos símbolos sociais. Para o autor a diegese é “uma construção imaginária, um mundo fictício que tem leis próprias mais ou menos parecidas com as leis do mundo natural, ou pelo menos com a concepção, variável, que dele se tem” (Aumont, 1993, p. 248). Dessa forma, na análise e identificação dos personagens, espaços e tempos, o estado denotativo da diegese afeta diretamente o estado conotativo das convenções, códigos e símbolos sociais, pois as condições de ícone da imagem são intrínsecas a leitura simbólica do filme (Aumont, 1993).

Assim, ao identificar no documento os principais conceitos a partir das categorias QUEM, QUANDO, ONDE, O QUE, a análise documentária representa os assuntos do conteúdo do filme que são inerentes a diegese, de tal modo, se estabelece uma linguagem particularizada no domínio temático do documento cinematográfico. Em vista disso, o conjunto de termos de indexação tem origem na diegese da narrativa cinematográfica. Dessa forma, a indexação de termos que remetem a lugares reais ou imaginários, personagens reais ou imaginários, temas fantasiosos/ficcionais ou reais, provém de conceituações que fazem parte do domínio de expressões artísticas e manifestações culturais, visto que participam do universo diegético da história. Dessa forma, ao invés de qualificar os termos a partir da divisão ficção e documentário, o vocabulário para imagens em movimento cinematográficas pode ser sistematizado num SOC com categorias para os termos que denotam conceitos de categorias diegéticas, como personagem, espaço e tempo diegéticos. A diegese como aspecto fundamental do domínio da linguagem de obras cinematográficas pode ser usada em hierarquias gerais das grandes áreas do conhecimento, como a categoria Artes, como mostra o quadro 7:

Quadro 7 - Organização de imagem em movimento com categorias diegéticas

Lugares geográficos diegéticos	Eventos diegeticos	Ambientes diegéticos
Brasil – Diegético São Paulo - Diegético	Ditadura Militar – Diegético Independência do Brasil – Diegético	Universo paralelo Realidade alternativa Mundo encantado

Fonte: Elaborado pela autora.

Assim como, é possível incluir a diegese como parte da linguagem especializada de instrumentos terminológicos específicos para imagens em movimentos, como microtesauros de assuntos dos conteúdos das imagens em movimento cinematográficas.

Quadro 8 - Indexação de filmes a partir de categorias diegéticas

Categorias Diegéticas		Genéricos	Específicos
Locais da ação	Exemplo	Cidade	São Paulo
Personagens	Exemplo	Personalidade Histórica	D.Pedro

Fonte: Baseado em Macambyra (2020).

O Quadro 8, não categoriza os termos como reais ou imaginários, mas parte da estrutura e da temática intrínseca ao tipo de expressão artística que é refletida no suporte e na intenção do documento pela diegese. Dessa forma, é importante desenvolver pesquisas de SOC para imagem em movimento na web.

3.2 SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO DA IMAGEM EM MOVIMENTO NA WEB

A publicação de bases de conhecimento de instituições do patrimônio cultural no ambiente web, permite a interoperabilidade de dados entre diferentes instituições e a diminuição de silos de informações nas bases de dados e na recuperação da informação. Assim, embora procedimentos de interoperabilidade estão sendo desenvolvidos para sistemas de documentação de bibliotecas e museus, a maioria dos arquivos e cinematecas com documentos audiovisuais, possuem silos de dados. E “apesar de projetos interessantes de protótipos e modelos, permanece o desafio de entender as necessidades específicas da comunidade arquivística cinematográfica e construir sistemas sustentáveis e úteis em torno desses conceitos” (Heftberger; Duchesne, 2021, p. 47, tradução nossa).

De tal modo, a digitalização de acervos culturais e a publicação na web, produziram um grande volume de informações oriundas de bibliotecas, arquivos e museus. Criando um interesse acadêmico e cultural perante esses dados. No âmbito dos arquivos com documentos audiovisuais, como fotografia e a imagem em movimento, algumas iniciativas de representação da informação e do conhecimento no ambiente web, giram em torno de solucionar desafios. Como promover o acesso à informação, a interoperabilidade de dados entre diferentes instituições e problemas de representação temática, como a indexação automática e manual.

Como tratado na seção anterior, existem necessidades específicas na comunidade de acervos de imagem em movimento. Esses processos envolvem a representação temática, como

a ausência de tratamento documental (Domínguez-Delgado; López-Hernández, 2017b) e a abundância de informações não representadas (Gracy, 2018). Ao mesmo tempo, enfrentam o desafio da polissemia e os aspectos de denotação e conotação da imagem (Lima, 2016), bem como o problema de estruturar SOC para documentos audiovisuais (Macambyra, 2020). E, por fim, é impossível adaptar automaticamente os processos da análise documental de documentos textuais para imagens (Smit, 1996). Em contrapartida, existem inúmeros avanços no desenvolvimento de manuais para catalogação e padrões de metadados descritivos para arquivos de documentos audiovisuais, como o manual de catalogação da FIAF e o padrão de metadado CEN. Por conseguinte, a representação temática dos assuntos do conteúdo dos filmes necessita de classificação e indexação. Igualmente, na seleção dos conceitos para recuperação, é fundamental utilizar descritores de um vocabulário controlado.

A imagem em movimento pode ser considerada um dado não estruturado a partir do ponto de vista da **ciência de dados** e da **ciência da computação**. Visto que, embora as imagens em movimento contenham uma abundância de informações visuais, essas informações não estão organizadas em um formato tabular ou estruturado como em um banco de dados, planilha ou tabela. Dados estruturados organizam informações em tabelas com colunas e cabeçalhos, facilitando consultas, análises estatísticas e uso de algoritmos. Imagens em movimento, ao contrário, são sequências contínuas de píxeis que não possuem estrutura tabular. Portanto, para analisar esse tipo de dados, são necessárias técnicas de processamento de imagem, vídeo, visão computacional e aprendizado de máquina. Assim, do ponto de vista da computação, imagens em movimento são consideradas não estruturadas devido à sua natureza contínua e complexa. São considerados dados não estruturados recursos textuais, como e-mails, documento de texto, conteúdo de rede social, além desses, também inclui recursos não textuais, como imagem digital, áudio e vídeo. Dessa forma, é possível dividir os dados dessas instituições em estruturados, semiestruturados e não estruturados.

Em vista disso, os dados provenientes de bibliotecas, arquivos e museus, podem ser categorizados em três grupos principais com base em seu tipo (Zeng, 2019, p. 3, tradução nossa) como mostra o quadro 9:

Quadro 9 - Dados provenientes de bibliotecas, arquivos e museus

Dados estruturados	Dados semiestruturado	Dados não estruturados
Os dados estruturados em bibliotecas, arquivos e museus, incluem bibliografias, bancos de dados de indexação e abstração, índices de citação, catálogos de todos os tipos, portais de coleções especiais, repositórios de metadados, conjuntos de dados de pesquisa com curadoria e autoridades de nomes. Os dados estruturados são normalmente mantidos em bancos de dados em que todos os pares de chave/valor têm identificadores e relações claras e seguem um modelo de dados explícito (Schöch, 2013 <i>apud</i> Zeng, 2019, p.3, tradução nossa).	Os dados semiestruturados em bibliotecas, arquivos e museus, compreendem as seções não estruturadas dentro das descrições de metadados (por exemplo, notas em registros bibliográficos, as descrições de conteúdo contidas em recursos de busca de arquivos, índice analítico e resumos de relatórios em repositórios digitais).	Os dados não estruturados são caracterizados como "todo o resto". Nas bibliotecas, arquivos e museus, eles podem ser encontrados em documentos e outros objetos portadores de informação (textuais ou não textuais, digitalizados ou não digitalizados), em todos os tipos de formatos. Em recursos de dados atendidos por bibliotecas, arquivos e museus, os dados não estruturados geralmente estão disponíveis na maior quantidade em comparação com os dados estruturados, têm a maior diversidade em tipo, natureza e qualidade e são os mais desafiadores de processar.

Fonte: Zeng (2019, p. 3, tradução nossa).

Os problemas de organização do conhecimento e representação da informação trazidos pelos dados não estruturados, impossibilita o acesso às informações presentes nos diferentes tipos de conteúdo. As recomendações da Web Semântica auxiliam com soluções para o tratamento da imagem e do vídeo, pois, para recursos não baseados em texto, a anotação semântica e as bases de conhecimento possuem grande potencial para se aprofundar nesses tipos de dados (Zeng, 2019).

Nesse sentido, nos últimos anos, ocorreu um significativo interesse em organizar e representar informação documental da imagem em movimento na web. A partir da estruturação de dados *linked data* (triplas RDF) e linguagem de ontologia para web (OWL). Visando aperfeiçoar a interoperabilidade de informação e o enriquecimento semântico de dados. De tal forma, cresceu o número de pesquisas e projetos-pilotos de instituições, bibliotecas, arquivos e museus com acervos de imagem em movimento no domínio da Web Semântica e das humanidades digitais.

As humanidades digitais reforçam o encontro entre tecnologia da informação, dados e objetos de pesquisa, para explorar problemas da área de humanidades com novas abordagens e metodologias, de maneira interdisciplinar com a ciência da computação. Desse modo, no

contexto das instituições, os projetos e pesquisas no domínio da imagem em movimento e as humanidades digitais, centralizam seus processos no uso de ferramentas (frameworks e esquemas de dados), players de exibição (interfaces com telas para visualização online do documento primário). Bem como, plataformas na web (bases de conhecimento e/ou ambiente colaborativo) e aplicações (programa ou grupo de programas projetado para usuários finais) digitais. Para estabelecer interação baseadas nas demandas dos usuários, principalmente pesquisadores de informação presente em acervos e arquivos.

Nas tecnologias da Web Semântica e dos dados abertos e conectados, existem o modelo de dados SKOS para SOC, incluindo esquemas de classificação e categorização, cabeçalhos de assuntos e arquivos de autoridade, vocabulários estruturados, como tesouros, e esquemas como redes semânticas e ontologias.

O SKOS pode ser definido como um vocabulário de valor, que coleta, enriquece e publica em RDF, os dados do domínio. Estruturando uma lista controlada de valores permitidos para registros de metadados. Entretanto, vocabulários SOC, não são usados apenas em descrições de metadados, mas também para mapear e reusar conceitos entre diferentes esquemas (estrutura taxonômica). A representação temática e os SOC no contexto LOD, incluem o uso de vocabulários de valor para determinar conceitos relevantes para descrever metadados de assuntos. Podendo ser metadados no nível da imagem, ou seja, que descreve a imagem inteira ou metadados no nível da anotação, que descreve pontos específicos da imagem.

No contexto da imagem em movimento, as iniciativas em torno da Web Semântica estão no início, mas já apresentam grupos de estudos e workshops organizados pelas principais instituições e autoridades da área. Como a FIAF, que visa mapear o modelo conceitual do manual de catalogação em ontologias para representar metadados descritivos. Visto que, os metadados descritivos provenientes de catalogação, podem ser coletados em diversos formatos, normalizadas por meio de uma estrutura LOD de processamento e mesclagem de dados. Para transformar declarações sobre as informações documentárias do acervo em dados abertos conectados. A partir da construção de ontologias para mapear o domínio dos dados dos arquivos. De forma que, tais ações permitem realizar consultas nos conjuntos de dados com SPARQL, que apresentam resultados, como, por exemplo, a comparação do número de acervos por instituição, por formato, por obra e por suporte de filmes, entre outros. Exemplos de experiências estão presentes no *Linked Open Data Task Force* da FIAF CDC (Comissão de Catálogo e Documentação), que estão construindo caminhos para que essas abordagens sejam aplicadas nos arquivos de filmes. De maneira que, no projeto Ontologia de Murnau (*InterMurnau*), foi desenvolvida uma ontologia a partir da exportação dos bancos de dados de

um grupo de instituições colaboradoras. A fim de que foram coletados nos catálogos especificamente dados relacionados ao cineasta Murnau, por existir uma ampla gama de instituições com variações e edições de suas obras.

Essa experiência teve como foco a catalogação, que abrange as características físicas do documento audiovisual e embora não tenham sido mapeados diretamente como modelos, a mesclagem das entidades coletadas na base de dados dos arquivos, foi baseada no modelo de catalogação FIAF/FRBR e no padrão de metadados da CEN 15907.

O projeto ADA é um projeto-piloto desenvolvido pelo grupo de pesquisa “*Audio-visual rhetorics of affect*” da *Freie Universität Berlin* e do *Hasso-Plattner-Institut*, para auxiliar o acesso e a pesquisa de imagem em movimento. O projeto tem dois objetivos principais, a criação de um vocabulário de anotação padronizado para ser aplicado em anotações semânticas e a classificação semiautomática de padrões audiovisuais por meio de modelos de treinamento em dados de anotação. Nesse sentido, o vocabulário de anotação para estudos de filmes e anotações semânticas de material audiovisual é baseado nos princípios do LOD.

O OASIS: *Open Annotation Semantic Indexing System* é um sistema para anotação semântica de imagens pictóricas, que inclui as disciplinas organização do conhecimento e indexação da imagem. E trabalha o conceito de anotação profunda da imagem, com metadados no nível da anotação, que descreveria pontos específicos da imagem através da anotação de subimagem.

O projeto I-Media-Cities²³ (IMC) é uma plataforma de anotação automática de conteúdo multimídia, uma iniciativa de instituições europeias que possuem obras audiovisuais ficcionais e não ficcionais relacionado as cidades da Europa. O modelo de metadados descritivos é proveniente da documentação dos arquivos com base na norma CEN EN19507 e a representação temática ocorre através do enriquecimento automático e manual de dados, que são normalizados através de um vocabulário controlado.

3.2.1 Linked Open Data na FIAF

A *International Federation of Film Archives* (FIAF), a Federação Internacional de Arquivos de Filmes, tem se dedicado à preservação e acesso ao patrimônio cinematográfico mundial desde 1938. Ela reúne as principais instituições sem fins lucrativos do mundo neste campo. O Manual de catalogação de imagens em movimento da FIAF 2016 inclui todo o ciclo

²³ <http://imediacities.eu>

de vida do conteúdo (proveniência, preservação, acesso), dessa forma os metadados resultantes atendem as necessidades de todos os tipos de usuários finais.

No ano de 2015, a Comissão de Catalogação e Documentação da FIAF iniciou uma investigação²⁴ de como é possível usar o LOD em arquivos de filmes, quais seriam os problemas e vantagens de adotar as recomendações da Web Semântica para dados abertos e conectados.

Dessa forma, a FIAF estabeleceu um grupo de estudo para pesquisar a extensão do manual de catalogação de imagens em movimento e dos glossários da FIAF, empregando dados abertos conectados. Entre os objetivos estão estabelecer identificadores estáveis e persistentes (por exemplo, filmografias nacionais, Wikidata, EIDR-*Universal unique identifier for movie and television assets*, ISAN-*International Standard Audiovisual Number*), para os recursos de cada modelos. Ao mesmo tempo, uma perspectiva do grupo de pesquisa, é usar fontes de dados de pesquisa e descobrir quais recursos LOD são relevantes para a comunidade de filiados da FIAF, ademais, quais não estão disponíveis para reuso.

Heftberger e Duchesne (2020), fizeram uma pesquisa aplicada sobre as práticas de catalogação na era dos dados abertos conectados, com o uso da Wikidata e Wikibase para arquivos de filmes. E aplicaram como metodologia a comparação entre a ontologia Wikidata/Wikibase e o metadado CEN 15907. De tal modo, obtiveram como resultados preliminares dois futuros modelos:

Uma instância centralizada da Wikibase (também conhecida como FIAF DATA), onde os dados serão mapeados por FIAF/ou um grupo centralizado. A segunda opção seria criar um repositório de software github que contém uma imagem do Docker da Wikibase pré-configurada, de modo que cada arquivo possa facilmente iniciar sua própria instância da Wikibase com uma ontologia aprovada (FIAF CDC), na forma de propriedades predefinidas, um modelo csv para facilitar a contribuição de dados de arquivo e algum script para ajudar no mapeamento para identificadores compartilhados (por exemplo, Wikidata) (Heftberger; Duchesne, 2020, tradução nossa).

Os pesquisadores concluíram que os arquivos de filmes podem usar o LOD disponível na nuvem de *datasets*²⁵, mas também como os arquivos podem contribuir para os bancos de dados existentes, como Wikidata, EIDR, etc.

²⁴ Extending the FIAF Moving Image Cataloguing Manual and FIAF Glossaries using Linked Open Data. Disponível em: <http://www.ace-film.eu/wp-content/uploads/2017/08/FIAF-CDC-LoD-ACE-Bologna-June-2017.pdf>

²⁵ Nuvem LOD. Disponível em: <https://lod-cloud.net/>

Do mesmo modo, tratam as possibilidades de combinar a análise audiovisual (por exemplo, aprendizado de máquina e visão computacional) com indexação e mapear a reutilização do material digital automaticamente. “Como o LOD pode ajudar a minerar nossos próprios bancos de dados?”, foi um questionamento produzido pelo grupo de estudo.

Em relação à organização do conhecimento foi discutido que os possíveis pontos de partida dos dados abertos conectados, são os registros filmográficos e registros de autoridade. Assim como, a indexação de assuntos, vocabulários controlados e informações sobre as coleções (dados de proveniência).

Do ponto de vista da comissão, espera-se que o LOD ajude na escassez de recursos manuais na catalogação. O reuso de dados e registros de autoridades é uma forma de eliminar trabalho redundante, para isso é possível importar dados e/ou otimizar fontes de dados colaborativamente.

Outro projeto desenvolvido, que foi uma extensão dos trabalhos da força tarefa LOD da FIAF, é a ontologia *Intermurnau*, que apesar de não ser um mapeamento, foi modelada baseada no manual da FIAF, que usa o modelo conceitual FRBR e no padrão de metatados CEN 15907. A ontologia foi realizada a partir da normalização de dados catalográficos de instituições de arquivos cinematográficos que possuem em sua base de dados e documentação, informações provenientes de acervos do cineasta alemão F. W. Murnau²⁶.

Durante a modelagem da ontologia, os autores normalizaram os diferentes métodos de catalogação e mapearam os dados de um grupo de arquivos cinematográficos. Contudo, os pesquisadores encontraram várias dificuldades em relação à compreensão dos atuais esquemas e estruturas de dados de filmes usados nas instituições. Embora “existam vários modelos prescritivos (como o CEN 15907 e o manual de catalogação de imagens em movimento FIAF), a adoção na comunidade é muitas vezes em curso, parcial ou aspiracional” (Heftberger; Duchesne, 2021, p. 49, tradução nossa). Consequentemente, foi coletado amostras de dados de um conjunto de arquivos, a partir disso, foram avaliados os esquemas em uso nas instituições e construído um único Wikibase²⁷ experimental. Unindo todas as contribuições de dados em uma única plataforma.

A obra do realizador alemão Friedrich Wilhelm Murnau foi escolhida como a base para o conjunto de amostras devido à sua condição de realizador com reputação internacional. Além

²⁶ F. W. Murnau foi um dos mais importantes realizadores do cinema mudo, do cinema expressionista alemão e do movimento Kammerspiel (Wikipédia).

²⁷ Wikibase é uma suíte de software de código aberto para criar bases de conhecimento colaborativas, abrindo as portas para a web com o Linked Open Data, permitindo armazenamento, gerenciamento e acesso a dados estruturados. É um serviço da comunidade Wikimedia Foundation.

disso, acervos com seus filmes estão numa ampla gama de instituições e por isso, existem muitas variações e edições de suas obras. Assim, foram convidadas oito instituições para contribuir com exportações de seus bancos de dados brutos com registros de F. W. Murnau. Como os dados²⁸ refletiram diferentes estruturas filmográficas e vocabulários técnicos, foi realizado um processo de normalização utilizando *scripts* computacionais para vincular esses conjuntos de dados individuais a uma ontologia autônoma.

Depois que os conjuntos de dados foram normalizados para um modelo padronizado, as entidades correspondentes foram detectadas e reconciliadas com uso de algoritmos para determinar o nível de semelhança dos nomes das obras e dos agentes. Para que desta forma fossem sobrepostas e trabalhado apenas com os títulos originais. Assim, “o aspecto mais desafiador do processo de correspondência foi reconciliar as camadas ‘variante’ e ‘manifestação’ do modelo, pois ‘obras’ eram uniformes e facilmente identificáveis, e ‘itens’ objetos automaticamente discretos” (Heftberger; Duchesne, 2021, p. 50, tradução nossa).

O resultado desse processo de fusão foi um único arquivo RDF chamado "*Intermurnau*", serializado no formato "Turtle" e pronto para ingestão na instância do Wikibase²⁹. A estrutura de processamento de dados foi a LOD, pois ele simplificou as operações de normalização e reconciliação.

A pesquisa analisa como projetos internacionais estão compartilhando coleções audiovisuais na web e publicando seus acervos online em ambientes digitais com métodos e metodologias que envolvem grandes conjuntos de dados estruturados e desestruturados

3.2.2 Anotação semântica da imagem em movimento

O enriquecimento semântico baseado em anotação semântica automática e manual de imagem em movimento na web, considera os SOC e as humanidades digitais. Um problema nesse cenário, é a alta demanda de recursos para indexação audiovisual nas instituições. A complexidade da análise e representação da imagem em movimento, produzem silos de informação e lacunas na representação do assunto do conteúdo de imagens. E na modelagem de metadados provenientes dos documentos de bibliotecas, arquivos e museus.

²⁸ Esses dados incluíam várias codificações diferentes: Excel, CSV, XML e JSON.

²⁹ A própria instância do Wikibase foi construída usando uma imagem do Docker cedida pela Wikimedia Deutschland, carregada em um servidor hospedado na nuvem.

A pesquisa bibliográfica realizada aponta métodos e procedimentos para melhorar os processos de indexação para a recuperação de imagens em movimento ou a criação de SOC para padronização terminológica. Todavia ainda existem muitas lacunas nos instrumentos das instituições para representar o assunto das imagens. Além disso, transpondo os mesmos problemas para o ambiente digital, as lacunas teóricas-conceituais e metodológicas se tornam ainda mais complexas. Sendo necessário problematizar e teorizar questões sobre organização da informação audiovisual em ambiente web, igualmente, tratar os dados vinculados, a interoperabilidade e o enriquecimento semântico.

Portanto, aproximações para resolver esses problemas envolvem os processos e ferramentas de enriquecimento semântico de dados baseado em anotação de imagens de arquivos, bibliotecas e museus, que auxiliam na representação da informação e na indexação de imagens no ambiente web. Logo, é de interesse realizar investigações em virtude do papel das tecnologias web na organização e recuperação de informação audiovisual. E devido às possibilidades de gerar plataformas e ecossistemas de pesquisa pautados nas tecnologias da informação e nos novos paradigmas da pesquisa acadêmica.

As pesquisas sobre enriquecimento semântico de dados baseado em anotação semântica de imagens em movimento no contexto da Web Semântica, contemplam os SOC, tesouros semânticos, vocabulários e ontologias. Ademais, envolvem os SOC e indexação da imagem, a investigação epistemológica da representação da imagem em movimento, a organização do conhecimento, a modelagem de dados conectados, interoperabilidade e representação de entidades. Além disso, buscam solucionar lacunas referente ao tratamento documental de imagem via processos automatizados na web.

Destarte, as metodologias, os conceitos e as tecnologias no contexto das humanidades digitais precisam ser potencializadas nas instituições brasileiras de conhecimento e memória. Além da criação de ecossistemas digitais que envolvem projetos de patrimônio, cultura e pesquisa acadêmica com dados estruturados, desestruturados e semiestruturados na web.

A correspondência de texto, principalmente de palavras, foi a principal forma de processamento e recuperação de informações “apenas recentemente, a pesquisa baseada em conteúdo (de acordo com significados semânticos) foi implementada e usada em grande escala, conforme simbolizado por *Knowledge Graphs*” (Zeng, 2019, p. 20, tradução nossa).

A mesma autora, afirma que tecnologias recentes estão sendo aplicadas para aprimorar a exploração e descoberta de recursos baseados em texto, “em vez de depender da pesquisa de texto completo, métodos como análise baseada em semântica, extração, mineração e marcação são usados para melhorar a descoberta de informações de dados não estruturados”. Para

recursos como imagem fixa e movimento, a “anotação semântica e as bases de conhecimento fundamentadas em ontologia, revelaram um grande potencial para se aprofundar em dados e apoiar a pesquisa em ciências humanas digitais” (Zeng, 2019, p. 20, tradução nossa).

O dicionário define anotação, do latim *lat annotatĭo*, como: 1. Ato ou efeito de anotar; 2. Registro escrito conciso; apontamento, nota; 3. Comentários e observações sobre obra literária e artística. “Em geral, “anotar” é o ato de expressar conhecimento sobre um recurso, anexando nomes, atributos, comentários, descrições, etc., a um documento ou a uma parte selecionada em um recurso” (Zeng, 2019, p. 20, tradução nossa). Ainda, segundo a autora, a anotação fornece informações adicionais (metadados) sobre um recurso existente.

Nesse sentido, o conceito de “anotação semântica” potencializa anotação, visto que “enriquece os dados não estruturados a um conhecimento de domínio estruturado, desde que aplicados efetivamente os padrões W3C, os quais são semanticamente marcados e expressos com relacionamentos digitados, estruturando os dados para processos automáticos de identificação de relações” (Clarke, 2015, p. 8, tradução nossa).

Entre os processos de anotação, enriquecimento e estruturação de dados não estruturados, a anotação de imagem profunda (em inglês *Deep image annotation*) é um conceito emergente no tratamento de imagens para enriquecimento semântico que tem como objetivos melhorar a eficiência da recuperação e o acesso de imagens, sendo uma base para integrar automaticamente recursos de imagem e aprimorar a descoberta do conhecimento (Zeng, 2019).

Uma vasta quantidade de informação reside em imagens, fotografias, pinturas, diagramas e desenhos que podem ser vistos, mas não pesquisados, que não são descobertas, estão inacessíveis devido aos métodos de consulta tradicionais (Zeng, 2019).

Ao abordar as finalidades de organização do conhecimento e descoberta de imagens, Clarke (2015, p. 24) forneceu uma linha de objetivos: cada imagem deve ter um URI e um conjunto extensível de metadados.

Assim, Clarke (2015, p. 3, tradução nossa) destaque a inacessibilidade dos detalhes da subimagem em comparação com a recuperação da totalidade da imagem, sendo que são bem estabelecidas as técnicas usadas para dar suporte à pesquisa e navegação “com base em metadados aplicados à imagem inteira, mas o acesso de pesquisa e navegação a pontos ou regiões específicas nas imagens é um campo relativamente novo”.

A anotação de subimagem pode ser comparada a índices analíticos e de assuntos, por permitir a pesquisa e navegação nas imagens. Nesse sentido, os SOC juntamente com a anotação de imagem, como a indexação semântica e dados vinculados, “podem ajudar a criar

ferramentas para integrar o processo de anotação de imagem com os métodos da Organização do Conhecimento (KO)” (Clarke, 2015, p. 11, tradução nossa).

A convergência das disciplinas Organização do conhecimento, indexação semântica e anotação da imagem, foi buscada em um projeto de nome OASIS: *Open Annotation Semantic Indexing System*, um sistema para anotação semântica de imagens. A equipe do OASIS determinou várias diretrizes de design de alto nível (Clarke, 2015, p. 7, tradução nossa):

1. O sistema deve ser capaz de processar imagens de qualquer tamanho usando imagens de multirresolução
2. Todos os metadados de imagem e anotação devem ser armazenados como Dados Vinculados em um banco de dados gráfico RDF
3. Cada indivíduo a anotação deve ser atribuído a um HTTP-URI exclusivo
4. Imagens e anotações devem ser semanticamente indexadas usando vocabulários controlados. O padrão principal escolhido para o modelo de dados de alto nível é o W3C Open Annotation Data Model. A especificação Open Annotation foi selecionada para oferecer suporte a *Linked Data* e também porque o modelo fornece opções de extensibilidade para anotações futuras de conteúdo em outros tipos de mídia.

No projeto OASIS, os metadados de imagem e anotação e a terminologia de indexação associada, são armazenados como dados conectados em um banco de dados RDF, oferecendo várias vantagens estratégicas:

1. A portabilidade de dados é alcançada usando um modelo de dados de padrões abertos
2. A padronização de dados é promovida pela capacidade de adotar ontologias e predicados de autoridades externas, como Dublin Core, CIDOC, FOAF, GeoNames, OWL, PROV e SKOS
3. A reutilização de dados é promovida pela capacidade de adotar e integrar vocabulários controlados com curadoria externa e conjuntos de dados, como LC, Getty, GeoNames e DBpedia
4. A análise de dados é facilitada pela capacidade de construir consultas baseadas em padrões do banco de dados gráfico
5. O compartilhamento de dados é habilitado pela opção de publicar um endpoint SPARQL (Clarke, 2015, p. 8, tradução nossa).

As anotações são URIs HTTP exclusivos atribuídos, assim, atribuir um identificador exclusivo para anotações individuais é imprescindível para indexação de subimagens e o acesso às informações.

Um conjunto extensível de metadados pode então ser criado para cada anotação única, incluindo as coordenadas do ponto ou área delimitada, o nível

de zoom em que o elemento deve ser visualizado, um rótulo e um texto descritivo (Clarke, 2015, p. 9, tradução nossa).

Tanto as imagens (total ou subimagem) e anotações podem ser indexadas semanticamente usando a terminologia de vocabulário controlado.

A indexação semântica significa que propriedades como tópico, local, criador, podem ser usadas para indexar itens de conteúdo (ou seja, imagens ou anotações de subimagem) para terminologia. Com o Linked Data, os relacionamentos de indexação são estruturados na linguagem RDF como triplos sujeito-predicado-objeto além disso são vinculados a fontes confiáveis (como Dublin Core Terms, FOAF, SKOS e OWL) (Clarke, 2015, p. 10, tradução nossa).

A “terminologia de vocabulário controlado significa especificamente que a indexação é realizada usando entidades conceituais em vez de *strings* de texto simples” (Clarke, 2015, p.10, tradução nossa). As entidades conceituais também possuem URIs HTTP e metadados descritivos associados de KOS LOD (Zeng; Mayr, 2018). Ou seja, usar um vocabulário controlados de termos, como um tesouro usando relacionamentos hierárquicos, associativos e de equivalência.

Ao usar dados vinculados e vocabulários controlados de autoridade (GeoNames, Getty AAT e IconClass), os conceitos usados para “descrever imagens e anotações de subimagem podem se tornar *gateways* para conectar a conjuntos de dados externos adicionais que também estão relacionados aos mesmos conceitos” (Clarke, 2015, p. 10, tradução nossa).

O quadro 10 mostra os diferentes níveis para anotação profunda da imagem, tanto no nível da imagem total, quanto no nível da anotação de uma região da imagem.

Quadro 10 - Metadados e níveis de anotação

METADADOS EM NÍVEL DE IMAGEM	METADADOS DE NÍVEL DE ANOTAÇÃO
A imagem é anotada na sua totalidade. Descreve a imagem total.	São anotados pontos de interesse específicos ou regiões de interesse em uma imagem.
As propriedades descritivas: título de imagem, uma descrição longa, data e criador.	Propriedades descritivas: um rótulo de anotação e uma descrição longa
Tecnologias para gravar metadados capturados automaticamente por câmeras	Metadados para as coordenadas espaciais do ponto ou área e o nível de zoom definido para o ponto ou região.

Fonte: Clarke (2015, p. 11-12, tradução nossa).

Em comparação com obras textuais, é possível listar os recursos como um índice, decompondo a imagem em seções com base nas anotações de regiões e partes, “assim listas

alfabéticas de recursos anotados fornecem um índice para uma imagem ou podem ser organizados em hierarquias” (Clarke, 2015, p. 12, tradução nossa).

Assim, a organização de características em hierarquias, de modo a decompor as características complexas da imagem, para a descoberta dos conteúdos. Dessa maneira, as características visuais de uma imagem “organizadas em hierarquias com base em conjuntos e subconjuntos temáticos ou associativos, ou para representar a decomposição espacial de uma imagem” (Clarke, 2015, p. 12, tradução nossa).

A indexação semântica significa especificamente que uma relação semanticamente expressiva é usada para indexar um item de conteúdo (ou seja, imagem ou anotação de subimagem) para terminologia descritiva (realizada a partir do controle de vocabulário proporcionado por um SOC sempre que possível).

São apresentados exemplos de ferramentas e padrões de anotação de imagens proprietárias e de código aberto que permitem que as imagens sejam anotadas de modo total ou parcial, além de rótulos e descrições de texto, incluem de acordo a Clarke (2015, p. 6, tradução nossa):

IIPImage³⁰
 MarQueed³¹
 SZoter³²
 Zoomify³³

Não existe um padrão internacional de domínio cruzado para anotação de imagens, mas existem várias iniciativas que visam padronizar metadados de imagens e anotações em domínios específicos de acordo (Clarke, 2015).

3.2.3 AdA ontologia

Outra iniciativa, voltada para anotação de vídeo, é a filmontologia da AdA ³⁴ (*Affektrhetoriken Des Audiovisuellen/Audio-Visual Rhetorics of Affect*), um vocabulário sistemático e modelo de dados de termos e conceitos analíticos de filmes para anotações semânticas de vídeo. O vocabulário é desenvolvido em colaboração com pesquisadores, acadêmicos de cinema e cientistas da computação, para apoiar estudos de filmes no contexto

³⁰ <http://iipimage.sourceforge.net>

³¹ <https://www.marqueed.com>

³² <http://www.szoter.com>

³³ <http://www.zoomify.com>

³⁴ AdA Filmontology. Disponível em: <https://projectada.github.io/ontology/>

digital. Para isso, foram constituídos métodos padronizados e sistemáticos para realizar o processo de anotação conjunta de corpora audiovisual, que possibilite análises comparativas e sistemáticas de filmes. A conceituação da ontologia AdA se refere as especificidades técnicas dos filmes, fundamentais nos processos de análise fílmica.

Em vista disso, o vocabulário é baseado em uma conceituação analítica e metodológico sobre a análise fílmica, modelada em ontologia OWL legível por máquina para publicar anotações como LOD, para o intercâmbio e reconciliação de dados. Ademais, é composto por mais de 400 valores individuais, atribuído a mais de 80 tipos de anotação organizadas em 8 níveis gerais de descrição, como Acústica, Montagem, Composição de Imagem ou Câmera.

A ontologia AdA é estruturada em três tipos diferentes de conceitos analíticos de filmes, como mostra o quadro 11 e a figura 2:

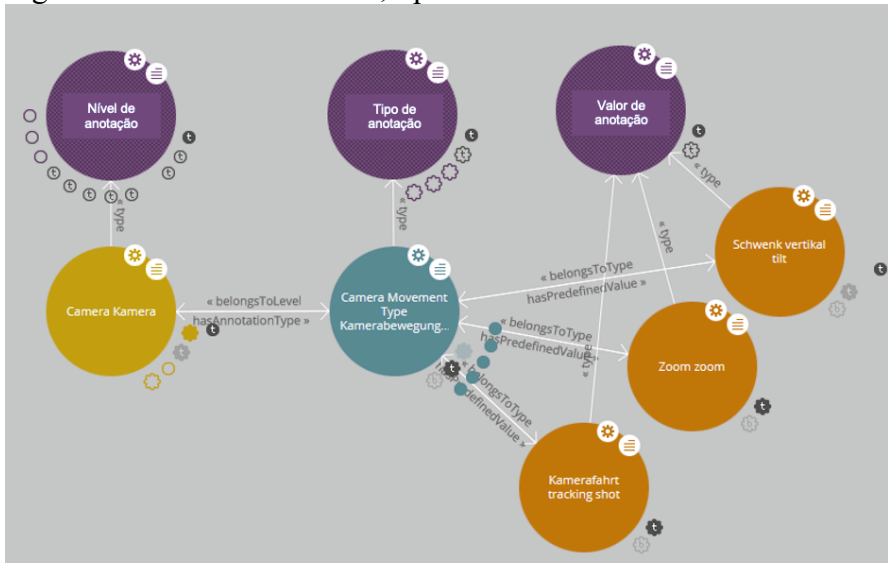
Quadro 11 - Classes AdA ontologia

Níveis de Anotação	Categorias gerais de descrição (e.g, acústica, câmera).
Tipos de Anotação	Conceitos de análise de filmes sob os quais um filme é analisado (e.g, clima musical, velocidade de movimento da câmera).
Valores de Anotação	Um conjunto de valores predefinidos atribuídos a cada tipo de anotação (e.g, neutro, tenso, feliz, triste, agressivo para humor musical).

Fonte: AdA Filmontology Website (2021).

A figura 2 mostras a estrutura de níveis, tipos e valores da ontologia:

Figura 2 - Estrutura de níveis, tipos e valores visualizados no LodLive



Fonte: AdA Filmontology Github (2021).

A filme-ontologia AdA consiste em **8 Níveis de anotação**, **78 Tipos de anotação** e **501 Valores de anotação**. Cada nível, tipo e valor tem um identificador de recurso exclusivo (URI), um nome e uma descrição em inglês e alemão. Os tipos recebem códigos de cores para melhor a diferenciação em ferramentas de software de anotação. A análise da ontologia AdA no Protégé³⁵, a partir do arquivo *ada_ontology.owl*³⁶ baixado do repositório público no Github, identificou os tipos de entidades e relacionamentos estruturados no modelo de dados. O quadro 12 mostra as entidades da AdA ontologia:

Quadro 12 - Entidades da ontologia AdA

Entidade (rdfs:label)	Tipo	Domínio (rdfs:label)	Alcance (rdfs:label)
Annotation Level ³⁷ @en	owl:Class	-	-
Annotation Type ³⁸ @en	owl:Class	-	-
Annotation Value ³⁹ @en	owl:Class	-	-
hasAnnotationType ⁴⁰ @en	owl:ObjectProperty	Annotation Level @en	Annotation Type @en
hasPredefinedValue ⁴¹ @en	owl:ObjectProperty	Annotation Type @en	Annotation Value @en
belongsToLevel ⁴² @en	owl:ObjectProperty	Annotation Type @en	Annotation Level @en
belongsToType ⁴³ @en	owl:ObjectProperty	Annotation Value @en	Annotation Type @en

Fonte: Elaborado pela autora.

³⁵ O Protégé é um editor de ontologias gratuito e de código aberto e um sistema de gerenciamento de conhecimento.

³⁶ Os dados são disponibilizados sob os termos da Licença Creative Commons Attribution-ShareAlike 3.0 License.

³⁷ <http://ada.cinepoetics.org/ontology/2021/05/19/AnnotationLevel>

³⁸ <http://ada.cinepoetics.org/ontology/2021/05/19/AnnotationType>

³⁹ <http://ada.cinepoetics.org/ontology/2021/05/19/AnnotationValue>

⁴⁰ <http://ada.cinepoetics.org/ontology/2021/05/19/hasAnnotationType>

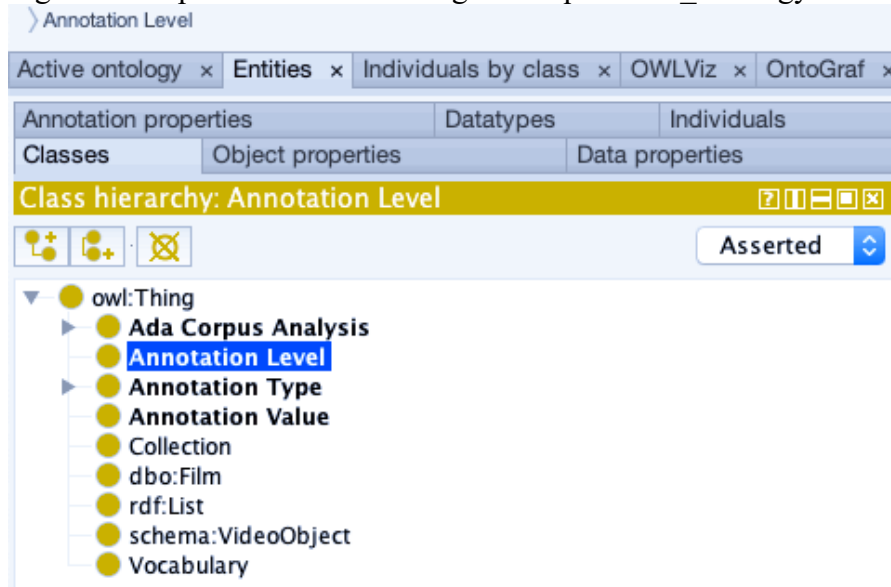
⁴¹ <http://ada.cinepoetics.org/ontology/2021/05/19/hasPredefinedValue>

⁴² <http://ada.cinepoetics.org/ontology/2021/05/19/belongsToLevel>

⁴³ <http://ada.cinepoetics.org/ontology/2021/05/19/belongsToType>

A ontologia AdA no editor Protégé possui uma taxonomia das classes (owl:class), a figura 3 mostra a classe *Annotation Level*:

Figura 3 - Captura de tela do Protégé do arquivo ada_ontology.owl



Fonte: AdA Filmontology Github (2021).

Na classe *Annotation Level* existe 8 instâncias de classe como mostra a figura, destacado em azul a instância “*Acoustics*”, como mostra a figura 4:

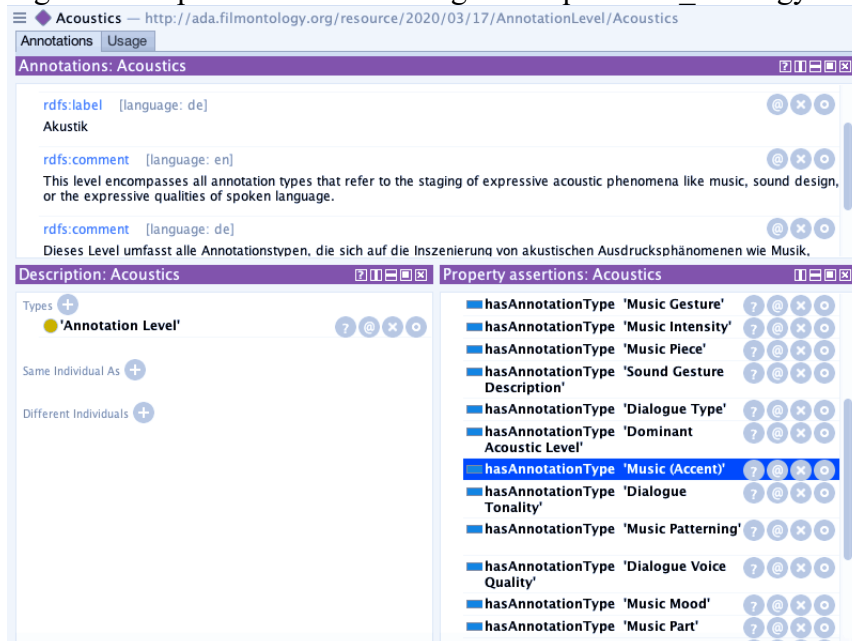
Figura 4 - Captura de tela do Protégé do arquivo ada_ontology.owl



Fonte: AdA Filmontology Github (2021).

A instância *Acoustics* é um tipo de *Annotation Level* e tem como propriedade *hasAnnotationType*, que pode ter como valor, por exemplo, a instancia “*Music (Accent)*”, como mostra a figura 5.

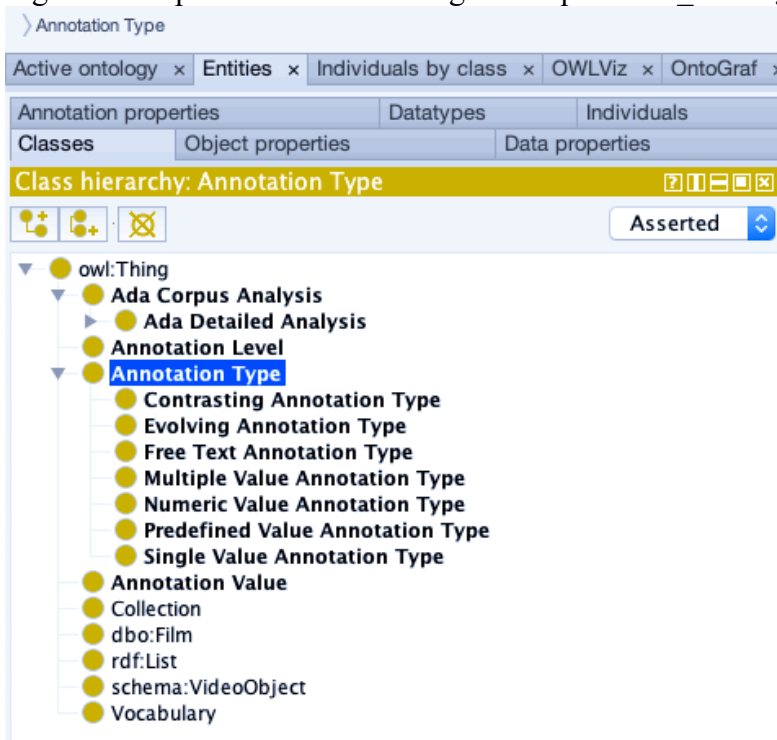
Figura 5 - Captura de tela do Protégé do arquivo *ada_ontology.owl*



Fonte: AdA Filmontology Github (2021).

Ademais, “*Music (Accent)*” é uma instância da a **superclasse** *Annotation Type* (fig.6) e de algumas de suas subclasses:

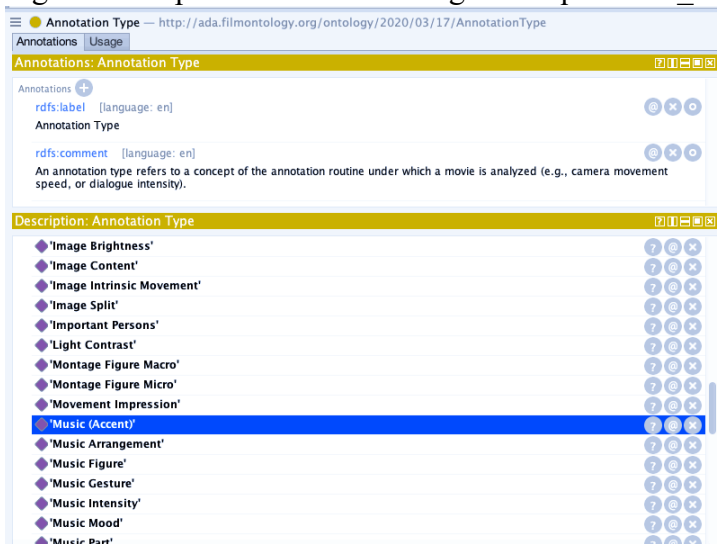
Figura 6 - Captura de tela do Protégé do arquivo ada_ontology.owl



Fonte: AdA Filmontology Github (2021).

A figura 7 mostra as instâncias da classe *Annotation Type*, entre elas “*Music (Accent)*”:

Figura 7 - Captura de tela do Protégé do arquivo ada_ontology.owl

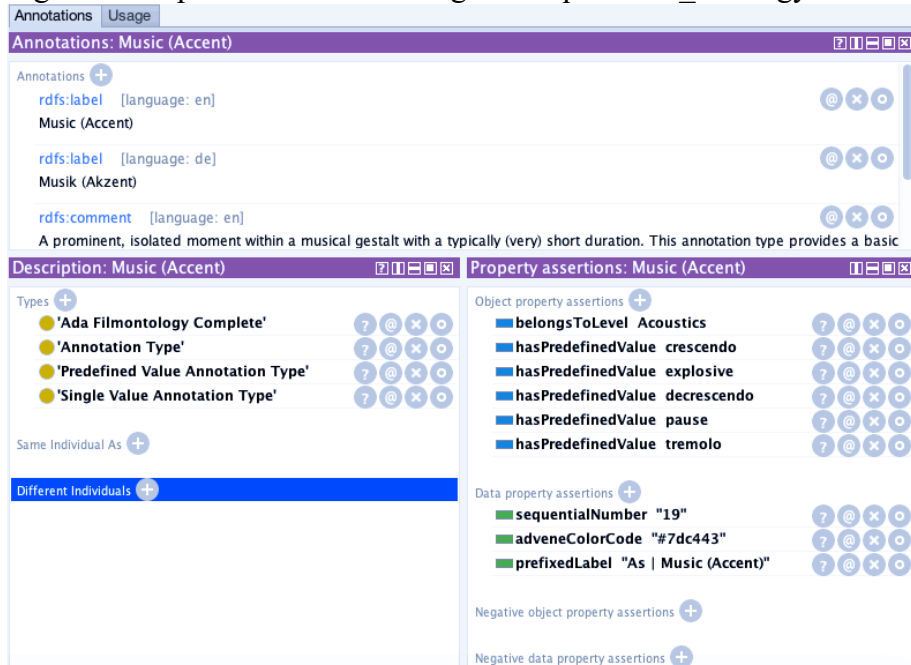


Fonte: AdA Filmontology Github (2021).

A instancia “*Music (Accent)*” é um tipo de *Annotation Type* e tem a propriedade `hasPredefinedValue` com os seguintes valores pré-definidos: *crescendo*, *explosive*,

decrecendo, *pause* e *tremolo*. Igualmente, há uma propriedade `belongsToLevel` que conecta a instância a sua categoria principal “*Acoustics*”, como mostra a figura 8.

Figura 8 - Captura de tela do Protégé do arquivo `ada_ontology.owl`



Fonte: AdA Filmontology Github (2021).

A entidade “*Acoustics*” é uma instância da classe **Nível de anotação** (*Annotation Level*) e por isso, ela pode ser usada com a propriedade `hasAnnotationType` cujo valores são entidades da classe **Tipo de anotação** (*Annotation Type*), como mostra o quadro 13.

Quadro 13 - Exemplo do valor Acoustics

Acoustics	
http://ada.cinepoetics.org/resource/2021/05/19/AnnotationLevel/Acoustics	
rdfs:comment	Este nível abrange todos os tipos de anotação que se referem à encenação de fenômenos acústicos expressivos como música, design de som ou as qualidades expressivas da linguagem falada.
rdfs:label	Acoustics @en
rdf:type	owl:NamedIndividual AnnotationLevel
hasAnnotationType	DialogueEmotion DialogueIntensity DialogueTonality DialogueType DialogueVoiceQuality DominantAcousticLevel MusicAccent MusicArrangement MusicFigure MusicFigurePatterning MusicGesture MusicIntensity MusicMood MusicPart MusicPiece MusicTonality SoundGestureDescription SoundGestureDynamics SoundSegment Volume

Fonte: AdA Filmontology Website (2021).

O valor “*Music (Accent)*” é uma instância da classe **Tipo de anotação** (*Annotation Type*), e por isso, ela pode ser usada com a propriedade `hasPredefinedValue` cujo valores são entidades pré-definidas da classe **Valor de anotação** (*Annotation Value*), como mostra o quadro 14.

Quadro 14 - Exemplo do valor Music

Music (Accent)	
http://ada.cinepoetics.org/resource/2021/05/19/AnnotationType/MusicAccent	
rdfs:comment	Um momento proeminente e isolado dentro de uma gestalt musical com uma duração tipicamente (muito) curta. Este tipo de anotação fornece uma classificação básica de acentos musicais.
rdf:type	owl:NamedIndividual AdaFilmontologyComplete AnnotationType PredefinedValuesAnnotationType SingleValueAnnotationType
belongsToLevel	Acoustics
haspredefinedValue	MusicAccent_crescendo MusicAccent_decrescendo MusicAccent_explosive MusicAccent_pause MusicAccent_tremolo

Fonte: AdA Filmontology Website (2021).

A entidade “*crescendo*” é uma instância da classe **Valor de anotação** (*Annotation Value*) e por isso, faz parte do domínio da propriedade `belongsToType` com valor de *MusicAccent*, como mostra o quadro 15.

Quadro 15 - Exemplo do valor Crescendo

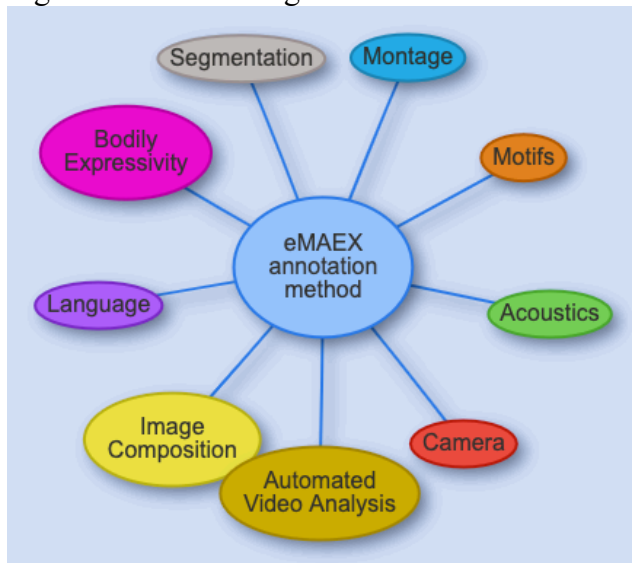
Crescendo	
http://ada.cinepoetics.org/resource/2021/05/19/AnnotationValue/MusicAccent_crescendo	
rdfs:comment	Intensificação notável da música, (e.g) aumentando o volume ou aumentando o tom.
rdfs:label	crescendo @en
rdf:type	owl:NamedIndividual AnnotationValue
belongsToType	MusicAccent

Fonte: AdA Filmontology Website (2021).

Dessa forma, a análise da visualização interativa (fig.9) da filmontologia AdA na ferramenta OntoViz⁴⁴, mostra 9 níveis de anotação: Montagem, Linguagem, Expressão Corporal, Análise de Vídeo Automática, Câmera, Composição da Imagem, Segmentação, Acústica e Padrões.

⁴⁴ <https://ada.cinepoetics.org/ontoviz/>

Figura 9 - Filmontologia AdA na ferramenta OntoViz

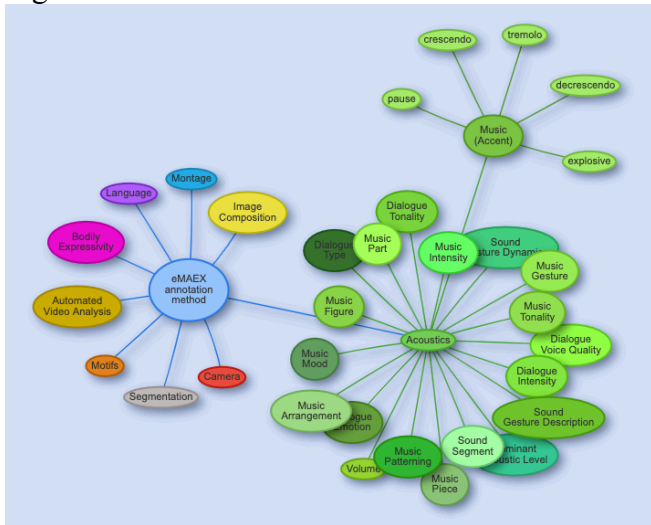


Fonte: AdA Filmontology Website (2021).

De tal modo, a entidade “*Acoustics*” é uma instância da classe **Níveis de anotação** e está conectada com a instância “*MusicAccent*” através das propriedades `hasAnnotationType` e `belongsToLevel`. Do mesmo modo, “*MusicAccent*” é uma instância da classe **Tipo de anotação** e está conectada com “*crescendo*” através da propriedade `hasPredefinedValue` e `belongsToType`. Devido a isso, a entidade “*crescendo*” é uma instância da classe **Valores de anotação** e um valor predefinido atribuído a instância “*MusicAccent*”.

A propriedade `hasAnnotationType` tem como domínio (*domain*) a classe Nível de anotação (*Annotation Level*) e como alcance (*range*) a classe Tipo de anotação (*Annotation Type*), além disso, é o inverso da propriedade `belongsToLevel`. A propriedade `hasPredefinedValue` tem como domínio (*domain*) a classe Tipo de anotação (*Annotation Type*) e como alcance (*range*) a classe Valor de anotação (*Annotation Value*)

Figura 10 - Valor Acoustics na ferramenta OntoViz



Fonte: AdA Filmontology Website (2021).

A filme-ontologia ADA é codificada como ontologia OWL usando o RDF e o *Resource Description Framework Schema* (RDFS), de maneira que, as declarações são formuladas na forma de **sujeito, predicado, objeto** (triplos RDF).

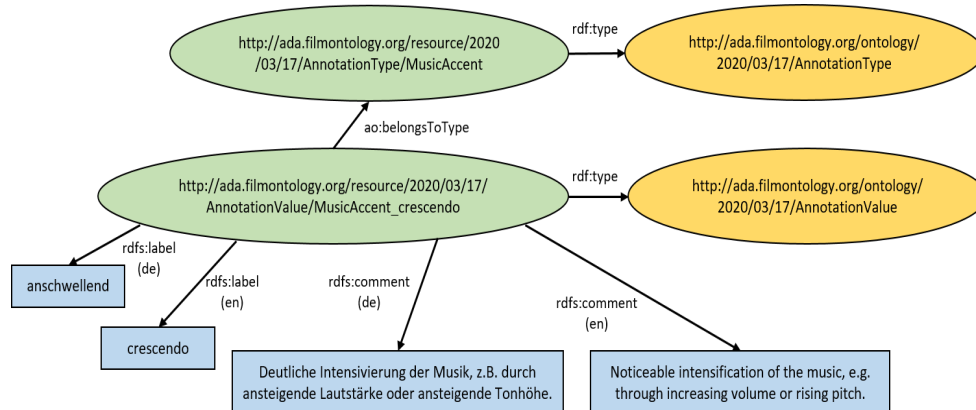
Portanto, como mostra a figura 10, a instância “*MusicAccent*⁴⁵” é um tipo de *AnnotationType* e “*MusicAccent_Crescendo*” é um tipo de *AnnotationValue*. Conseqüentemente, a propriedade `ao:belongsToType`, que conecta essas duas instâncias de classe, tem como domínio a classe *AnnotationValue* e como alcance a classe *AnnotationType*.

As instâncias “*MusicAccent*” e “*Acoustics*”⁴⁶ se conectam através da propriedade `ao:belongsToLevel`, que tem como domínio *AnnotationType* e alcance *AnnotationLevel*, logo a primeira (*MusicAccent*) refere-se a um conceito que pertence a uma categoria (*Acoustics*), está por sua vez, é uma categoria que agrupa um conjunto de tipos de anotação semelhantes.

⁴⁵ Analisamos a filme-ontologia ADA no editor Protégé. “*MusicAccent*” tem como definição (rdfs:comments): Um momento proeminente e isolado dentro de uma gestalt musical com uma duração tipicamente (muito) curta. Este tipo de anotação fornece uma classificação básica de acentos musicais (tradução nossa).

⁴⁶ No editor protégé “*Acoustics*” tem como definição (rdfs:comments): Este nível abrange todos os tipos de anotação que se referem à encenação de fenômenos acústicos expressivos como música, design de som ou as qualidades expressivas da linguagem falada (tradução nossa).

Figura 11 - Definição de um valor de anotação como grafo RDF



Fonte: AdA Filmontology Website (2021).

A ontologia AdA é o resultado das anotações filme-analíticas realizadas por especialistas da área usando o eMAEX, onde são realizadas anotações semânticas diretamente no material audiovisual. O sistema tem entradas para os metadados descritivos, para anotar o material com dados estruturados compatíveis com os conceitos do vocabulário de anotação e da ontologia de domínio AdA. Dessa forma, os exemplos a seguir possuem as seguintes entradas:

NÍVEL DE ANOTAÇÃO	ACOUSTICS
TIPO DE ANOTAÇÃO	MUSICACCENT
VALOR DA ANOTAÇÃO	CRESCENDO

O eMAEX é um “método sistemático pelo qual as emoções do espectador são estudadas como expressões multimodais cinematográficas de imagens de movimento; a análise de mídia eletrônica de imagens de movimento expressivo” (AdA, 202, tradução nossa). Por conseguinte, o recurso do método de anotação eMAEX é usado como ponto de entrada para anotações que se referem a fragmentos de um material audiovisual utilizando *timecodes*, realizadas por especialistas da área. Criando-se dessa forma um conjunto de metadados para um corpus formado por material audiovisual, a depender do projeto e do tema do projeto. Por exemplo, a equipe do projeto FU Berlin compilou um corpus de vídeo com longas-metragens, documentários, notícias de televisão e vídeos da web sobre a crise financeira global após 2007, a fim de investigar retóricas audiovisuais de afeto. Como os filmes possuem direitos autorais, o projeto compartilhou apenas os metadados semânticos do corpus.

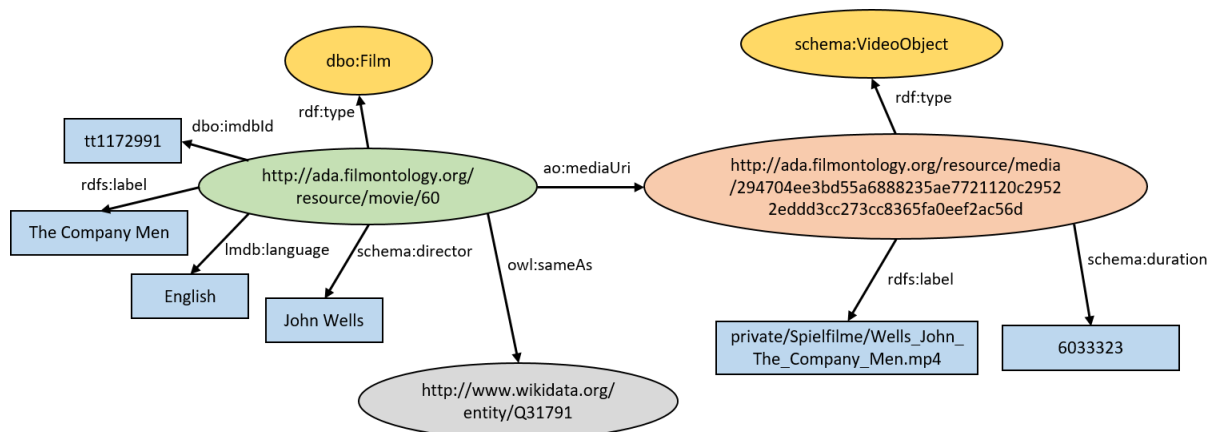
Os metadados do corpus possuem os **metadados gerais dos filmes (Obras)**, que contêm informações como o identificador, título, diretor e resumo do filme. E os **metadados do arquivo (Mídia)**, com o tempo de execução exato, o nome do arquivo e um identificador

determinado pela soma de verificação SHA256 do arquivo. Por exemplo, os metadados do longa-metragem “The Company Men” estão disponíveis da seguinte forma:

Identifer	60
Title	The Company Men
Year	2010
Language	English
Writer	John Wells
Director	John Wells
Runtime	104 min
Abstract	The story centers on a year in the life of three men trying to survive a round of corporate downsizing at a major company - and how that affects them, their families, and their communities.
Actors	Ben Affleck, Tommy Lee Jones, Chris Cooper, Suzanne Rico
IMDB	http://www.imdb.com/title/tt1172991
Release date	11.02.2011
Genre	Feature Film

Os metadados do corpus são codificados como dados RDF usando vocabulários LOD, como *DBpedia Ontology*, *Schema.org* e *Linked Movie DataBase*. Ainda, os recursos podem ser vinculados a entidades Wikidata e DBpedia. A figura 12 mostra a codificação RDF dos metadados gerais do filme “*The Company Men*” como um grafo RDF:

Figura 12 - Metadados do corpus AdA codificados como dados RDF



Fonte: AdA Filmontology Website (2021).

Os metadados do corpus também estão disponíveis para download no repositório GitHub do projeto AdA, como exportação RDF no formato Turtle. Dessa forma, é possível visualizar o arquivo Turtle (.ttl) com a serialização RDF do recurso “*The Company Men*”, como

definido na imagem, junto com os vocabulários e ontologias usadas nos metadados semânticos do corpus:

Figura 13 - Captura de tela do arquivo metadata_corpus.ttl baixado no github

```

armov:60 a
  rdfs:label "The Company Men" ;
  ao:mediaUri arm:294704ee3bd55a6888235ae7721120c29522eddd3cc273cc8365fa0eef2ac56d ;
  linkedmdb:language "English" ;
  linkedmdb:writer "John Wells" ;
  dbowork:runtime "104" ;
  dbo:abstract "The story centers on a year in the life of three men trying to survive a round of corporate dowa";
  dbo:actor "Ben Affleck, Tommy Lee Jones, Chris Cooper, Suzanne Rico" ;
  dbo:filmVersion "Bluray, EAN 0886979286993" ;
  dbo:imdbId "tt1172991" ;
  dbo:releaseDate "2011-02-11" ;
  dbo:year "2010" ;
  dcterms:isPartOf <http://ada.filmontology.org/resource/Corpus> ;
  schema:director "John Wells" ;
  schema:genre "Feature Film" ;
  schema:url "http://www.imdb.com/title/tt1172991" ;
  owl:sameAs <http://www.wikidata.org/entity/Q31791> , <http://dbpedia.org/resource/The_Company_Men> .

```

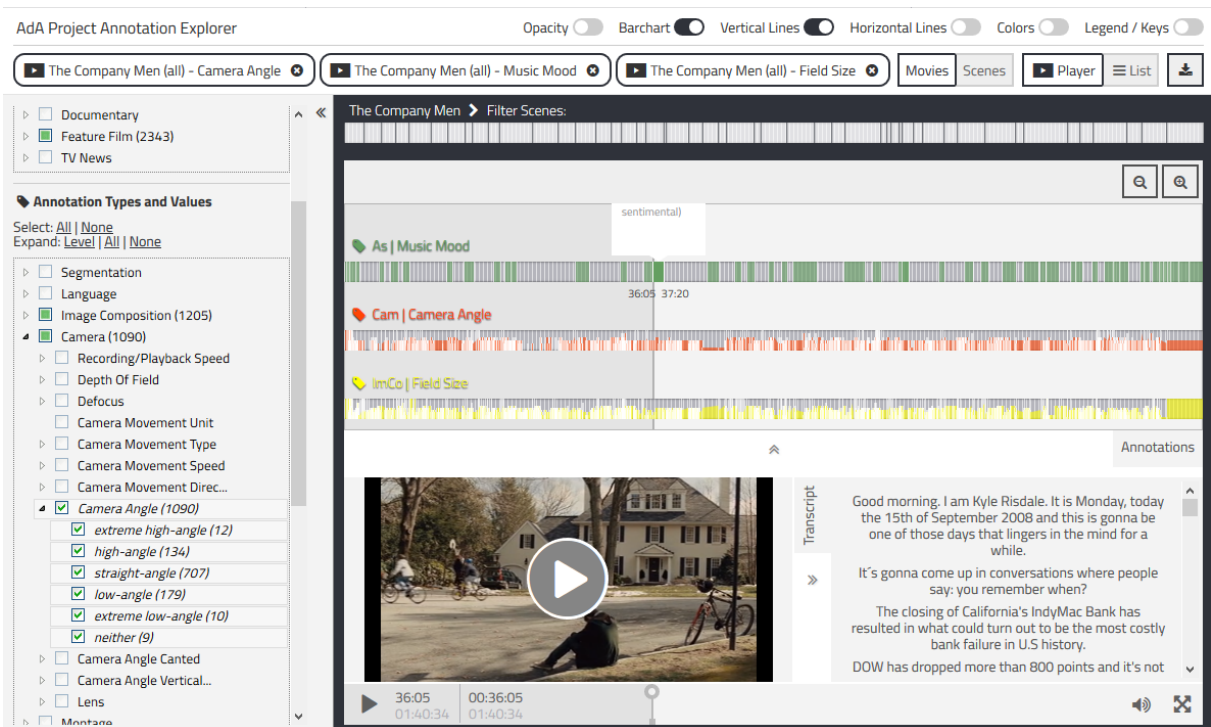
Fonte: AdA Filmontology (2021).

O processo de anotação é realizado com Advene, um kit de ferramentas de software gratuito para anotação de documentos audiovisuais, a interface permite a importação da filmontologia AdA e a exportação de anotações de vídeo compatíveis com W3C como o RDF. Cada fragmento se refere a uma anotação de um filme e os conjuntos de dados criados no projeto AdA formam milhares de anotações com referências baseadas no timecode do material de vídeo original. O conteúdo é anotado com os conceitos analítico do filme conforme definido pelos tipos e valores de anotação na filmontologia AdA. Cada anotação também contém metadados sobre o autor e a data de criação. Por exemplo, as seguintes informações estão disponíveis para caracterizar o movimento de câmera no minuto “00:41” do longa-metragem “The Company Men”:

Annotation ID	ed63d084-717f-11e9-99b8-0242ac130003
Begin timecode	00:41:29.900
End timecode	00:41:50.620
Annotation Type	Camera Movement Type
Annotation Value	tracking shot
Author	anton
Date	2018-05-04 22:10:22

Os dados são gerenciados usando o RDF *triplestore* OpenLink Virtuoso⁴⁷ e o LodView⁴⁸, um software para desreferenciação de IRI compatível com o padrão W3C. Também foi desenvolvida uma visualização interativa da filmontologia AdA que pode ser acessada na ferramenta OntoViz. A filmontologia AdA está disponível para download no repositório GitHub.

Figura 14 - Plataforma online para exploração de anotação da AdA



Fonte: AdA Filmontology Website (2021).

Para visualizar a ontologia AdA foi desenvolvido uma plataforma online para exploração de anotação, para a consulta de dados de anotação na aplicação web, como mostra a figura 14.

3.2.4 Plataforma I-Media-Cities

O projeto I-Media-Cities⁴⁹ (IMC) é mais uma iniciativa liderada por 9 instituições culturais europeias (arquivos cinematográficos e audiovisuais de 8 países), 5 instituições de pesquisa e 2 fornecedores tecnológicos concebidos para enriquecer, partilhar, fornecer acesso

⁴⁷ <https://virtuoso.openlinksw.com/>

⁴⁸ <https://github.com/LodLive/LodView>

⁴⁹ <http://imediacities.eu>

e uso de conteúdo digital relacionado às cidades. Os tipos de materiais são obras audiovisuais ficcionais e não ficcionais, que descrevem as cidades em todas as perspectivas, incluindo a transformação física e a dinâmica social a partir do final do século XIX, “é o conteúdo central do ambiente do IMC, através do qual é possível coletar, processar, integrar e compartilhar esta grande coleção de "imagens em movimento", através de um sistema associado de metadados (Caraceni *et al.*, 2018, tradução nossa).

A plataforma realiza o processo de anotação automática de conteúdos multimídia, enriquecidos com metadados e que serão posteriormente anotados manualmente. As instituições culturais, históricas e museológicas em toda a Europa estão cada vez mais detendo grandes coleções digitais, digitalizadas ou digitais, que podem ter um impacto significativo em muitos campos de pesquisa, tanto em Ciências Sociais como Humanas. Esse cenário ajuda a construir conexões entre essas instituições e as demandas do campo de pesquisa, diálogos ainda em desenvolvimento, mas cada vez mais explorado, no sentido que ambas possam aprimorar o uso de tecnologias da informação para criar oportunidade e melhorar os aspectos das pesquisas em humanas e o acesso às coleções (Caraceni *et al.*, 2018).

Projetos como o IMC “ajudam a construir pontes entre os cidadãos europeus e o seu próprio patrimônio cultural e estimular a reutilização para muitos fins diferentes, tanto dentro como fora do contexto do arquivo” (Caraceni *et al.*, 2018, tradução nossa).

A plataforma IMC recebe conteúdos na forma de arquivos de vídeo, imagens, texto e XML com todos os metadados fornecidos pelos arquivos audiovisuais. Uma vez que os conteúdos são carregados dentro da plataforma, várias imagens automáticas e ferramentas de análise de vídeo, orquestradas através de um pipeline, analisam e extraem informações e metadados - na imagem, no frame ou no nível da tomada - que podem ser produzidos em XML ou JSON. Cineca é o principal parceiro tecnológico da IMC e a equipe envolvida no projeto vem do departamento de Aplicações e Inovação em SuperComputação (<http://www.hpc.cineca.it/>). Um dos objetivos mais desafiadores que os centros de HPC e, portanto, a Cineca, estão enfrentando hoje em dia é gerenciar o dilúvio de dados vindo de diversos grupos e projetos de pesquisa e para promover e apoiar uma nova maneira de conduzir pesquisas, uma nova ciência dominada por dados e centrada em dados, baseada em computação intensiva em dados (Caraceni *et al.*, 2018, tradução nossa).

Plataformas sociais contribuíram para o enriquecimento de dados na web, com a explosão informacional, aumentando o volume de dados compartilhados e armazenados. “Comentários sobre fotos ou vídeos compartilhados, avaliações de produtos e até mesmo menções de recursos da web em redes sociais podem ser considerados anotações” (Caraceni *et al.*, 2018, tradução nossa). No domínio do patrimônio cultural surgiu a demanda de não apenas

digitalizar e disponibilizar o acesso online do acervo, mas de inovação nas metodologias, métodos e procedimentos que permitem a criação e organização do conhecimento e a representação e transmissão da informação. Isto se refletiu na criação de serviços on-line centrados no usuário, onde itens de conteúdo são disponibilizados e é possível, por exemplo, fazer a anotação manualmente.

As anotações têm sido uma das principais ferramentas de colaboração utilizadas por arquivistas, pesquisadores e usuários gerais com experiências e experiências diferentes para criar, explorar e interpretar coleções de patrimônio cultural digital de forma colaborativa (Caraceni *et al.*, 2018, tradução nossa).

Dessa forma, projetos como o IMC ajudam a resolver problemas referente a atividade de indexação manual de imagens em movimento. Assim, (Caraceni *et al.*, 2018, tradução nossa) pontuam que a “anotação manual é um processo extremamente trabalhoso, especialmente em situações em que os vídeos precisam ser anotados no nível do segmento ou do quadro”.

O aspecto inovador do projeto IMC reside na criação de uma plataforma capaz de realizar um processo de anotação automática de itens de conteúdo multimídia que já foram enriquecidos com metadados e que serão posteriormente anotados manualmente (Caraceni *et al.* 2018, tradução nossa).

Ainda segundo o autor, as vantagens das anotações são a recuperação eficiente de dados em nível de segmento e quadro de vídeo, o uso de ferramentas automáticas de análise de imagem e vídeo, que analisam e extraem informações no nível de imagem, quadro ou tomada. Além disso, existem as ferramentas usadas para a anotação automática, como segmentação de tempo de vídeo, qualidade de vídeo, movimento de câmera, detecção de objeto e reconhecimento em uma seleção de elementos predefinidos.

O modelo de metadados IMC está dividido entre metadados descritivos provenientes dos arquivos, principalmente com base na norma CEN EN19507 e dados gerados através do enriquecimento automático ou manual de dados, foi modelado usando o W3C *Web Annotation Data Model*.

Os metadados do projeto IMC são criados por humanos e também por ferramentas automáticas de análise de imagem e vídeo.

Avanços recentes no campo da visão computacional e aprendizado de máquina permitem que essas ferramentas extraiam informações e metadados interessantes de imagens e vídeos, mais rápido do que qualquer ser humano

poderia, o que é especialmente importante com a quantidade crescente de dados digitais ou digitalizados (Caraceni *et al.*, 2018).

As ferramentas automáticas que produzem metadados de objetos adicionais no IMC são: segmentação temporal de vídeo, estimativa de movimento da câmera, qualidade da mediação e detecção de vídeo e reconhecimento de objetos.

Os metadados gerados por ferramentas de análise automática não podem ser usados diretamente no projeto IMC. Existem dois desafios a serem enfrentados quando os metadados são gerados por ferramentas automáticas: baixo nível semântico de anotações de dados estruturados e excesso de dados gerados.

Os “metadados são considerados de alto nível se forem diretamente interpretáveis e compreensíveis por uma pessoa ao observá-los ou consultá-los” (Caraceni *et al.*, 2018, tradução nossa). Desse modo, “anotar é, no básico, o ato de criar associações entre pedaços de informação. Às vezes, anotar significa adicionar um comentário sobre alguns recursos, ou mais, em geral, fazer uma relação entre dois recursos” (Caraceni *et al.*, 2018, tradução nossa).

No IMC, pode ser considerado um segmento da obra audiovisual para anotar um assunto específico em uma cena. O modelo de dados de anotação IMC é a recomendação do W3C chamada Modelo de dados de anotação da web (Web Annotation Data Model).

O modelo de anotação IMC combina as habilidades humanas e computacionais, com o uso integrado de anotação automática e manual. Assim, algoritmos fazem a análise visual de forma automática, com informações sobre a segmentação, qualidade, movimento da câmera e detecção de objetos do vídeo, as quais são posteriormente combinadas com anotações manuais. A abordagem de anotação combinada permite validar os resultados ou completar, ou adicionar um tipo diferente de informação.

Devido à grande variação nas fontes potenciais de informação usadas para descrever o conteúdo no IMC, fornecer ao usuário termos de anotação em uma linguagem comum predefinida, como um vocabulário controlado, taxonomia, etc., aumenta o potencial da anotação única, uma vez que o uso do mesmo termo anotado permite uma classificação comum do próprio conteúdo (Caraceni *et al.*, 2018, tradução nossa).

O IMC faz a divisão de diversos temas no campo da investigação de cidades para descrever a cidade tal como é apresentada em imagens e vídeos. “Para cada subdomínio, os elementos descritores e vocabulários relacionados são selecionados entre as quantidades de

dados estruturados utilizáveis para a descrição do patrimônio cultural” (Caraceni *et al.*, 2018, tradução nossa).

Um conjunto de vocabulários adotados para descrever o que está acontecendo em um fragmento de filme que mostra um ambiente urbano vem do projeto Ephemeral Films do Austrian Film Museum e do Ludwig Boltzmann Institute (<http://efilms.ushmm.org/>). Esses vocabulários podem ter até três níveis hierárquicos e mapeados a partir de vários vocabulários, mais notavelmente os da Biblioteca do Congresso. O objetivo do sistema de anotação IMC é “fornecer conexão a domínios de conhecimento abertos, criando assim a base para um sistema de anotação semântica interligado a outros conjuntos de dados abertos disponíveis na web” (Caraceni *et al.*, 2018, tradução nossa), como o Geonames⁵⁰ já mencionados e Wikidata⁵¹.

⁵⁰ <http://www.geonames.org>

⁵¹ <https://www.wikidata.org>

4 SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO NA WEB

A organização e a representação do conhecimento no século XXI estão presentes não mais apenas em ambientes físicos, como bibliotecas, arquivos, museus e demais unidades de informação, todavia foi somado o ambiente web. E existem diferentes desafios de organização da informação e do conhecimento, perante o enorme volume de dados e informações que são criadas, compartilhadas, armazenadas e representadas diariamente no universo digital.

No surgimento da recuperação e armazenamento de informação em computadores a partir da década de 50, as principais preocupações eram os altos custos e as limitações da tecnologia. Após superadas estas questões foi possível a organização do conhecimento encontrar um espaço significativo para a linguagem, os códigos culturais e o sujeito da informação (Barreto, 2007), aliada as questões dos processos de organização do conhecimento e dos SOC. De acordo a Saracevic (1996), Capurro (2003), Barreto (2007) e Buckland (2013), depois que apareceram soluções para os problemas de armazenamento e a recuperação da informação, como os altos custos e as limitações da tecnologia, foi possível que as redes da ciência da informação, se voltassem para a linguagem e os códigos culturais. O sujeito da informação, como focos de interesse e de possibilidade de superar desafios da área.

Todavia, no século XXI, com a superação dos desafios tecnológicos, como o espaço de armazenamento, ocorreu o aumento da velocidade da rede de computadores e com o desenvolvimento da WWW, a preocupação é com o excesso de dados e informações, que trouxeram desafios de representação e organização do conhecimento nesses espaços. Nesse contexto é importante para a pesquisa os estudos sobre representações do conhecimento que buscam representar o conhecimento por meio de SOC na web. “Os diferentes aspectos da representação do conhecimento, como os estudos da cognição e modelos de representação em sistemas de IA, envolvem a linguística, filosofia, psicologia, ciência da informação e inteligência artificial” (Bräscher, 2014, p. 175, tradução nossa).

Nesse contexto é importante para a pesquisa os estudos sobre representações do conhecimento que buscam representar o mundo real por meio de SOC. “Uma relação semântica contém pelo menos dois lugares vagos ou slots a serem preenchidos. Um conceito que participa da relação deve ter certas características semânticas ou pertencer a uma determinada categoria de conceitos” (Bräscher, 2014, p. 176).

No cenário contemporâneo, duas abordagens estão experimentando estratégias que combinam conhecimento legível por humanos e por máquinas, para lidar com o grande volume de dados: a ciência de dados aliada a inteligência artificial, que provem de métodos estatísticos

para coletar, tratar, analisar e prever, grandes conjuntos de dados. E, por outro lado, a Web Semântica com dados estruturados e conectados, junto a ontologias que lidam com a lógica descritiva (LD) e bases de conhecimento processáveis por máquina. Sendo um conjunto de práticas e uma pilha de tecnologias, que fazem parte do domínio da organização do conhecimento e a base teórica metodológica dessa pesquisa.

A Web Semântica é uma extensão da web tradicional, uma expansão da web de documentos, uma mudança na abordagem convencional com identificadores para documentos da web (páginas web) e protocolos que tratam formalmente a totalidade de uma página da web mediante cabeçalhos HTML, protocolos HTTP e hiperlinks. Para um enfoque no conteúdo representado no interior do documento web, realizando a estruturação em esquemas de dados legíveis por máquinas e conectados a outros recursos de informação com sistemas de identificadores globais e o uso do RDF para realizar declaração e descrever recurso informacionais.

Os protocolos da web afetam o uso de tesouros, pois antes da web, eles eram eficazes em ambientes controlados na forma de SOC claramente definidos para acessar informações, mas no ambiente web, a heterogeneidade de formatos e estruturas, com o grande volume de dados e recursos, impossibilitou esse modelo anterior. Logo, “os caminhos de acesso que levam aos elementos de informação não são delimitados ou estabelecidos previamente e a constante atualização de conteúdo impossibilita praticamente a aplicação de tesouros diretamente na web”. (Pastor-Sanchez; Mendez; Muñoz, 2009, tradução nossa).

Inevitavelmente, o futuro do tesouro na Web semântica passa necessariamente pela permeabilidade com outras propostas (ontologias e topic maps) ou pelo seu uso combinado, ao estabelecer uma mudança de paradigma na criação de tesouros, com foco nos tesouros conceituais (Matthews *et al.*, 2001 *apud* Pastor-Sanchez *et al.*, 2009, tradução nossa).

De maneira que, os desafios atuais expandem o contexto dos documentos e índices, sendo necessário entender como representar o conhecimento em outros campos, como, por exemplo, a estrutura dos registros e arquivos de bancos de dados, a estrutura de dados em softwares, a estrutura sintática e semântica da linguagem natural, a representação do conhecimento em inteligência artificial (Bräscher, 2014).

Os SOC na sua diversidade de formas têm em comum auxiliarem a organização do conhecimento e da informação na recuperação de informação legível a agentes humanos e máquinas. No ambiente web os SOC buscam criar semântica via recursos informacionais, contudo, a estrutura da internet e as tecnologias digitais alteraram drasticamente o cenário de

uso dos SOC nas bibliotecas físicas e arquivos físicos para o uso nos bancos de dados, bases de conhecimento e metadados no contexto digital. Esse fato está provocando a necessidade de compreender como é possível criar modelos de representação do conhecimento que sejam eficientes para esses sistemas digitais e na web. Dessa forma, a transformação digital alterou os modos de produção, coleta, compartilhamento de dados e de informação, conseqüentemente, modificou a conjuntura da organização do conhecimento.

O maior obstáculo para uso de sistemas semânticos tem sido a complexidade para a aquisição de conhecimento, pois a capacidade que possui os seres humanos de reutilizar conhecimento prévio não é uma habilidade dos computadores (Sowa, 2011). Dessa forma, de acordo a Sowa (2011, tradução nossa), como exemplo, a “tarefa de traduzir informações de um livro didático para uma forma semântica computável requer as habilidades combinadas de um linguista, lógica, cientista da computação e especialista no assunto”. E apresenta-se como grande desafio para projetar ferramentas automatizadas que combinem as diferentes habilidades de diversos especialistas.

4.1 LINGUAGEM NATURAL E ARTIFICIAL NA ORGANIZAÇÃO DO CONHECIMENTO

A linguagem é fundamental para a organização do conhecimento no contexto da Ciência da Informação, pois através da linguagem, é possível analisar, descrever e representar informações de forma estruturada e organizada. Na área da organização do conhecimento, a linguagem é utilizada para criar sistemas de classificação, indexação e recuperação de informações. Por exemplo, as classificações bibliográficas como a Classificação Decimal de Dewey ou a Classificação Decimal Universal utilizam a linguagem para classificar os documentos em categorias temáticas. A linguagem também é usada na construção de taxonomias, vocabulários controlados, tesouros e ontologias. Do mesmo modo, a linguagem é fundamental para a construção de metadados para a catalogação de informação. Segundo Petter (2002), todas as formas de linguagem, incluindo as verbais e as não-verbais, são sistemas de signos usados para a comunicação, essa característica comum possibilitou a criação da ciência que estuda todo e qualquer sistema de signos, denominada Semiologia para Saussure ou Semiótica para Pierce. “A Linguística é, portanto, uma parte dessa ciência geral; estuda a principal modalidade dos sistemas sígnicos, as línguas naturais, que são a forma de comunicação mais altamente desenvolvida e de maior uso” (Petter, 2002, p. 15).

A linguística moderna analisou o signo linguístico a partir da estrutura da língua, de modo que, a língua foi analisada como um arranjo sistemático, um sistema de signos de uma estrutura e parte de uma convenção social (comunidade linguística). “A Linguística, que faz parte da Semiótica, estuda a principal modalidade dos sistemas signicos, a das línguas naturais” (Lopes, 1987, p. 17).

A língua é para Saussure "um sistema de signos" - um conjunto de unidades que se relacionam organizadamente dentro de um todo. É "a parte social da linguagem", exterior ao indivíduo; não pode ser modificada pelo falante e obedece às leis do contrato social estabelecido pelos membros da comunidade (Petter, 2002, p. 10).

A linguística estruturalista define o conceito de língua como um sistema semiótico simbólico, baseado no convencionalismo do signo, portanto, a língua é um conjunto de símbolos convencionados socialmente e considera que a relação entre o signo e seu referente é arbitrário e não natural. Ainda, a língua é particular e variável, pois cada língua tem suas próprias características e variações regionais, sociais e históricas; é um sistema de signos, por ser composta por um conjunto de elementos como fonemas, morfemas, palavras, frases e discursos. Por fim, a língua é um arranjo sistemático de partes, visto que, é organizada de acordo com princípios e regras específicas que determinam como os elementos podem ser combinados e usados para expressar ideias e significados. Essa organização sistemática é o que possibilita a comunicação entre os falantes de uma língua. “A língua é considerada uma estrutura constituída por uma rede de elementos, em que cada elemento tem um valor funcional determinado” (Petter, 2002, p. 10).

De acordo a Fiorin (2002) as palavras formam um sistema autônomo que é independente do que elas nomeiam, de modo que cada língua consegue categorizar o mundo diversamente. Como, por exemplo, o inglês e o português, têm palavras diferentes para expressar a mesma realidade, atuando como recortes diferentes da mesma realidade. Os signos definem-se uns em relação aos outros, um signo delimita o outro e o valor de um signo é dado por outro signo.

Além disso, um signo é sempre interpretável por outro signo: no interior do mesmo sistema pelos sinônimos, pelas paráfrases, pelas definições; em outro sistema, em outra língua, por exemplo, pela tradução. A dificuldade de traduzir indica que não há univocidade na relação entre os nomes e as coisas (Fiorin, 2002, p. 70).

A expressão *aliquid pro aliquo* (alguma coisa em lugar de outra), era usada no período medieval para referir-se aos signos, essa definição expressa que o signo não substitui o objeto

da realidade, mas o representa (Fiorin, 2002). “Saussure vai precisar bem esse fato, quando diz que o signo lingüístico não une um nome a uma coisa, mas um conceito a uma imagem acústica” (Fiorin, 2002, p. 71).

A imagem acústica /gatu/ não evoca um gato particular, mas a idéia geral de gato, que tem um valor classificatório. Na criação desse conceito, a língua não leva em conta as diferentes raças, os tamanhos diversos, as cores várias etc. Faz abstração das características particulares de cada gato, para instaurar a categoria da /felinidade/ (Fiorin, 2002, p. 72).

Peirce apresentou o conceito de ícone na Semiologia, onde há uma semelhança visual entre o sinal e o significado. Segundo Lopes (1987), as fotografias, cópias, impressões digitais e outros tipos de imagens apresentam uma relação necessária entre a parte que expressa o conteúdo (o significante) e o conteúdo expressado (o significado). O autor considera a imagem ou ícone são sinais não-sígnicos, afirmando que “a relação entre o significante e o significado do ícone, não se dá, nele, nenhum tipo de semiose, pois inexistente aí a convenção, produto da intencionalidade comunicativa dos homens” (Lopes, 1987, p. 46). Todavia, Fiorin (2015) afirma que algumas linguagens, como as linguagens visuais, podem apresentar signos onde a relação entre significante e significado tem iconicidade, em suma, existe uma correspondência e conexão direta, a relação é motivada. Por essa razão, a arbitrariedade do signo não se aplica a todas as linguagens. “A foto de uma paisagem é um signo em que significado e significante estão unidos por semelhança” (Fiorin, 2002, p. 77). Para esta pesquisa, as imagens são um tipo de signo visual que possuem significação e podem ser utilizadas para transmitir mensagens; a semiótica é a ciência que estuda os signos em geral, e o signo visual é um tipo específico de signo. Em resumo, a conexão entre significação e signo visual, está presente nas convenções da comunicação visual. De fato, os signos visuais são uma forma de representação que pode transmitir significados, mesmo que de maneira diversa da linguagem verbal.

Segundo Lopes (1987), a semiótica é o estudo dos signos e da comunicação usando signos e trata os processos de significação (criação, interpretação e comunicação) em diversos tipos de representação verbais, visuais, sonoros, entre outras; se dedica aos sistemas sígnicos que representam a realidade cultural, tendo como foco compreender como diferentes sistemas linguísticos criam e recriam a realidade fenomênica, sem se preocupar em estudar a realidade natural. Ainda segundo o autor, os signos são suportes exteriores e materiais da comunicação entre as pessoas e, ao mesmo tempo, são meios para exprimir a relação entre o homem e o mundo, desse modo, esse sistema atribui às linguagens a função de sistemas modelizantes. É a

“ciência que estuda os sistemas de signos, quaisquer que eles sejam e quaisquer que sejam as suas esferas de utilização, chama-se Semiologia ou Semiótica” (Lopes, 1987, p. 15).

A primeira fase da teoria peirceana da informação (entre os anos de 1865 e de 1867) é restrita a mensagens verbais e o enfoque está ligado à lógica proposicional tradicional, já a segunda fase da teoria peirceana da informação (depois de 1900) engloba os aspectos pragmáticos, cognitivos e semióticos da informação, nos quais a informação está presente nos contextos verbal e não-verbal, ademais, a questão de significado está atrelada a comunicação (Nöth; Gurick, 2011). A primeira teoria da informação desenvolvida por Peirce se baseia na diferença lógica e semântica entre os termos denotação e significação. Denotação, referência, extensão e largura são conceitos que descrevem objetos empíricos aos quais os símbolos se aplicam, enquanto significação, sentido, intenção, compreensão e profundidade se referem ao conhecimento verbal associado aos símbolos (Nöth; Gurick, 2011).

Em suma, a denotação de um termo é o conjunto de objetos que ele pode representar ou referir na realidade. Por exemplo, o termo "gato" denota todos os gatos que existem na realidade. Em contrapartida, a significação de um termo é o conjunto de atributos ou predicados que podem ser atribuídos a ele na forma de uma proposição verdadeira. Por exemplo, a proposição "O gato é um animal de estimação" atribui o predicado "animal de estimação" ao sujeito "gato". São conceitos fundamentais na lógica que explicam como as palavras são usadas para descrever a realidade e formar proposições verdadeiras ou falsas. Além disso, a distinção entre denotação e significação é importante para a compreensão da semântica das linguagens naturais e artificiais.

A quantidade relativa de denotação e significação de um termo pode ser determinada por sua extensão e predicados atribuídos, respectivamente. “A quantidade de significação contida num termo cresce com o número de predicados atribuídos a este termo, enquanto a quantidade de denotação de um termo (sua extensão) cresce com o número de coisas reais que o termo representa” (Nöth; Gurick, 2011, p. 8). Por tais motivos, o símbolo "cachorro" tem uma extensão ampla por denotar a classe de todos os cachorros, mas sua significação é limitada; em contraste, um símbolo que denota um cachorro específico tem uma significação com muitas características, mas sua extensão é mais limitada. Assim, quanto mais significado um símbolo contém, menor será sua extensão e vice-versa.

Charles Sanders Peirce e Charles Morris propuseram uma descrição dos sistemas de signos baseada em três pontos de vista, sintaxe, semântica e pragmática, descrevendo os níveis de semiose (Lopes, 1987, p. 17):

- a) É o estudo da função sintática e envolve a análise das relações entre os diferentes signos, ou seja, as relações que um signo qualquer mantém para com os demais signos pertencentes ao mesmo enunciado. “Seria o estudo da função sintática”.
- b) As relações de um signo para com o seu objeto, em suma, a conexão do signo para com o seu denotatum. “Seria o estudo da função semântica”.
- c) Parte da perspectiva do signo para com os seus usuários, isto é, a vinculação do signo com o remetente e o destinatário. “Seria o estudo da função pragmática”.

A sintaxe lógica propositadamente se concentra na estrutura lógico-gramatical da linguagem, na sintática da semiose, de modo que, está relacionada ao arranjo das palavras na frase e nas relações lógicas e gramaticais, também se refere ao conjunto de regras de frases e orações em determinada língua. “Não está interessada nas propriedades individuais dos veículos sígnicos ou em quaisquer outras relações deles que não sejam sintáticas, isto é, relações determinadas pelas regras sintáticas” (Morris, 1985, p. 17). O sistema semiótico formado pela sintaxe, semântica e pragmática, corresponde a três níveis da semiose. “Na sua qualidade de níveis, os subcomponentes sintático, semântico e pragmático estão hierarquizados: o nível semântico engloba o nível sintático e é, por sua vez, englobado pelo nível pragmático” (Lopes, 1987, p. 18).

O objeto de estudos da Semiótica é a semiose, ou a forma de significação. Segundo Morris (1985, p. 7), a “semiose é o processo em que algo funciona como um signo”, de tal forma, os três componentes da semiose são, respectivamente, o veículo sígnico (aquilo que atua como um signo), o designatum (aquilo a que o signo se refere) e o interpretante (o efeito sobre um intérprete); o intérprete pode ser incluído como um quarto fator.

Conforme Lopes (2015), é possível realizar a compreensão de um objeto através de seu signo devido a uma série de traços pertinentes na noção desse objeto. Em suma, no conceito de cadeira, o número de pés é variável, mas “possuir pés” é uma característica constante, portanto, é um traço pertinente. De tal maneira, ao designar as características de uma cadeira, existirão traços que não são de fato pertinentes e que podem variar de uma cadeira para outra. Contudo, o conjunto de todos os traços pertinentes que fazem de "cadeira" uma unidade léxica constitui o semema de cadeira. Por conseguinte, cada elemento constante do conjunto de traços pertinentes é o sema. “Quais são os traços distintivos que compõem o "objeto mental" cadeira e fazem de "cadeira" uma unidade léxica?” (Lopes, 1987, p. 265).

Um panorama histórico sobre o estudo do significado pode ser iniciado desde Platão, passando por Aristóteles; no século XIX Breal cunhará o nome semântica ao sugerir o estudo das leis que governam o significado. As definições do termo semântica na literatura indicam

tratar do estudo do significado ou ciência que estuda a significação, mas não existe uma definição consensual para o conceito de significado, com variações que transitam entre abordagens filosóficas do realismo e relativismo (Almeida; Souza, 2011).

A semântica é a ciência das significações das línguas naturais. Na semântica as relações e as diferenças entre termos linguísticos se estruturam em dois âmbitos distintos, a oposição entre essas duas ordens ressalta a natureza de cada uma, sendo ambas formas de atividade mental indispensáveis para a vida da língua (Saussure *et al.*, 2006). A relação sintagmática une os termos por meio de um sistema, uma cadeia, uma combinação, um encadeamento efetivo, em contrapartida, a relação associativa (pragmática) está baseada na mnemônica visual e nos elementos familiares associados da memória.

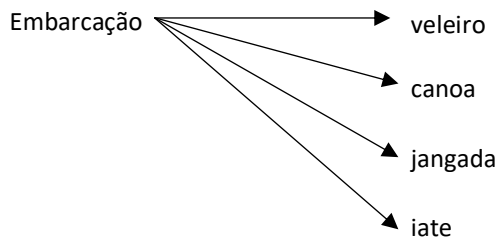
O eixo sintagmático estabelece uma relação entre pelo menos dois termos, onde cada termo somente tem valor se opondo ao termo precedente ou ao seguinte, ou os dois, esse encadeamento de valor, estabelece uma relação linear da língua, impossibilitando que seja possível pronunciar dois elementos ao mesmo tempo, pois “estes se alinham um após outro na cadeia da fala. Tais combinações, que se apoiam na extensão, podem ser chamadas de sintagmas” (Saussure *et al.*, 2006, p. 142).

No eixo paradigmático ou relação associativa, as palavras apresentam associações na memória formando grupos com relações muito diversas, como, por exemplo, “a palavra portuguesa ensino ou ensinamento, as palavras associadas serão ensinar, e depois armamento, desfiguramento, etc., e por fim educação, aprendizagem, etc” (Saussure *et al.*, 2006, p. 143). No exemplo, as coordenações são de uma espécie bem diferente das relações sintagmáticas, não tendo por base a extensão; elas fazem parte da língua de cada indivíduos, são as relações associativas.

Saussure foi pioneiro da semântica estrutural ao definir o signo linguístico através da união do conceito (**significado**) com sua imagem acústica (**significante**) e concluiu que os termos implicados nos signos linguísticos devem ser situados pelas classes da *langue* (paradigmática)/língua e nos enunciados da *parole* (sintagmática)/fala.

Segundo Lopes (2015, p. 259), os campos semânticos “são estruturados a partir de associações mnemônicas cujo ponto de partida é aleatório. Dada uma palavra qualquer, podemos associá-la mentalmente com várias outras, através do mecanismo dos parciais semânticos iguais”, por exemplo, a palavra embarcação (as flechas indicam associação) na figura 15:

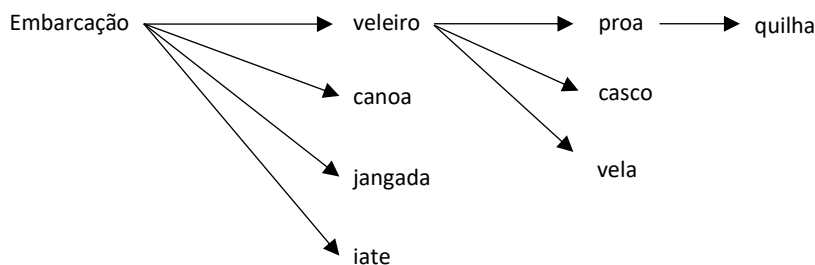
Figura 15 - Paradigma semântico



Fonte: Lopes (2015, p. 259).

As palavras ligadas à palavra embarcação formam um paradigma semântico, ao pertencerem a uma mesma classe de sentidos. O processo que une essas palavras na memória da língua é um processo aberto e dinâmico, sendo possível que qualquer uma das palavras dessa classe paradigmática seja uma nova origem para a criação de outros campos semânticos. O processo continuado aplicando as mesmas relações conjuntivas, permite criar a seguinte série, por exemplo, na figura 16:

Figura 16 - Campo semântico 2



Fonte: Lopes (2015, p. 259).

As duas primeiras colunas verticais são dois campos semânticos diferentes, a primeira coluna designa o campo semântico de embarcação e a segunda, o de partes constituintes dos veleiros. Os elementos da primeira coluna apresentam dois marcadores semânticos, o primeiro é obrigatório, /embarcação/ e o segundo é um marcador das características do elemento a ser descrito. Por exemplo, no quadro 16:

Quadro 16 - Marcadores semânticos

PRIMEIRA COLUNA	1º MARCADOR SEMÂNTICO	2º MARCADOR SEMÂNTICO
veleiro =	/embarcação/ +	/a vela.;
Jangada =	/embarcação/ +	/de paus roliços/

Fonte: Lopes (1987, p. 260).

Os elementos da primeira coluna pertencem à mesma classe paradigmática, porque todos podem designar uma embarcação “e, no tocante a essa designação, a presença de um deles, numa frase, exclui automaticamente a presença de qualquer outro membro da mesma classe; na frase” (Lopes, 1987, p. 260).

Certamente, um dos objetivos principais do trabalho de documentação é a representação do conhecimento presente num documento, viabilizando sua recuperação; através da representação sintética do conhecimento; o processo de indexação reduz os textos a tags de significado, que, geralmente, é uma unidade de caráter terminológico (CABRÉ, 1999). Na descrição documental essas unidades passam do caráter comunicativo ao operacional e consideradas descritores controlados por um tesouro.

O uso de uma terminologia padronizada para indexar documentos confere à documentação uma sistematicidade e uma univocidade que o uso de palavras da língua comum nunca poderia alcançar. Portanto, a documentação se serve da terminologia, e mais especificamente da terminologia padronizada, para descrever ou representar o conteúdo dos documentos (Cabré, 1999, p. 233).

A elaboração de conceitos é realizada empregando a linguagem natural visando definir termos específicos presentes no discurso especializado. O termo Terminologia com maiúscula determina o campo (área) de estudo que trata dos termos utilizados na comunicação técnico-científica.

O objetivo principal da Terminologia é analisar os elementos da língua natural e da comunicação especializada e utilizar termos na representação de conceitos e sistemas de conceitos; auxiliando a comunicação entre especialistas e entre especialistas e o público.

4.2 SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

De acordo com Hjørland (2008), dois aspectos ou significados diferentes de *Knowledge Organization* (KO) devem ser distinguidos: um restrito e um amplo.

No sentido estrito, Organização do Conhecimento (KO) refere-se a atividades como descrição, indexação e classificação de documentos realizadas em bibliotecas, bancos de dados bibliográficos, arquivos e outros tipos de “instituições de memória” por bibliotecários, arquivistas, especialistas em informação, especialistas em assuntos, bem como por algoritmos de computador e leigos (Hjørland, 2008, p. 86, tradução nossa).

Em contrapartida, o entendimento mais amplo, considera que a organização do conhecimento permeia todas as esferas da sociedade, todos os tipos de conhecimento e suas representações. A KO no sentido amplo é uma preocupação de disciplinas como a sociologia e a metafísica (Mazzocchi, 2018). O termo KO é “sobre descrever, representar, arquivar e organizar documentos e representações de documentos, bem como assuntos e conceitos tanto por humanos como por softwares”, portanto, está focado nos aspectos qualitativos dos processos de organização do conhecimento e no desenvolvimento dos SOC (Hjørland, 2008, p. 86, tradução nossa).

Por conseguinte, os processos de organização do conhecimento são a catalogação, análise de assunto, a indexação, a marcação e classificação por humanos ou computadores (Mazzocchi, 2018). Em relação aos processos de organização do conhecimento, algumas facetas aparecem mais desenvolvidas nas regras e manuais de instituições internacionais de catalogação da imagem em movimento, como a catalogação e a classificação. Entretanto, o trabalho que estamos desenvolvendo tem foco nos SOC, ao tratar os processos de indexação para imagens em movimento.

O termo SOC “pretende abranger todos os tipos de esquemas para organizar a informação e promover a gestão do conhecimento” (Hodge, 2000, p. 3, tradução nossa). A autora inclui como tipo de SOC, os esquemas de classificação, cabeçalhos de assunto, arquivos de autoridade, redes semânticas e ontologias. Os SOC são mecanismos centrais em todas as bibliotecas, museus e arquivos.

As características do KO podem ser estendidas para os SOC, que também apresentam aspectos gerais e específicos, no entanto, seu significado está mais relacionado as características restritas e aos ambientes da biblioteconomia e Ciência da informação, pois “o termo SOC é usado principalmente para se referir a itens funcionais projetados para organizar o conhecimento e a informação, e tornar mais fácil seu gerenciamento e recuperação” (Mazzocchi, 2018).

Assim, no contexto mais estrito, o termo SOC é usado principalmente para organizar o conhecimento e a informação, e otimizar a gestão e recuperação.

Os sistemas de organização do conhecimento (SOC) são a seleção de conceitos com uma indicação de relações semânticas selecionadas. Os exemplos são sistemas de classificação, listas de cabeçalhos de assuntos, tesouros, ontologias e outros sistemas de metadados (Hjørland, 2016, tradução nossa).

Um SOC serve como uma ponte entre a necessidade de informação do usuário e o material da coleção, de maneira que, orienta o usuário em um processo de descoberta. Em tal medida, o SOC permita o usuário encontrar um item desconhecido via diferentes formas como a navegação, pesquisa direta, página da web ou na busca de um site (Hodge, 2000). Não existe um único SOC que seja de comum acordo a todas as demandas, organizações e usuários, cada contexto prioriza o que e como organizar, numa enorme variedade de formas organizar o conhecimento. A visão tendenciosa ocidental influência, é claro, também como os SOC e suas estruturas relacionais são desenvolvidos. Como já notado, os princípios da lógica clássica determinam como as classificações e tesouros são construídos (Mazzocchi, 2018, tradução nossa).

Porém, mesmo com uma grande diversidade, os SOC apresentam algumas características em comum que são críticas para seu uso na organização de bibliotecas digitais.

- O SOC impõe uma visão particular do mundo sobre uma coleção e seus itens.
- A mesma entidade pode ser caracterizada de maneiras diferentes, dependendo do SOC que é usado.
- Deve haver semelhança suficiente entre o conceito expresso em um SOC e o objeto do mundo real ao qual esse conceito se refere para que uma pessoa com conhecimento possa aplicar o sistema com confiabilidade razoável. Da mesma forma, uma pessoa que busca material relevante usando um SOC deve ser capaz de conectar seu conceito com sua representação no sistema. (Hodge, 2000, p. 4, tradução nossa).

Os SOC são sistemas que cumprem o importante papel de padronização da terminologia adotada para a organização e recuperação da informação, por delimitar o uso de termos e definir os conceitos e suas relações de acordo ao seu domínio específico, com conhecimento compartilhado e consensual (Bräscher, 2014).

Hodge (2000) agrupou os SOC de acordo com três categorias abordando os SOC também em ambientes digitais, mostrado no quadro 17:

Quadro 17 - Categorias dos SOC

Categorias de SOC	Características gerais das categorias	Tipos específicos de SOC
LISTAS	Sistemas lineares e menos estruturados; ênfase nas listas de termos (frequentemente fornecidas com definições)	Arquivos de autoridade Glossários Dicionários Gazetteers
CLASSIFICAÇÕES E CATEGORIAS	Sistemas estruturados hierarquicamente; ênfase na criação de conjuntos de assuntos	Títulos de assuntos Esquemas de classificação Taxonomias Esquemas de categorização (os últimos três termos são frequentemente usados de forma intercambiável)
LISTAS DE RELACIONAMENTO	Sistemas complexos e altamente estruturados; ênfase nas conexões entre termos e conceitos	Tesauros Redes Semânticas Ontologias

Fonte: Hodge (2000 *apud* Mazzocchi, 2017).

A abordagem de Hodge (2000) sobre as categorias dos SOC foi explorada com mais profundidade no quadro 18:

Quadro 18 - Classificação de Hodge de SOC

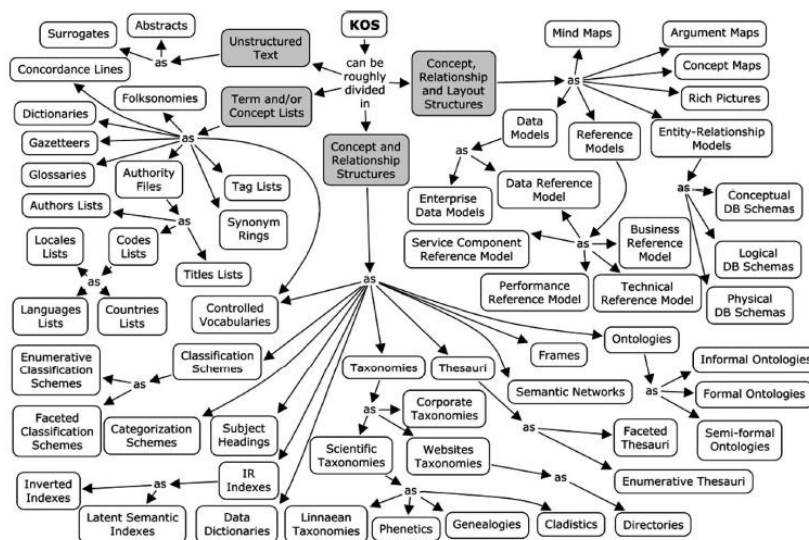
LISTAS DE TERMOS	
Arquivos de autoridade	Arquivos de autoridade são listas de termos usados para controlar os nomes variantes de uma entidade ou o valor de domínio de um campo específico. Os exemplos incluem nomes de países, indivíduos e organizações. Os termos não preferenciais podem ser vinculados às versões preferenciais. Este tipo de SOC geralmente não inclui uma organização profunda ou estrutura complexa.
Glossários	Um glossário é uma lista de termos, geralmente com definições. Os termos podem ser de um campo de assunto específico ou de um trabalho particular. Os termos são definidos dentro de um ambiente específico e raramente incluem significados variantes.
Dicionários	Os dicionários são listas alfabéticas de palavras e suas definições. Sentidos variantes são fornecidos quando aplicável. Os dicionários são mais gerais em escopo do que os glossários. Eles também podem fornecer informações sobre a origem de uma palavra, variantes (por ortografia e morfologia) e múltiplos significados entre as disciplinas. Enquanto um dicionário também pode fornecer sinônimos e através das definições, palavras relacionadas, não há nenhuma estrutura hierárquica explícita ou tentativa de agrupá-los por conceito.
CLASSIFICAÇÕES E CATEGORIAS	
Cabeçalho de Assuntos	Esse tipo de esquema fornece um conjunto de termos controlados para representar os assuntos dos itens em uma coleção. As listas de títulos de assuntos podem ser extensas e abranger uma ampla gama de assuntos; no entanto, a estrutura da lista de títulos de assunto é geralmente muito rasa, com uma estrutura hierárquica limitada. Em uso, os cabeçalhos de assunto tendem a ser coordenados, com regras de como eles podem ser unidos para fornecer conceitos mais específicos.
Esquemas de Classificação, Taxonomias e Esquemas de Categorização	Esses termos são frequentemente usados de forma intercambiável. Embora possa haver diferenças sutis de exemplo para exemplo, todos esses tipos de SOC fornecem maneiras de separar entidades em "baldes" ou níveis de tópicos amplos. Alguns exemplos fornecem um arranjo hierárquico de notação numérica ou alfabética para representar tópicos amplos. Esses tipos de SOC podem não seguir as regras de hierarquia exigidas no ANSI NISO Thesaurus Standard (Z39.19) (NISO 1998) e carecem dos relacionamentos explícitos apresentados em um tesouro. As categorias de assunto são frequentemente usadas para agrupar termos do dicionário de sinônimos em conjuntos de tópicos amplos que estão fora do esquema hierárquico do dicionário de sinônimos.
LISTAS DE RELACIONAMENTO	
Tesouro	Os tesouros são baseados em conceitos e mostram as relações entre os termos. Relacionamentos comumente expressos em um dicionário de sinônimos incluem hierarquia, equivalência (sinonímia) e associação ou parentesco. Esses relacionamentos são geralmente representados pela notação BT (termo mais amplo), NT (termo mais restrito), SY (sinônimo) e RT (termo associativo ou relacionado).
Redes Semânticas	Com o advento do processamento de linguagem natural, houve desenvolvimentos significativos em redes semânticas. Esses SOC estruturam conceitos e termos não como hierarquias, mas como uma rede ou uma teia. Os conceitos são pensados como nós e os relacionamentos se ramificam a partir deles. Os relacionamentos geralmente vão além do padrão BT, NT e RT. Eles podem incluir relações específicas de todo-parte, causa-efeito ou pai-filho. A rede semântica mais conhecida é a Word Net da Universidade de Princeton, que agora é usada em uma variedade de mecanismos de busca.
Ontologias	A comunidade de gestão do conhecimento está desenvolvendo ontologias como modelos conceituais específicos. Eles podem representar relacionamentos complexos entre objetos e incluir as regras e axiomas que faltam nas redes semânticas. Ontologias que descrevem o conhecimento em uma área específica são muitas vezes conectadas com sistemas de mineração de dados e gestão do conhecimento.

Fonte: Hodge (2000, tradução nossa).

Souza, Tudhope, Almeida (2012, tradução nossa) definiram as principais diferenças das taxonomias SOC propostas por Hodge (2000) e Wright (2008) num mapa conceitual (fig. 15).

Em vista disso, os autores incluíram nos tipos de SOC os resumos, linhas de Concordância e índices de recuperação da informação, pois segundo o autor “i) são usados para organização do conhecimento e recuperação da informação; ii) promovem a gestão do conhecimento; iii) são estruturas de representação do conhecimento baseadas na terminologia”. Foram excluídos os formatos padrão (como HTML, SGML, etc.) e produtos como WordNet, pois os primeiros são ferramentas de representação de SOC e o segundo é um banco de dados lexical.

Figura 17 - Um conjunto provisório de tipos de SOC



Fonte: Souza, Tudhope e Almeida (2012).

A elaboração do SOC envolve a análise de conceitos e respectivas características para o estabelecimento da posição que cada conceito ocupa em determinado domínio, bem como de suas relações com os demais conceitos que compõem um sistema conceitual (Bräscher, 2014, p. 175). Os SOC podem ser vistos como ferramentas semânticas, ou seja, conjuntos de termos/conceitos com relacionamentos entre si, conseqüentemente, proveem uma seleção de termos/conceitos e as relações entre eles.

O campo de KO em SI está, portanto, preocupado com a construção, uso e avaliação de ferramentas semânticas para RI. Esse insight traz a semântica para a vanguarda do SI [...]. Quais abordagens foram utilizadas no campo do KO ao longo de sua história? Como eles se relacionam com a teoria semântica? [...] Dado que SOC são essencialmente ferramentas semânticas, abordagens diferentes para KO devem refletir diferentes abordagens para semântica? (Hjørland, 2007, tradução nossa).

Os SOC “são representações de domínios de conhecimento que definem o significado dos termos no contexto desses domínios, estabelecem relações conceituais que auxiliam no

posicionamento de um conceito no sistema conceitual e são utilizados como instrumentos de organização e recuperação da informação” (Bräscher, 2014, p. 175). Na elaboração de SOC, o principal objeto de estudo são os conceitos e não a expressão utilizada para se referir a eles (termos, códigos, fórmulas ou outros símbolos que os representam) (Bräscher, 2014, p. 176).

De acordo com Souza, Tudhope e Almeida (2012), os SOC têm recebido atenção especial, principalmente no contexto da Web Semântica, devido à necessidade de desambiguação do vocabulário e à exigência de estruturas altamente formalizadas para permitir a compreensão semântica por parte das máquinas. O W3C promove a publicação desses sistemas na Web Semântica, visando facilitar a interoperabilidade semântica, a recuperação de informações e o acesso a recursos de informação. Além das linguagens e esquemas gerais utilizados para representação de conhecimento na web, como RDF e OWL, existem padrões específicos, como o SKOS (W3C, 2009), que procuram preencher a lacuna entre os SOC utilizados na Biblioteconomia e estruturas formais baseadas em lógica, permitindo a expressão de tesouros e esquemas no ambiente da Web Semântica.

4.2.1 Teoria do conceito e elementos da lógica

A autoridade para cunhar conceitos e apoio terminológico, são problemas apresentados na organização do conhecimento de imagens em movimento, pelas dificuldades específicas do contexto de análise das imagens. Assim, as ciências sociais no desenvolvimento das primeiras-base terminológicas em meados dos anos 70, apesar de especificidades diferentes, também apresentaram desafios terminológicos, no entanto, devido ao uso da linguagem natural. Por tais motivos, a Teoria do conceito, desenvolvida pela cientista Ingetraut Dahlberg, surgiu como uma teoria que converteu a condição do conceito em unidade de conhecimento, tanto para conceitos empíricos, como para os conceitos abstratos, presentes na codificação na linguagem natural em artificial. Por causa disso, a Teoria do conceito pode auxiliar as dificuldades acerca dos conceitos empíricos e abstratos das imagens em movimento, através dos processos de significação do triângulo do conceito: na denotação e predição de referentes e na designação das características do referente em formas verbais.

O artigo “*A Referent-Oriented, Analytical Concept Theory for INTERCONCEPT*” foi desenvolvido por Dahlberg no contexto do projeto INTERCONCEPT e apresentado no “*COCTA Panel 3.3, 9th International Conference of Sociology*” em Uppsala no ano de 1978. Segundo a autora, o ponto de partida do *Committee on Conceptual and Terminological Analysis* (COCTA) foram as “exigências dos cientistas sociais como usuários de conceitos, à luz dos

arcabouços analíticos contemporâneos fornecidos pela lógica e pela filosofia da ciência” (Dahlberg, 1978, p. 142, tradução nossa). O projeto INTERCONCEPT da UNESCO tinha como demanda estabelecer normas para o Programa de Informação Geral e principalmente para implementar o Programa de Informação em Ciências Sociais. Nessa perspectiva, o foco do estudo de Dahlberg (1978) apresentou soluções para os problemas em aproximar o COCTA e o INTERCONCEPT, banco de terminologia proposto e patrocinado pela UNESCO para cientistas sociais. Pois com essa conexão seria possível criar condições para a formação de uma rede de acadêmicos, organizados principalmente por áreas temáticas. Assim, a autora delimitou os principais problemas meta conceituais entre o serviço terminológico INTERCONCEPT e as orientações do usuário de ciências sociais do COCTA.

De tal modo, Dahlberg (1978) destacou que o INTERCONCEPT apresentava limitações em relação ao trabalho terminológico patrocinado pela ISO/TC 37, que tinha um viés exclusivo para as necessidades da tecnologia e das ciências naturais. Visto que, as metodologias de "padronização" ou "normalização" eram marcadas por diferenças entre as ciências naturais e as ciências humanas, porque as ciências naturais detinham a autoridade de cunhar novos conceitos. O que era justificado pela forma empírica das novas descobertas das ciências naturais e pelo motivo de que era possível a identificação dos cientistas que cunhavam esses novos conceitos. Por outro lado, a “capacidade das ciências naturais de cunhar neologismos para novos conceitos fornece um motivo e uma possibilidade de normalização não disponível aos cientistas sociais”, o uso de termos derivados da linguagem natural pelos cientistas sociais resultou em uma proliferação extrema dos significados das palavras mais empregadas, produzindo assim “uma confusão polissemântica que parece desafiar todos os esforços normalizadores” (Dahlberg, 1978, p. 142, tradução nossa). Assim, segundo a autora, para construir uma ponte entre as demandas dos cientistas sociais, representados pelo COCTA, e o INTECONCEPT era preciso definir o significado de 'conceito', 'termo', 'definição', 'intensão', 'extensão', 'referente', ou seja, os meta conceitos do trabalho conceitual e terminológico. Igualmente, propôs uma sugestão de modelo de análise da natureza e da estrutura dos conceitos para apoiar os procedimentos em relação ao COCTA e o INTERCONCEPT no processo de construção dos trabalhos.

Ingetraut Dahlberg, um cientista da informação e filósofa e, claro, a fundadora da ISKO, estabeleceu as bases para que a KO fosse reconhecida como um campo ou ciência distinta, que se preocupa com a ordem do conhecimento (ou seja, uma metaciência) (Mazzocchi, 2017, tradução nossa).

Dessa forma, Dahlberg inicia o artigo com a seguinte pergunta: o que são conceitos? Para responder à autora propõem que se considere um conceito como uma "unidade de conhecimento", enfatizando a capacidade de compreensão comum do conhecimento. A partir disso, adiciona a dimensão das ideias no processo de validação do conhecimento e problematiza a **condição de verdade do conhecimento** por meio da cunhagem de conceitos, como sendo um aspecto específico do ambiente controlado e empírico das ciências naturais. Visto que, assim como as proposições e afirmações verdadeiras são condições para ocorrer o conhecimento, seja a partir de documentos ou processos individuais de cabeça de cada pessoa, o mesmo ocorre nas afirmações verdadeiras e proposições científicas consideradas verdadeiras. Em seguida, a autora evoca o conceito de *Begründungszusammenhang*, onde foi proposto considerar a ciência como o conjunto de tais proposições sobre uma área de estudo reunidas em uma relação fundacional. Assim, se as ciências são construídas em proposições e estas por componentes formados por unidades de conhecimento, então tais unidades podem ser verificadas de forma científica.

Assumindo que o homem tem a capacidade de fazer afirmações corretas tanto sobre coisas reais (itens empíricos) quanto sobre idéias existentes apenas em seu cérebro, podemos estabelecer um "modelo para construção de conceitos" como mostrado na Fig. 112. (Dahlberg, 1978, p. 143, tradução nossa).

O modelo para construção de conceitos de Dahlberg (fig.18), começa pelo item, que se apresenta como o referente, podendo ser um objeto ou conjunto de objetos num universo de itens existentes. Em seguida são verificadas as afirmações corretas sobre esse referente por meio de evidências ou por acordo intersubjetivo, se caso forem consideradas afirmações verdadeiras, estas são sintetizadas numa forma verbal, como um termo ou um nome. Esse termo ou nome passa a comunicar o conceito no universo do discurso.

Figura 18 - Modelo para construção de conceitos de Dahlbe

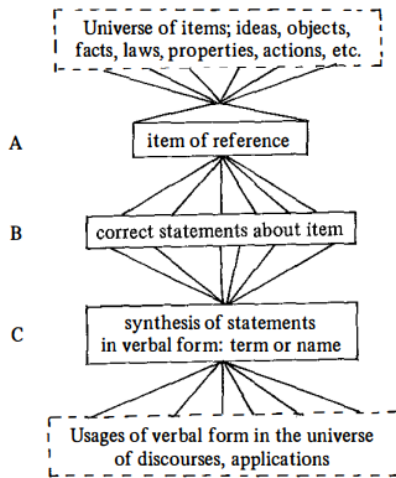


Fig. 1: Model for concept construction

Fonte: Dahlberg (1978, p. 143).

O modelo de construção de conceitos demonstra como é gerado um conceito que ainda não existe, sendo considerados componentes imprescindíveis para a formação do conceito: seu referente (A), julgamentos sobre o referente (B) e a forma verbal (C) a qual é a representação do conceito. Assim, “cada afirmação correta sobre A produz um elemento de conhecimento sobre A e a soma de afirmações corretas sobre A fornece a unidade de conhecimento sobre A” (Dahlberg, 1978, p. 143, tradução nossa). De tal maneira, o universo de itens e o uso da forma verbal no universo do discurso, são processos abertos das etapas de construção do conceito, que apenas se fecha no momento que o item de referência é selecionado. Visto que, as declarações corretas sobre o referente (B) e a sintetização do conceito em termo (C), são estabelecidos pelo item de referência (A).

De tal modo, a autora define que “um conceito é uma unidade de conhecimento, compreendendo declarações verificáveis sobre um item de referência selecionado, representado verbalmente”. Do mesmo modo, cada um dos três componentes (A), (B) e (C) da unidade de conhecimento são definidos como: o componente (A) é o referente, sendo o desígnio das declarações verificáveis e (B) é o conjunto de declarações verificáveis, ou seja, os atributos do item de referência. Por conseguinte, (C) é o componente que resume ou sintetiza e representa um conceito no processo de comunicação na forma verbal. Importante ressaltar que as declarações (B) são as características de um conceito e os atributos do item de referência, como, por exemplo, uma propriedade, um estado, uma dimensão ou similar, portanto, se apresenta como a soma das características de um conceito e o conteúdo de um conceito. “Uma característica é o componente de um conceito derivado de um enunciado sobre seu referente,

ou (sobre enunciados como conteúdos)”. Logo, a definição reformulada de conceito se estabelece a seguir:

Um conceito é uma unidade de conhecimento que compreende as características de um referente por um termo ou um nome.

Quanto aos passos envolvidos na construção de um conceito, podemos falar deles como

o passo referencial (A)

o passo predicacional (B)

o passo representacional (C).

A partir dessas conclusões, Dahlberg estruturou o **triângulo do conceito** (fig.19) e as etapas ordenadas de cada componente, inserindo o referente (A) no topo do triângulo e a representação das características (B) no canto inferior esquerdo, simbolizando a prioridade dos enunciados na conceituação. Por fim, introduziu o passo representacional na forma verbal (C) no canto inferior direito, como a última parte a ser determinada.

Figura 19 - Triângulo do conceito

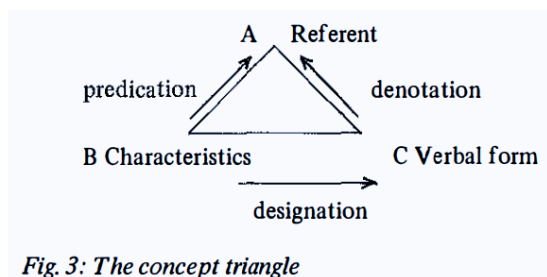


Fig. 3: The concept triangle

Fonte: Dahlberg (1978).

Na teoria do triângulo do conceito, a divisão entre 'conceitos teóricos' e 'conceitos empíricos', é desnecessária, pois não teria efeito na classificação de conceitos em categorias mutuamente exclusivas, visto que, diversos conceitos são tanto empíricos quanto teóricos. Assim, a categorização adequada de conceitos depende de uma categorização de referentes:

Um 'unicórnio' não tem contrapartida empírica no mundo animal de hoje; no entanto, pode muito bem ser considerado como um conceito de um objeto imaterial, ou seja, de um no mundo das peças de teatro, contos de fadas ou fábulas existentes. Do nosso ponto de vista, esta distinção é desnecessária, não ajuda a classificar ou classificar conceitos em categorias mutuamente exclusivas. Uma categorização adequada de conceitos pode seguir uma categorização de referentes. E aqui podemos perceber que já existem várias propostas por parte da classificação da informação (Dahlberg, 1978, p. 144, tradução nossa).

Dessa forma, a autora indica a proposta de Teune (fig.20), que tem influência das categorias sistematizadas por Aristóteles, para classificar itens de referência em categorias e subcategorias, esquema que fornece uma base sistemática de análise de conceitos, abrangendo qualquer campo de assunto.

Figura 20 - Referentes conceituais

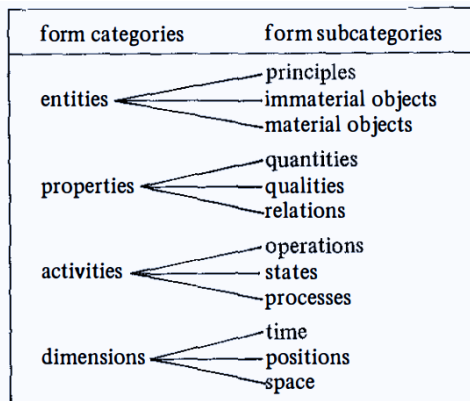
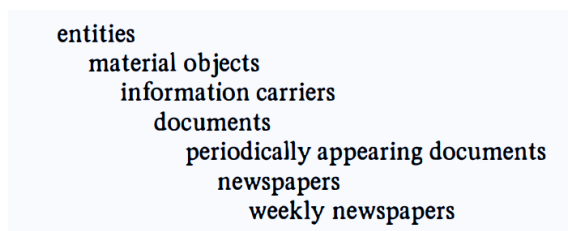


Fig. 4: Concept referents

Fonte: Dahlberg (1978).

A análise dos referentes gera hierarquias na categorização, auxiliando na identificação dos conteúdos e da estruturação do conceito. A predicação dos atributos do referente forma uma hierarquia de características na forma de escada de características que também são conceitos. De forma que, se estabelece classes para as características de um conceito (Fig.21).

Figura 21 - Hierarquia de características na forma de escada



Fonte: Dahlberg (1978).

A figura mostra que o próximo nível na hierarquia de características implica que uma afirmação verdadeira possa ser feita sobre os itens encontrados em níveis inferiores da sequência. “Se um assim chamado 'conceito mais amplo' não vale para todas as suas espécies e indivíduos mais restritos, então a hierarquia pode ser considerada defeituosa” (Dahlberg, 1978, p. 145, tradução nossa). Assim, todos os conceitos abaixo de “*entities*” são entidades, todos os conceitos abaixo de “*material objects*” são objetos materiais e assim sucessivamente até o

conceito mais estrito. Em vista disso, nessa perspectiva ocorrem 4 categorias e suas subcategorias de conceitos derivadas das predicções de um referente e uma característica é apenas um elemento do conceito. Além do mais, existem ao todo três tipos gerais de predicções que produzem três tipos de características apresentadas no quadro 19:

Quadro 19 - Três tipos gerais de predicções e três tipos de características

<i>Tipo de Predição</i>	<i>Exemplo</i>	<i>Tipo de Característica</i>	<i>Definição</i>
Todos os referentes de um determinado tipo.	Todos os 'documentos' são portadores de informação.	Características essenciais	Características que se aplicam a todo referente de um dado tipo.
Alguns dos referentes de um determinado tipo.	Alguns 'documentos' pode-se dizer que eles têm uma periodicidade.	Características acidentais	Características que se aplicam apenas a alguns referentes de um determinado tipo
Apenas um referente em questão.	Jornal publicado em hamburgo, chamado <i>Die Zeit</i> .	Características de individualização	Características aplicáveis a um único referente.

Fonte: Dahlberg (1978).

Além da influência do referente na elaboração dos enunciados sobre um conceito, é possível distinguir entre as características que constituem um conceito, como as características necessárias, possíveis, implícitas e as individuais, assim como no quadro 20:

Quadro 20 - Características que constituem um conceito

Características de categoria de forma
Aqueles referentes a categorias de forma e subcategorias de forma.
Características ontológicas
As que se referem a categorias do ser, como, por exemplo, "ser um objeto material", "ser um ser vivo", "ser um ser humano".
Características constituintes do conceito (essenciais, acidentais e individualizantes)
Aquelas que se referem a todos, ou a alguns casos ou a um único caso de um determinado item de referência.
Características consecutivas ou implícitas
As decorrentes das características essenciais e acidentais por implicação e possível diversidade.

Fonte: Dahlberg (1978).

Como apresentado no triângulo do conceito, o terceiro componente de cada conceito é sua forma verbal: o termo ou nome, a autora destaca que 'termo' era usado para representar tanto o nome de um conceito e o próprio conceito. Portanto, a reorientação de um conceito, que distingue a forma verbal (termo) do conhecimento que ele designa, é apresentada como uma prática recente para a época de publicação do artigo de Dahlberg. E usada para a construção de novos conceitos, que exigia a criação de uma forma verbal para acompanhar o novo conceito.

De modo geral, é inevitável mais de uma forma de verbalização de um conceito, mas existem princípios para normalizar a linguagem e facilitar a comunicação:

Princípio 1: Conformidade com o referente

A forma verbal de um conceito deve obedecer ao referente de um conceito.

- a) Se o referente de um conceito é de natureza tão geral, então o conceito também terá um grau de generalidade, que representa uma categoria de forma ou uma subcategoria.
- b) Se o referente é uma especificação de algum grau e representa um subconjunto. Exemplo: termo geral: 'desenvolvimento', termos especiais: 'desenvolvimento social', 'desenvolvimento urbano', 'desenvolvimento infantil'.
- c) Se o referente for um indivíduo, então a forma verbal deve refletir isso através do nome próprio desse indivíduo

Princípio 2: Reflexão de características

Uma forma verbal de um conceito deve refletir a natureza de um conceito, ou seja, suas características necessárias ou características que constituem o conceito.

Princípio 3: Duração mínima do termo

Uma vez que a forma verbal de um conceito entrará no discurso e será usada quantas vezes for necessário, segue-se que a duração de tal forma deve ser reduzida ao mínimo.

Princípio 4: Derivabilidade verbal

Selecionar forma verbal em um determinado caso que possua potenciais derivacionais, como por exemplo, metal em 'metálico', 'metalífero', 'metalino', 'metalista', 'metalar', etc.

Princípio 5: Internacionalidade

Por razões de compreensão internacional é aconselhável criar termos que contenham elementos latinos ou gregos.

Os metaconceitos relativos a tipos de termos:

- **Monosemas**: sempre que um único termo representa um único conceito,
- **Polissemas**: sempre que um único termo representa vários conceitos não tão distantes entre si;
- **Homônimos**: são termos que têm a mesma forma, mas que representam conceitos bem diferentes;
- **Sinônimos**: são termos diferentes para um mesmo conceito;
- **Quase- ou quase-sinônimos**: são termos que representam conceitos bastante semelhantes.
- **Hiperônimos**: são termos que se referem a um conceito mais geral do que **hipônimos** que são termos que se referem a conceitos específicos de um conceito geral.

Partindo da definição de conceito como o conjunto de suas características, e de que as relações de um conceito com outros conceitos, partem de seu conteúdo, foi definido que “se dois ou mais conceitos têm pelo menos uma característica em comum, então claramente deve existir uma relação entre esses conceitos” (Dahlberg, 1978, p. 148, tradução nossa).

Em tais termos, pelo menos quatro tipos de conceito podem ser relacionados as relações quantitativas, que medem a quantidade e semelhança de características em um conceito:

- **Inclusão de conceito:** todas as características de um conceito estão contidas no maior número de características de outro conceito
- **Interseção de conceito:** as características de dois conceitos se sobrepõem
- **Disjunção de conceito:** as características de dois conceitos não têm nada em comum.

As relações qualitativas podem ser subdivididas formalmente de acordo com os referentes de conceito, podendo encontrar também quatro tipos de relações baseadas em características, nomeadamente:

- **Relação hierárquica** = a relação entre gênero-espécie, espécie-espécie e espécie-indivíduo
- **Relação de partição** = a relação entre um todo e suas partes, entre as partes e entre partes e sub'partes.
- **Relação de oposição** = a relação de contradição, contrariedade e PIN (positivo·neutro-indiferente).
- **Relação funcional** = a relação entre os componentes de uma afirmação/proposição, dependendo das valências semânticas de um conceito relacionado à atividade (por exemplo, 'produzir' exige que um produtor, um consumidor, um produto, etc. estejam envolvidos).

Além das características e relações de um conceito, existem também a 'intensão' e 'extensão' de um conceito:

(13) A intenção de um conceito é a soma total de suas características, que é a mesma que a soma total de conceitos de suas hierarquias de conceitos mais sua(s) característica(s) específica(s).

(14) A extensão de um conceito é a soma total de seus conceitos especiais e seus conceitos individuais, ou seja, o conjunto de conceitos para os quais a intenção de um conceito é verdadeira

A intenção de um conceito inclui todos os seus conceitos mais amplos (de acordo com a relação hierárquica, bem como de acordo com sua relação qualitativa formal) enquanto sua extensão cobre todos os seus mais estreitos (Dahlberg, 1978, p. 1148).

As seguintes relações baseadas na posse de características comuns são logicamente possíveis (Dahlberg, 1978, p. 104), portanto as relações lógicas são:

identidade A (x, x, x) B (x, x, x) As características são as mesmas;

implicação A (x, x) B (x, x, x) O conceito A está contido no conceito B;

intersecção A (x, x, o) B (x, o, o) Os dois conceitos coincidem algum elemento;

disjunção	$A(x, x, x) \vee B(o, o, o)$	Os conceitos se excluem mutuamente. Nenhuma característica em comum;
negação	$A(x, x, o) \wedge B(o, x, o)$	O conceito A inclui uma característica cuja negação se encontra em B.

A palavra Lógica é usada em muitos contextos, segundo Copi (1978, p. 19 *apud* Kobashi, 2018), é o estudo dos métodos e princípios que distinguem o raciocínio correto do incorreto. A Lógica possui diferentes definições, como: “a) no uso comum, cotidiano, costumamos opor lógico a ilógico. b) uso técnico-científico (Kobashi, 2018, p. 16).

Segundo Kobashi (2018), no contexto de indexação e representação por meio de instrumentos de representação da informação, os princípios da lógica possibilitam que as atividades de análise e seleção de conteúdos sejam metódicas e consistentes. De tal forma, conhecer o percurso histórico ajuda a compreender o uso atual e a aplicabilidade da lógica à Organização e Representação da Informação. “A Lógica é uma disciplina que fornece os instrumentos para raciocinar e trabalhar com rigor e método. Por essa razão, ela é aplicada praticamente em todos os campos do conhecimento” (Kobashi, 2018, p. 16).

De acordo a Tugendhat e Wolf (1996) é possível classificar a história da lógica em três períodos:

- 1) **Lógica antiga:** abrange desde o fundador Aristóteles até o fim da Idade Média.
- 2) **Lógica moderna:** tem início com a Lógica de Port-Royal (1662) e foi um período caracterizado por questões ligadas à teoria do conhecimento e à psicologia.
- 3) **Lógica atual:** período que abrange a contemporaneidade, começa com as abordagens de Frege (1879), sendo usualmente caracterizada como lógica "matemática" ou "simbólica", referindo-se ao desenvolvimento da lógica com base em cálculos. O acontecimento fundamental desse período foi a separação nítida dos problemas lógicos dos psicológicos e retomado a pesquisa lógica no sentido estrito,

Inferir, segundo Copi (1978, p. 148 *apud* Kobashi, 2018, p. 30), “é extrair uma conclusão de uma ou mais premissas.”. A Lógica é uma disciplina que trata a análise de argumentos, um argumento consiste em uma conclusão e em evidências que a sustentam; a lógica se concentra na análise do argumento e não do pensamento. Portanto, em relação aos argumentos, a questão é se as premissas estão adequadamente relacionadas à conclusão.

Na Lógica:

- a) apenas o argumento é analisável. Não é possível analisar o pensamento. O pensamento é objeto da Psicologia;
- b) a correção lógica de uma inferência não depende da verdade das premissas [...];
- c) é possível inferir com base em pressupostos duvidosos ou falsos.
- Mais ainda:
- d) a Lógica oferece métodos para avaliar a correção dos argumentos;
- e) indica que inferências, formuladas como argumentos, podem ser aceitos como corretos;
- f) o poder de persuasão não é objeto da Lógica;
- g) os argumentos logicamente incorretos podem ser persuasivos enquanto argumentos logicamente corretos podem ser pouco persuasivos (Kobashi, 2018, p. 30).

Nos processos inferenciais imediatos e mediatos, as proposições são simbolizadas pelas letras S e P. As especificações de quantidade e qualidade das proposições são codificadas com as seguintes letras: A, I, E, O. A quantidade alude aos aspectos universais (todos) e particulares (alguns), já a qualidade se relaciona aos aspectos afirmativo e negativo das proposições. O quadro “Codificação de proposições” mostra a codificação adotada na Lógica (Kobashi, 2019):

- Na Lógica, as proposições são codificadas pelas letras A, E, I e O.
- A – todo S é P – proposição denominada UNIVERSAL AFIRMATIVA.
- E – nenhum S é P – proposição denominada UNIVERSAL NEGATIVA.
- I – algum S é P – proposição denominada PARTICULAR AFIRMATIVA.
- O – algum S não é P – proposição denominada PARTICULAR NEGATIVA.
- Fonte: (Kobashi, 2019, p. 32).

Na inferência imediata, a conclusão é obtida a partir de uma única premissa. A seguir é apresentado o "Quadro de Oposição", que permite visualizar as conclusões obtidas por inferência imediata (Kobashi, 2019):

Quadro 21 - Quadro de Oposição

PROPOSIÇÃO	A	E	I	O
Se A é verdadeira	VERDADEIRA	FALSA	VERDADEIRA	FALSA
Se E é verdadeira	FALSA	VERDADEIRA	FALSA	VERDADEIRA
Se I é verdadeira	?	FALSA	VERDADEIRA	?
Se O é verdadeira	FALSA	?	?	VERDADEIRA
Se A é falsa	FALSA	?	?	VERDADEIRA
Se E é falsa	?	FALSA	VERDADEIRA	?
Se I é falsa	FALSA	VERDADEIRA	FALSA	VERDADEIRA
Se O é falsa	VERDADEIRA	FALSA	VERDADEIRA	FALSA

Fonte: Kobashi (2019).

Segundo a definição proposta por Copi (1978, p. 148), a inferência mediata ocorre quando há mais de uma premissa, como no caso do silogismo que possui duas premissas. O silogismo é um exemplo clássico de inferência mediata,

Exemplo 1:

Premissa 1: Todos os gatos são felinos.

Premissa 2: O animal A é um gato.

Conclusão: Portanto, o animal A é um felino.

Exemplo 2:

Todos os animais que miam são gatos.

Cheese mia.

Portanto, Cheese é uma gata.

Exemplo 3:

Todas as aves são animais que têm asas.

O canário é uma ave.

Logo, o canário é um animal que tem asas.

O raciocínio concluído com base em premissas denomina-se inferência mediata. Como no exemplo, é um conjunto de proposições sequenciais que resulta conclusões válidas. A conclusão é interpretada a partir da primeira premissa com a intercessão da segunda premissa.

Compreender as inferências mediatas é especialmente relevante para a Organização e Representação de Informações, ao possibilitar a criação de ferramentas de tratamento de informação, como sistemas de classificação, ontologias, tesouros e taxonomias, que apresentam uma hierarquização de conceitos (Kobashi, 2018).

A hierarquização de conceitos no contexto da Lógica parte de inferências imediatas, pois quanto maior a extensão de um termo, maior a quantidade elementos designados pelo termo. Por outro lado, quanto maior a intensão de um termo, menor é a quantidade de elementos que podem ser designados por ele. Como, por exemplo, os termos ordenados do mais geral (maior extensão): Animal, mamífero, felino, gato, gato siamês.

De acordo a Kobashi e Francelin (2011), os princípios tradicionais de organização de conceitos ou “canônicos”, são baseados na Filosofia e na Lógica clássicas, têm função basicamente normativa; tendo Dalhberg desenvolvido seus sistemas de organização de conceitos no interior do cânone do racionalismo moderno.

De acordo a Sowa (2000) o sistema de *gráficos existenciais* de Peirce de 1897 apresenta a versão mais amplamente usada da lógica até o momento de sua publicação, determinando as formas mais simples e primitivas de expressar os elementos da lógica e desenvolveu uma notação gráfica para expressar as formas dos gráficos existenciais, que também podem ser

expressos em linguagem natural, em notação algébrica ou diversas representações lineares, gráficas ou mesmo faladas. Sowa (2000, tradução nossa) desenvolveu uma tabela listando as cinco primitivas semânticas de Peirce, cada uma ilustrada com um exemplo em inglês e “como esses cinco elementos são primitivos, eles não podem ser formalmente definidos em termos de algo mais primitivo; em vez disso, a coluna do meio da tabela indica brevemente seu "significado informal”. Como tratado no quadro 22:

Quadro 22 - Cinco primitivas semânticas

Tabela 1. Cinco primitivas semânticas		
Primitivo	Significado Informal	Exemplo em inglês
Existência	Algo existe.	<i>Tem um cachorro.</i>
Correferência	Algo é o mesmo que algo.	<i>O cachorro é meu animal de estimação.</i>
Relação	Algo está relacionado a algo.	<i>O cachorro tem pulgas.</i>
Conjunção	A e B	<i>O cachorro está correndo, e o cachorro está latindo.</i>
Negação	Não A.	<i>O cachorro não está dormindo.</i>

Fonte: Sowa (2000).

As cinco primitivas estão disponíveis em todas as linguagens naturais e em todas as versões da lógica de primeira ordem, são chamadas de primitivas semânticos porque expressam, além das relações sintáticas, as relações semânticas entre os signos e o mundo. É importante ressaltar que as notações expressam essas cinco primitivas a lógica de primeira ordem como um subconjunto, por exemplo, a lógica de primeira ordem é um subconjunto de SQL e da mesma forma diferentes linguagens podem usar notações diferentes para representar as cinco primitivas, como no quadro 23:

Quadro 213 - Cinco primitivas semânticas LN, LD e operadores lógicos

	Linguagem natural	Lógica descritiva	Operadores lógicos de consulta de banco de dados
<i>Existência</i>	Na maioria das línguas naturais, a existência está implícita na menção de algo.	<i>quantificador existencial</i> , a existência pode ser expressa por um símbolo explícito, como \exists .	EXISTS
<i>Correferência</i>	Para dizer que dois signos diferentes se referem à mesma coisa, as línguas naturais usam uma variedade de métodos, tanto explícitos quanto implícitos: pronomes, determinantes, flexões e formas do verbo <i>ser</i> .	A maioria das notações lineares para lógica usa variáveis e o sinal de igual, e notações gráficas usam linhas de conexão ou ligaduras.	Variáveis e sinal de igual
<i>Relação</i>	Palavras de conteúdo em linguagens naturais expressam alguma informação sobre pelo menos uma entidade, conhecida como o <i>referente</i> da palavra, mas também podem relacionar ou implicar outras entidades.	?	tabelas
<i>Conjunção</i>	Em ambas as linguagens naturais e artificiais, a conjunção pode ser expressa implicitamente fazendo uma declaração após a outra ou explicitamente por uma palavra como <i>e</i> ou um símbolo como \wedge	a conjunção pode ser expressa implicitamente fazendo uma declaração após a outra ou explicitamente por uma palavra como <i>e</i> ou um símbolo como \wedge	usa a palavra-chave AND
<i>Negação</i>	Todas as linguagens naturais e a maioria das versões da lógica fornecem palavras, inflexões ou símbolos para expressar negação. As maiores variações de uma língua para outra estão nos métodos para distinguir o contexto ou escopo do que é negado do que não é negado	Todas as linguagens naturais e a maioria das versões da lógica fornecem palavras, inflexões ou símbolos para expressar negação. As maiores variações de uma língua para outra estão nos métodos para distinguir o contexto ou escopo do que é negado do que não é negado	usa a palavra-chave NOT com parênteses para mostrar o escopo

Fonte: Elaborado pela autora.

A Web Ontology Language (OWL) e seus perfis são baseados em Lógicas descritivas (DL). Uma lógica descritiva (DL) modela conceitos, funções e indivíduos, e seus relacionamentos. Lógicas descritivas (DL) são uma família de linguagens formais de representação de conhecimento, quadro 24.

Quadro 224 - Lógicas descritivas (DL)

Conventional Notation			
Symbol	Description	Example	Read
\top	\top is a special concept with every individual as an instance	\top	top
\perp	empty concept	\perp	bottom
\sqcap	intersection or conjunction of concepts	$C \sqcap D$	C and D
\sqcup	union or disjunction of concepts	$C \sqcup D$	C or D
\neg	negation or complement of concepts	$\neg C$	not C
\forall	universal restriction	$\forall R. C$	all R-successors are in C
\exists	existential restriction	$\exists R. C$	an R-successor exists in C
\sqsubseteq	Concept inclusion	$C \sqsubseteq D$	all C are D
\equiv	Concept equivalence	$C \equiv D$	C is equivalent to D
\doteq	Concept definition	$C \doteq D$	C is defined to be equal to D
:	Concept assertion	$a : C$	a is a C
:	Role assertion	$(a, b) : R$	a is R-related to b

Fonte: Wikipédia (2023).

Em 2007, o grupo de trabalho do W3C OWL iniciou aprimoramentos e extensões para o OWL. Em 2009, foi concluído a recomendação OWL2, essa linguagem de ontologia é baseada em lógica descritiva e visa fornecer um formalismo lógico para ontologias e a Web Semântica.

No OWL os construtores lógicos são aplicados para criar descrições de classes complexas a partir de classes atômicas:

OWL Construtores de Classe Lógicas

logical AND (conjunction):	<code>owl:intersectionof</code>	\sqcap
logical OR (disjunction):	<code>owl:unionOf</code>	\sqcup
logical negation:	<code>owl:complementof</code>	\neg

Segundo Baader, Horrocks, Lutz e Sattler (2017), as lógicas de descrição (DLs) são uma família de linguagens de representação de conhecimento que podem ser usadas para representar conhecimento de um domínio de forma estruturada. O nome "lógicas de descrição" é motivado pelo fato de que, por um lado, as noções importantes do domínio são representadas por **descrições conceituais**, ou seja, expressões construídas a partir de **conceitos atômicos** (predicados unários) e **papéis atômicos** (predicados binários) usando os construtores de conceitos e papéis fornecidos pela DL específica. Por outro lado, as DLs diferem de suas antecessoras, como redes semânticas e frames, por serem equipadas com uma semântica baseada em lógica que, com algumas diferenças de notação, é, na verdade, a mesma semântica da lógica clássica de primeira ordem.

Ainda segundo os autores, as descrições lógicas dividem o conhecimento de um domínio em duas partes: a terminológica chamada TBox e a assertiva chamada ABox. A

combinação de uma TBox e uma ABox é chamada de base de conhecimento (KB, na sigla em inglês). A TBox representa o conhecimento sobre a estrutura do domínio (semelhante a um esquema de banco de dados), enquanto a ABox representa o conhecimento sobre uma situação concreta (similar a uma instância de banco de dados).

Por exemplo, no domínio universitário, as declarações na TBox capturam conhecimento, a partir da definição do conceito “professor” e das predições de que professor é alguém que leciona um curso, assim como, define um aluno como alguém que frequenta um curso e especifica a regra de que alunos não ensinam. Logo, “as declarações na ABox fornecem exemplos específicos, como a afirmação de que Mary é uma pessoa, CS600 é um curso e Mary ensina CS600” (Baader; Horrocks; Lutz; Sattler, 2017, p. 2). Essas declarações são expressas em uma semântica formal baseada em lógica, sendo possível representá-las na lógica de primeira ordem como:

Teacher \equiv Person \exists teaches.Course,
 Student \equiv Person \exists attends.Course,
 \exists attends. \neg Student,
 Mary : Person,
 CS600 : Course,
 (Mary, CS600) :teaches.

(Baader *et al.*, 2017, p. 2).

Equivalentemente, essas instruções podem ser escritas na sintaxe da lógica de descrição da seguinte forma:

Teacher \equiv Person \exists teaches.Course,
 Student \equiv Person \exists attends.Course,
 \exists attends.
 \neg Student,
 Mary : Person,
 CS600 : Course,
 (Mary, CS600) :teaches.

(Baader *et al.*, 2017, p. 2).

As primeiras três declarações desse conjunto de conhecimento constituem sua TBox, e as últimas três declarações constituem sua ABox. Na semântica das descrições lógicas existe um entendimento bem definido e compartilhado de quando uma declaração é inferida por uma base de conhecimento. No exemplo anterior, é inferido que Mary é uma professora. As tarefas

comuns de raciocínio incluem verificar a satisfatibilidade ⁵² de conceitos e a consistência de bases de conhecimento, determinar quando um conceito é mais específico do que outro e responder a diferentes tipos de consultas na base de conhecimento (Baader; Horrocks; Lutz; Sattler, 2017, p. 2).

4.3 SEMÂNTICA FORMAL

A linguagem natural possui muitas ambiguidades, através das homônimas e a sinonímias, que podem levar à ambiguidade. Diferentemente, as linguagens artificiais (ou documentárias), possibilitam que os termos sejam controlados para evitar as ambiguidades, através do controle de homônimas e sinônimas e o uso de termos monossêmicos. Dessa forma, para entender uma palavra é preciso realizar a desambiguação, na web ocorre o mesmo processo. Como, por exemplo, na busca por imagens da palavra-chave “manga” numa ferramenta de busca (*search engine*), como o Google, o resultado compreendia diversos conceitos diferentes:

Figura 22 - Resultado de busca baseada em palavra-chave



Fonte: Elaborado pela autora.

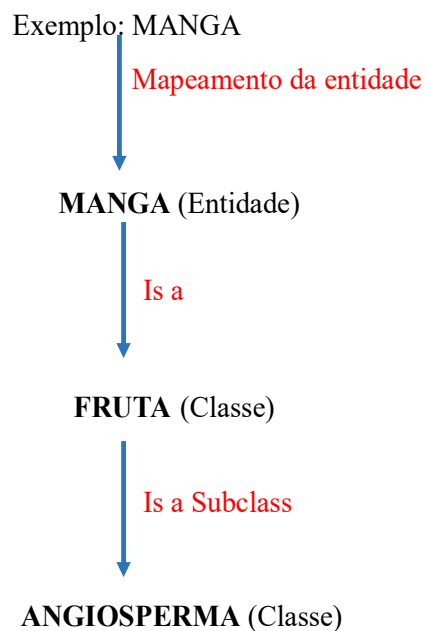
Desse modo, como mostra a figura 22, um exemplo são os homônimos gerados pelos diferentes significados da palavra manga. Que pode ser uma fruta ou parte do vestuário, a figura 23⁵³ mostra a desambiguação da palavra manga, ocorrendo o processo de mapeamento da

⁵² Na lógica matemática, uma fórmula é *satisfatível* se for verdadeira sob alguma atribuição de valores às suas variáveis.

⁵³ As figuras 22, 25 e 26, foram adaptadas das anotações do curso *Semantic Web Technologies*, do Dr. Harald Sack, no canal de cursos do Hasso Plattner Institute, da Universidade de Postdam. Disponível em: <https://youtu.be/TrFBgkLujrA>

entidade, seguido da identificação da classe e das subclasses da palavra e por consequência, a partir disso é possível definir suas propriedades (tem caroço, tem flor, é terrestre, etc.).

Figura 23 - Desambiguação e mapeamento de entidade



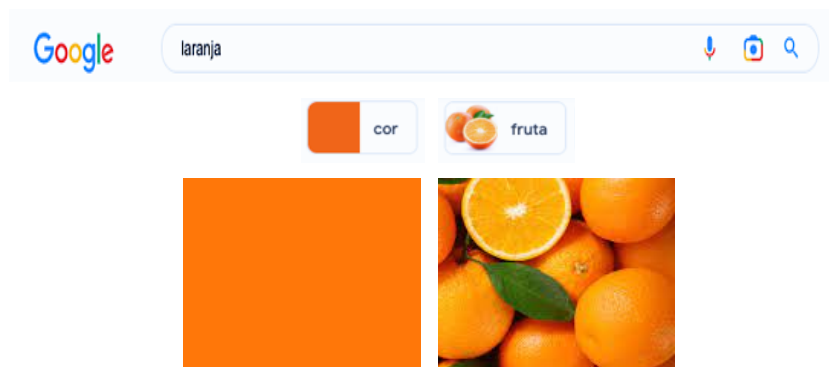
Fonte: Adaptado de Sack (2019).

Outras palavras que são homônimas, o que expressa que são iguais, mas com significados diferentes, é a palavra laranja (cor) e laranja (fruta).

Laranja ou Laranjas pode referir-se a:

- Laranja — fruto citrino
- Laranja (cor) — cor
- Laranja (indivíduo) — intermediário em operações financeiras ilícitas

Figura 24 - Interface do buscador



Fonte: Elaborado pela autora.

Além de termos diferentes com significados diferentes, mas com grafia e/ou fonéticas iguais, também existe o problema da polissemia, onde um termo tem mais de um significado, como a palavra terra.

Terra ou Terras pode referir-se a:

- Terra (planeta) — o terceiro planeta do sistema solar
- Terra (economia) — fator de produção
- Terra (elemento) — um dos quatro elementos, segundo os filósofos antigos

Do mesmo modo, existem termos diferentes que são homônimos, com a mesma grafia e fonética de terra, como:

- Terra (eletricidade) — potencial elétrico da superfície da Terra
- Terra (mitologia) — a deusa da Terra, segundo a mitologia romana
- Terra (satélite) — satélite que monitora as condições ambientais do planeta

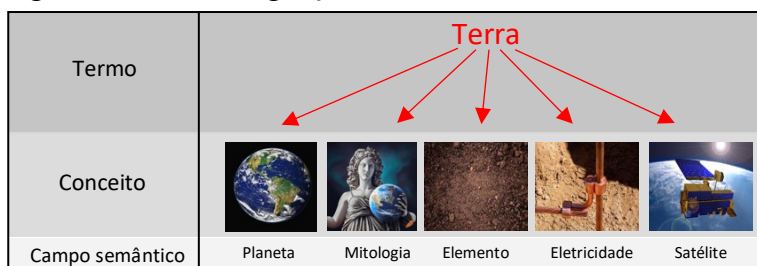
Por fim, o termo também pode ser referido para entidades nomeadas:

- Terra (DC comics) — personagem dos quadrinhos
- Terra (livro) — livro de Sebastião Salgado
- Terra (empresa) — provedor de acesso à internet

A ambiguidade da linguagem natural, mostra que os conceitos são fundamentais para representar o conhecimento via taxonomias e tesouros no ambiente web. Visto que, o termo apresenta ambiguidades, por conseguinte, a definição por trás do conceito é o que permite estabelecer um significado para a entidade. Além disso, a constante atualização do conteúdo e a ausência de um ambiente controlado na web, demanda que o significado de uma entidade seja baseado em conceitos e não termos. De maneira que, as alterações linguísticas podem ser

designadas por unidades terminológicas passíveis de atualização, como os termos ou rótulos (*labels*).

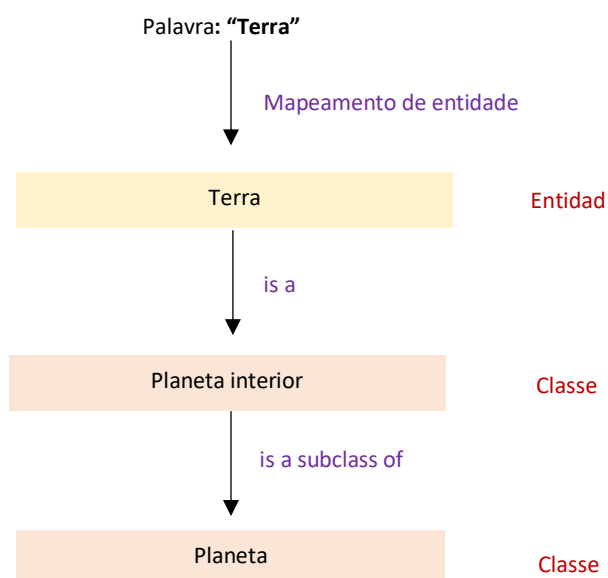
Figura 25 - Desambiguação de Terra



Fonte: Elaborado pela autora.

Assim, Terra pode ser um termo ambíguo pelo fato de ser um texto (palavra) em linguagem natural (LN). De maneira que, a LN exige a desambiguação para a interpretação de uma mensagem e das ambiguidades linguísticas. Numa busca semântica o significado (semântica) das entidades e classes, deve ser explicitamente definido a que se refere ou o que significa, através do processo de mapeamento da entidade, identificação da classe e das subclasses e definição das propriedades, na figura 26.

Figura 26 - Mapeamento e desambiguação de terra

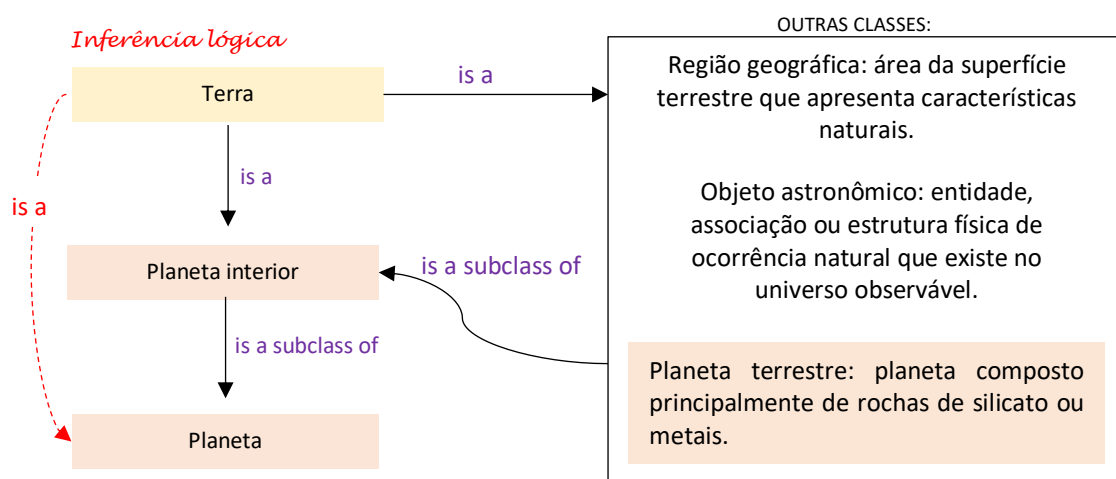


Fonte: Adaptado de Sack (2019).

Bem como, é necessário definir como a entidade se relaciona com outros objetos, para estabelecer a semântica e a representação do conhecimento. Somente a hierarquia de classe não

é suficiente para entender o significado de uma classe (categoria de conceitos), portanto, também é preciso definir as propriedades compartilhadas pelas entidades da classe, na figura 27.

Figura 27 - Propriedades compartilhadas pelas entidades da classe



Fonte: Adaptado de Sack (2019).

A sintaxe das linguagens artificiais baseadas em computador, é fornecida pela linguagem de representação na qual o SOC está codificado. Na linguística, a sintaxe é o estudo de como as palavras podem ser combinadas em sentenças (Cançado, 2005). No contexto dos conteúdos presente na web, a frase sintaticamente correta tem um significado implícito. Por conseguinte, na frase “Um planeta interior é uma terra”, o conhecimento implícito não está adequado, ou seja, a informação não é possível ser derivada por meio de deduções lógicas a partir das informações disponíveis.

Na ciência da computação, a sintaxe define a estrutura normativa dos dados e nas linguagens formais, é um conjunto de regras, pelas quais é possível criar expressões a partir de um conjunto fundamental de símbolos (alfabeto).

O significado da informação (Semântica) é realizado de modo explícito, por representações de conhecimento formais (estruturadas) e padronizadas (Ontologias); para ser possível, o processamento do significado da informação automaticamente, assim como, relacionar e integrar dados heterogêneos.

Segundo abordagem de Cançado (2005, p. 15) “a sintaxe, que é o estudo de como as palavras podem ser combinadas em sentenças; e a semântica é o estudo do significado das línguas ou ainda que é o estudo do significado das palavras e das sentenças”. De acordo a autora,

mesmo uma frase alheia totalmente a nossa experiência, como exemplo, “O macaco roxo tomava um sorvete no McDonald's”, pode ser entendida, devido ao conhecimento dos significados das palavras e o entendimento das instruções relativas à ordem da sentença, pois a teoria semântica envolve o significado das palavras e os algoritmos que combinam os significados para se chegar a um significado da sentença, sendo a palavra a menor unidade dessa composição, e as frases e sentenças a maior unidade de análise. “Uma teoria semântica deve ser capaz de enunciar de que modo a relação entre o significado de palavras e o significado de sentenças, depende da ordem das palavras ou de outros aspectos da estrutura gramatical da sentença” (Cançado, 2005, p. 19).

Cançado (2005, p. 147) delimita os principais aspectos que a caracterizam: i) a composicionalidade; ii) as propriedades semânticas; iii) a referência e a representação.

1) A composicionalidade e a expressividade linguística: uma teoria semântica deve atribuir o significado das palavras e das sentenças, identificar as características da relação entre o significado das palavras e das sentenças e estabelecer como a relação depende da ordem das palavras e da estrutura das sentenças.

2) As propriedades semânticas: uma teoria semântica deve caracterizar e explicar as relações sistemáticas entre palavras e entre sentenças de uma língua. Por exemplo, abordar as noções de implicações, sinonímias, contradições, ambiguidades, anomalias, etc.

3) A referencialidade e a representação: numa abordagem referencial, o estudo do significado diz respeito à ligação entre as expressões linguísticas e o mundo. Já para uma abordagem representacional do significado, devemos levar em conta a ligação entre a linguagem e os construtos mentais.

Uma importante iniciativa, no âmbito da abordagem referencial, é aquela conhecida como Semântica Formal, formada pelos seguintes aspectos:

- Composicionalidade
- Condição de verdade: um conjunto de regras para aplicação do conceito de verdade a sentenças de sistemas formais baseados em lógica, sem o risco da ambiguidade inerente à linguagem natural (Almeida, 2011, p. 32).
- concepção de modelos em semântica (Oliveira, 2001).

O presente trabalho tem como referencial conceitual no desenvolvimento de ontologias leves e pesadas (formais), as abordagens semânticas ocorridas no século XX: referencial e semântica formal. Estas abordagens estão destacadas no quadro 25.

Quadro 23 - Quadro sinótico das abordagens semânticas em Linguística

Evolução	Abordagem	Breve descrição
Antiguidade	Platônica	Origem dos nomes
	Aristotélica	Categorias
Século XIX	Diacrônica	Semântica como disciplina
	Sincrônica	Origem da orientação linguística moderna
	Referencial	Sentido, significado, referência
Século XIX e XX	Campos semânticos	Palavras com significados relacionados
	Análise componencial	Conjuntos de características semânticas
	Metalinguagem	Componentes como entidades lógicas
Século XX	Pragmática	Uso de jogo argumentativo
		Sentença tem função social
	Mentalista	Linguagem e esquemas mentais
		Outros níveis de representação mental
		Papéis temáticos dos argumentos de um verbo
	Referencial	Triângulo do significado
Semântica formal	Teorias de origem Lógica Filosófica	

Figura 4: Quadro sinótico das abordagens semânticas em Linguística

Fonte: Almeida (2011, p. 35).

Os tipos de abordagens semântica da Web Semântica em relação semântica linguística está no quadro 26:

Quadro 26 - Quadro sinótico das abordagens à semântica na WS

Abordagem	Breve descrição	Semântica linguística
Repr. do conhecimento	A semântica é formal e baseada em teorias lógico-filosóficas	formal
Repr. do conhecimento	A semântica é o significado de sentenças através de interpretação	formal
Semântica da <i>web</i>	A semântica possibilita interpretação por um computador	formal
Semântica mundo real	A semântica mapeia objetos do mundo para o sistema	indeterminada
Semântica axiomática	A semântica mapeia linguagens da WS para a Lógica	formal
Teoria dos Modelos	A semântica valida processos de inferência automáticos	formal
Semântica implícita	A semântica transmite o consenso obtido entre as pessoas	indeterminada
	Semântica inserida em padrões de dados não legível para máquinas	indeterminada
Semântica informal	A semântica é explícita e informal	indeterminada
Formal para humanos	Semântica explícita e expressa em linguagem formal, para pessoas	formal
Formal para máquinas	Semântica explícita e expressa em linguagem formal, para máquinas	formal
	Semântica definida por regras sintáticas mais interpretações	formal
Semântica nebulosa	Semântica baseada em estatística	formal

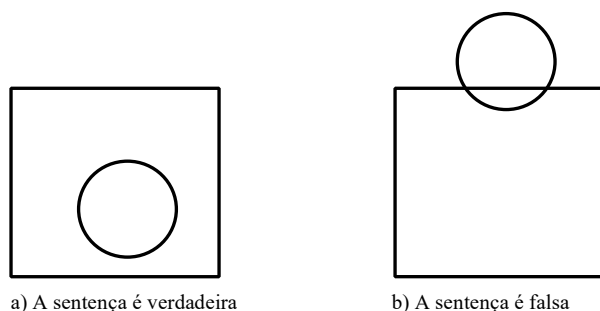
Figura 5: Quadro sinótico das abordagens à semântica na WS

Fonte: Almeida (2011, p. 39).

A Semântica formal para processamento por máquinas “trata da semântica explícita, especificada formalmente, a qual pode ser utilizada por computadores para processamento direto via mecanismos de inferências para derivar novos dados a partir de existentes” (Almeida; Souza, 2011, p. 38). Os principais aspectos que permeiam o estudo da Semântica Formal ao longo de sua evolução são: i) ênfase no princípio da composicionalidade; ii) uso da condição de verdade para explicar o significado; iii) concepção de modelos em semântica (Oliveira, 2001 *apud* Almeida; Souza, 2011, p.32). O princípio da composicionalidade estabelece que o significado das sentenças depende do significado das palavras, igualmente, o significado de uma sentença depende do conhecimento dos significados das partes e das regras que estabelecem a combinação sintática das partes (Almeida; Souza, 2011). O uso da condição de verdade para explicar o significado diz respeito a determinar em que condições tal sentença é

verdadeira. “Uma vez que a noção de verdade é central na lógica, a preocupação reside em explicar como as sentenças dos sistemas formais são consideradas verdadeiras ou falsas” (Almeida; Souza, 2011, p. 36). As condições para uma sentença ser verdadeira ou falsa, são demonstradas na figura 28, com a seguinte sentença: o círculo está no quadrado. Desse modo, a imagem com formas geométricas permite demonstrar se a sentença é verdadeira ou falsa.

Figura 28 - Condições verdade para a sentença do círculo está dentro do quadrado



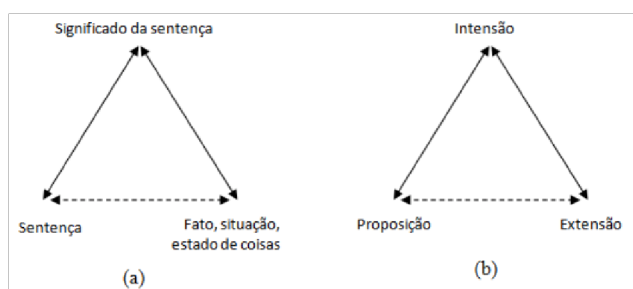
Fonte: Almeida e Souza (2011).

Assim, em “Semântica Formal, tais situações são denominadas situações possíveis (*possible situations*) ou estados das coisas (*state of affairs*)” (Almeida; Souza, 2011, p. 33). Nesse contexto, o termo situação significa “um cenário incompleto, uma parte do universo limitado no espaço e no tempo. Já o termo mundo é usado quando as pessoas têm uma ideia completa do que o mundo deve ser, com todos os seus componentes e habitantes, os quais não se alteram ao longo do tempo” (Almeida; Souza, 2011, p. 33).

A lógica permitiu a aplicação da noção de proposição ao estudo da sentença, que passa a corresponder ao significado da sentença, “o que corresponde a um deslocamento no triângulo do significado, do nível da palavra para o nível da sentença” (Nirenberg; Raskin, 2004 *apud* Almeida; Souza, 2011, p. 33), como mostrado na figura 29.

Desse modo, ocorre uma renomeação dos vértices do triângulo, portanto, “nenhum dos elementos do triângulo (b) se relaciona diretamente com a linguagem natural, uma vez que uma proposição é resultado da tradução de uma sentença declarativa em uma metalinguagem baseada em lógica” (Almeida; Souza, 2011, p. 33).

Figura 29 - Triângulo do significado (a) para sentenças e (b) usando termos lógicos



Fonte: Adaptado de Nirenberg e Raskin (2004).

O estudo da linguagem natural (linguística, semiótica e terminologia) é utilizado em Representação do Conhecimento, onde a linguagem corresponde, em geral, a um tipo de lógica (Almeida; Souza, 2011, p. 36). De acordo a Uschold (2005 *apud* Almeida; Souza, 2011), distingue e cita os tipos de semânticas presentes no mundo dos sistemas computacionais, classificando-as em semântica do mundo real, semântica axiomática e semântica Teórica dos Modelos. Assim, estabelece uma classificação a semântica para humanos e máquinas.

A representação formal do conhecimento se dá através da semântica formal para representar os dados desestruturados de forma legível pelas máquinas, nesse processo o significado semântico da entidade e das classes, é alcançado via uma especificação formal explícita de uma conceituação compartilhada. De tal modo, partindo de dados estruturados, informações podem ser explicitadas com a ajuda de sistemas lógicos de inferência, assim, para representar esses conceitos é preciso definir as propriedades e restringir os domínios e o intervalo dessas propriedades. Dessa forma, a Web Semântica é uma atualização da web em semântica explícita ou proposições semânticas. O quadro 27 apresenta os tipos de semântica que fazem parte dos sistemas de informação e comunicação, sendo a maioria proveniente da linguagem natural:

Quadro 24 - Classificação para a semântica de Uschold

SEMÂNTICA IMPLÍCITA	SEMÂNTICA INFORMAL	SEMÂNTICA FORMAL PARA PROCESSAMENTO HUMANO	SEMÂNTICA FORMAL PARA PROCESSAMENTO POR MÁQUINAS
O significado é comunicado a partir do entendimento comum obtido por consenso entre pessoas. (Ex. XML e as definições de marcação através do consenso da comunidade).	O significado é explícito e expresso informalmente. (Ex. glossários, em especificações de requisitos para sistemas, etc.).	Semântica explícita e expressa em linguagem formal, mas que serve à comunicação humana e não a sistemas.	Semântica explícita, especificada formalmente, utilizadas por computadores para processamento através de inferências para derivar novos dados a partir de existentes.

Fonte: Uschold (2005 *apud* Almeida; Souza, 2011, p. 38).

O quadro 28 apresenta definições diferentes para as categorias semânticas: semântica implícita, semântica formal e semântica nebulosa.

Quadro 25 - Categorias para a semântica de Sheth, Ramakrishnan; Thomas (2005)

SEMÂNTICA IMPLÍCITA	SEMÂNTICA FORMAL	SEMÂNTICA NEBULOSA
Padrões de dados não explicitados em sintaxe legível por computadores; (Ex. textos não estruturados e em repositórios de documentos com estrutura pouco formal) É necessário p/ análise: técnicas da Recuperação da Informação e da Linguística computacional.	Dados representados por linguagem formal baseada em estruturas sintáticas estabelecidas e por regras que definem as combinações possíveis, associadas às interpretações semânticas; declarações expressas para sistemas, utilizadas em Representação do Conhecimento, Inteligência Artificial e Bancos de Dados.	Exploração de relações não estabelecidas explicitamente; a formalização da combinação em linguagens passíveis de inferência; aplicada à extensão de modelos de Bancos de Dados e de Representação do Conhecimento, em casos em que a informação é imprecisa ou incompleta.

Fonte: Baseado em Sheth, Ramakrishnan e Thomas (2005 *apud* Almeida; Souza, 2011, p. 38).

Assim, a semântica do ponto de vista do uso pelo sistema no computador, legível por máquinas, difere da semântica que os humanos estabelecem na modelagem conceitual dos sistemas de informação. Nesse cenário, grande parte do conteúdo das imagens em movimento faz parte do grupo de informações não estruturadas, que não podem ser recuperadas ou inferidas mediante ontologias, ou semântica explícitas na web atual.

O maior obstáculo para uso de sistemas semânticos tem sido a complexidade para a aquisição de conhecimento, pois a capacidade que possui os seres humanos de reutilizar conhecimento prévio não é uma habilidade dos computadores. Dessa forma, de acordo a Sowa (2011, tradução nossa) a “tarefa de traduzir informações de um livro didático para uma forma semântica computável requer as habilidades combinadas de um linguista, lógico, cientista da computação e especialista no assunto” e apresenta-se como grande desafio para projetar ferramentas automatizadas que combinem as diferentes habilidades de diversos especialistas.

Os computadores podem processar números, estruturas de dados e até axiomas em lógica muito mais rápido do que as pessoas. Mas as pessoas aproveitam o conhecimento prévio que os computadores não têm. Hao Wang (1960), por exemplo, escreveu um programa que provou todos os 378 teoremas em lógica proposicional e de primeira ordem do Principia Mathematica. Em um computador de tubo de vácuo lento, o programa de Wang levou uma média de 1,1 segundo por teorema – muito menos tempo do que Whitehead e Russell, os dois lógicos brilhantes que escreveram o livro. Mas os teoremas do Principia requerem uma quantidade insignificante de conhecimento embutido – apenas cinco axiomas e algumas regras de inferência (Sowa, 2011, tradução nossa).

Segundo o autor, durante as décadas de 1970 e 1980, ocorreu uma maior atenção nos sistemas especialistas baseados em regras e programas para processamento de linguagens

naturais, mas era necessário que os aplicativos fossem alimentados como uma enorme quantidade de conhecimento prévio para produzir resultados, com Engenheiros do conhecimento e especialistas codificando o conhecimento em lógica formal, o que exigia longos anos de treinamento em lógica, ontologia, análise conceitual, projeto de sistemas e métodos para entrevistar os especialistas, além do pagamento de altos salários para especialistas como médicos e engenheiros, o que gerava altos custos para aplicativos comerciais convencionais.

Durante a década de 1990, a abundância de dados na World Wide Web forneceu dados brutos para métodos estatísticos e a compreensão da linguagem foi deixado em favor de métodos estatísticos para recuperação e extração de informações, mas os métodos estatísticos não geram uma representação semântica adequada para raciocínio adicional ou para explicações em linguagem comum.

No início do século XXI, a Web Semântica adaptou as tecnologias de IA da década de 1980 aos vastos recursos da web. Porém, as linguagens e ferramentas da Web Semântica eram desconhecidas dos desenvolvedores e distantes do software convencional.

Documentos que as pessoas escrevem para se comunicar com outras raramente são tão precisos quanto a lógica. No entanto, as pessoas podem ler esses documentos e relacioná-los com notações formais para ciência, matemática e programas de computador. Eles podem derivar qualquer informação que precisem, raciocinar sobre ela e aplicá-la em um nível apropriado de precisão. Essa flexibilidade é essencial para um sistema de aquisição de conhecimento — automatizado, semiautomatizado, ou pelo menos assistido por computador. No último meio século, pesquisadores de IA e linguistas computacionais tentaram atingir esse objetivo (Sowa, 2011, tradução nossa).

Os sistemas de Processamento de Linguagem Natural mais avançados empregam uma abordagem semântica baseada em algum tipo de lógica. Apesar de quatro décadas de pesquisa, nenhum sistema que se baseie nessa abordagem consegue ler uma página de um livro de ensino médio e aplicar os resultados de forma tão eficaz quanto um aluno comum. Isso levou até mesmo os pioneiros nos métodos lógicos a questionarem essa capacidade. Isso reforça a necessidade de rever os princípios fundamentais da linguística e, sobretudo, da semântica, requerendo a criação de distinções mais aprofundadas do que a teoria semântica atual (Sowa, 2011).

A Web Semântica foi inspirada por Tim Berners-Lee, mas foi projetada por um comitê do W3C. Ele evoluiu a partir de um discurso na primeira *World Wide Web Conference* (Berners-Lee, 1994):

Adicionar semântica à web envolve duas coisas: permitir documentos que tenham informações em formulários legíveis por máquina e permitir que links sejam criados com valores de relacionamento (Berners-Lee, 1994, tradução nossa).

Em 2001, alguns componentes da Web Semântica haviam sido especificados pelas recomendações do W3C, mas o único consenso sobre a arquitetura geral era o chamado “bolo de camadas” à esquerda da Figura 1. No entanto, a Web Semântica vem desenvolvendo um conjunto valioso de ferramentas, e elas devem ser melhor integradas tanto com a tecnologia de IA quanto com softwares mais convencionais. Uma maneira de começar é preencher a caixa para “unificar lógica” na Figura 1 com Common Logic (ISO/IEC 2007). A base semântica para a Common Logic é baseada em uma proposta de Hayes e Menzel (2001). Guha e Hayes (2002) adotaram essa semântica para RDF, e ela é compatível com todas as lógicas dos bolos de camada. A Common Logic também foi adotada como a lógica unificadora para os diagramas UML, que são amplamente utilizados para especificar software convencional (OMG, 2010). Uma lógica unificadora pode dar suporte a ferramentas de desenvolvimento com mapeamentos definidos com precisão entre componentes de diferentes comunidades com notações e metodologias diferentes.

Segundo Sowa (2000), a Internet é um sistema semiótico gigante. É uma coleção maciça dos três tipos de signos de Peirce: *ícones*, que mostram a forma de algo; *índices*, que apontam para algo; e *símbolos*, que representam algo de acordo com alguma convenção. Contudo, o autor afirma que as propostas atuais de ontologias e metadados negligenciaram algumas das características mais importantes dos signos: um signo tem três aspectos: é (1) uma *entidade* que representa (2) outra *entidade* para (3) um *agente*, fazendo com que algumas propostas de metadados não considerassem as próprias entidades que representam e os agentes (humanos, animais ou robôs) que as interpretam. O autor demonstra como a semiótica fornece diretrizes para organizar os signos com seus três ramos de sintaxe, semântica e pragmática e como as cinco primitivas semióticas fundamentais estão representadas em notações semanticamente equivalentes para lógica, como *linguagens naturais controladas* e linguagens de computador.

Ontologias contêm categorias, léxicos contêm sentidos de palavras, terminologias contêm termos, diretórios contêm endereços, catálogos contêm números de peças e bancos de dados contêm números, cadeias de caracteres e BLOBs (Binary Large Objects). Todas essas listas, hierarquias e redes são coleções de signos fortemente interconectadas. Mas as conexões primárias não estão nos bits e bytes que codificam os sinais, mas nas mentes das pessoas que os interpretam. O objetivo de várias propostas de metadados é tornar essas conexões mentais explícitas, marcando os dados com mais sinais. Esses

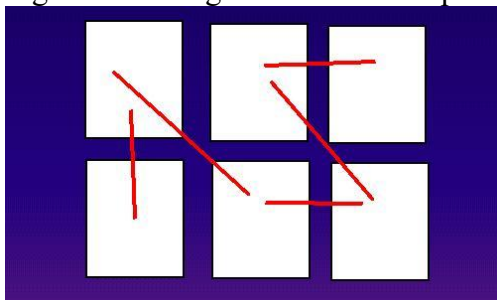
próprios sinais de meta-nível têm outras interconexões, que podem ser marcadas com sinais de meta-nível. Mas dados sem sentido não podem adquirir significado sendo marcados com metadados sem sentido (Sowa, 2000, tradução nossa).

No início dos anos 2000 a maioria dos processadores de texto lidavam apenas com um pequeno subconjunto de sintaxe, de tal modo que St. Laurent (1999 *apud* Sowa, 2000, tradução nossa) chamou os sistemas informatizados como desastre WYSIWYG (*What you see is what you get*): "Texto simples, por mais chato que seja, é muito mais fácil de gerenciar do que a saída de um processador de texto médio ou programa de editoração eletrônica", mas a pouca expressividade semântica fazia com que a sigla mais adequada fosse WYSIAYG (*What you see is all you get*), "o texto está tão sobrecarregado com tags de formatação que não há espaço para semântica ou pragmática" (Sowa, 2000, tradução nossa).

4.4 WEB SEMÂNTICA

Tim Berners-Lee na primeira Conferência Internacional da World Wide Web, no CERN, Genebra, Suíça, em setembro de 1994, na qual a formação do W3C foi anunciada, fez uma apresentação intitulada "A Necessidade de Semântica na web". Ele afirma que para o usuário o mundo do hipertexto é emocionante, mas que há muito pouca informação legível por máquina nele, pois o significado dos documentos é legível para aqueles que entendem a linguagem natural e o idioma do documento. Porém, "para um computador, então, a web é um mundo plano e chato, desprovido de significado" (Berners-Lee, 1994, tradução nossa).

Figura 30 - Imagem clássica do hipertexto



Fonte: Tim Berners-Lee (1994).

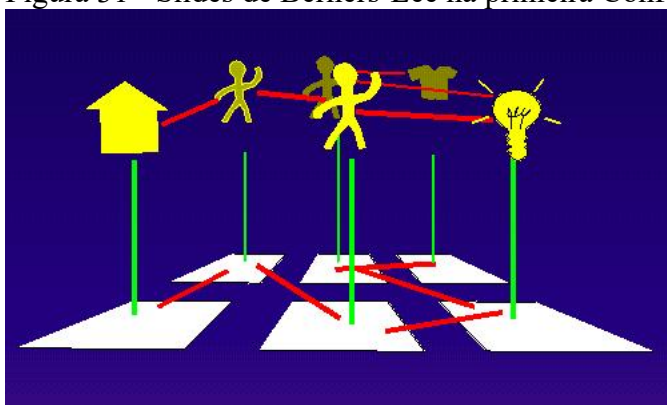
As características dos documentos na web, descrevem objetos reais e conceitos imaginários e estabelece relações entre conceitos. "Por exemplo, um documento pode descrever uma pessoa. O título de propriedade de uma casa descreve uma casa e também a relação de propriedade com uma pessoa." (Berners-Lee, 1994, tradução nossa). De tal modo, adicionar

semântica à web configura dois processos: documentos com informações em formatos legíveis por máquina e que a criação de links seja com valores de relacionamento. “Somente quando tivermos esse nível extra de semântica seremos capazes de usar o poder do computador para nos ajudar a explorar a informação em maior extensão do que nossa própria leitura.” (Berners-Lee, 1994, tradução nossa).

Um efeito importante do desenvolvimento de protocolos de segurança na web é que o espaço abstrato da informação da web está ligado à realidade. Ao assumir a responsabilidade verificável pelas declarações da web, uma parte garante um isomorfismo entre a web e a realidade. (Berners-Lee, 1994, tradução nossa).

Segundo o autor, as máquinas poderiam operar com informações da web para realizar atividades reais, como, por exemplo, um programa procurar uma casa e negociar a transferência da propriedade da casa para um novo proprietário.

Figura 31 - Slides de Berners-Lee na primeira Conferência Internacional da WWW



Fonte: Berners-Lee (1994).

Outro documento “Arquitetura de metadados”, publicado em 1997 por Tim Berners-Lee e escrito antes dos roteiros da Web Semântica, mas que é uma introdução as mesmas ideias, nele são colocados os conceitos de metadados em documentos da web. Portanto, o autor afirma que recursos são definidos por elementos informais legíveis por humanos na forma de documentos e por aspectos formais que são os seus metadados, “uma das características da World Wide Web é que os recursos, quando você os recupera, não ficam simplesmente sozinhos sem explicação, mas há informações sobre o recurso. Informações sobre informações são geralmente conhecidas como metadados” (Berners-Lee, 1997, tradução nossa). Ele define que metadados são informações compreensíveis por máquina sobre recursos da web, com semântica

e estrutura bem definidas, sendo a expressão legíveis por máquinas as informações geradas por agentes de software.

Os metadados foram chamados de "Metadados" porque começaram a vida e atualmente ainda são principalmente informações sobre recursos da Web, ou seja, dados sobre dados. No futuro, quando as linguagens e os mecanismos de metadados estiverem mais desenvolvidos, eles também deverão formar uma base sólida para uma rede de informações compreensíveis por máquina sobre qualquer coisa: sobre pessoas, coisas, conceitos e ideias. Mantemos esse fato em mente no projeto, embora o primeiro passo seja fazer um sistema de informação sobre informação (Berners-Lee, 1997, tradução nossa).

Dessa forma, define que metadados são dados sobre dados e os três modos de se obter metadados na web: a) no cabeçalho de um documento HTML, b) nos protocolos de transferência HTTP e por último, c) uma prática que era muito incomum e pouco usada na época, “a terceira maneira pela qual os metadados são encontrados é quando são pesquisados em outro documento.” (Berners-Lee, 1997, tradução nossa).

Em sua jornada de apoiar a construção do design e arquitetura do WWW, Tim Berners-Lee se preocupou com o significado de um documento na web. Por isso em 1999 publicou outro documento com título *Meaning* (Significado), com fundamentos sobre o significado de um documento no espaço URI e o introduz fazendo a seguinte pergunta: Qual é o significado de um documento? No texto ele aborda as diferenças de um documento físico em papel e a questão da ambiguidade da linguagem natural e como isso foi tratado em documentos da web, identificando coisas importantes por URIs.

Na Web Semântica, no entanto, o significado não é vago: a ideia é que as línguas devem ser definidas formalmente e com a maior precisão possível. A Web Semântica consiste em algumas linguagens "terminais" que são definidas apenas em termos de linguagem natural e algumas linguagens para as quais existem interpretações legíveis por máquina em outras linguagens formais. Enquanto os programas que processam documentos no primeiro tipo de idioma normalmente precisam ser codificados manualmente, os documentos no segundo conjunto podem ser processados automaticamente para convertê-los nos idiomas do primeiro conjunto (Berners-Lee, 1999, tradução nossa).

O autor alega que o conceito de “documento fundamentado” é importante para sistemas previsíveis, mas não servem para a web, bem como, as palavras de um idioma natural não são fundamentadas num conjunto exclusivo, ao contrário, a linguagem natural é dinâmica e se altera ao longo do tempo.

O mundo é parecido com a web, e qualquer tentativa da web de restringi-lo a ser parecido com uma árvore está fadado a forçar uma deturpação da realidade. Esta é a visão Wittgenstein do significado. Compreender essa visão às vezes confunde as pessoas sobre a maneira muito sistemática pela qual o significado nos protocolos da Internet é definido por camadas e camadas de especificações. (Berners-Lee, 1999, tradução nossa).

Na perspectiva do autor, as duas visualizações se misturam, mas os protocolos da internet estabelecem restrições sociais que restringem o uso conforme as especificações. Como, por exemplo, preencher os campos de um formulário do governo ou de empresas na web é um modelo fornecido que não pode ser alterado, é um contrato social aceito pelo usuário. Ainda conclui, devido a mudanças no desenvolvimento da Web Semântica, todas suas fundamentações serão sistemas legados desaparecidos da web.

Naquele momento os termos usados para restrições sociais eram definidos por softwares e inseridos neles os documentos RDF. O que foi visto como um modo de inicializar a Web Semântica com termos úteis, pois a forma da web era o que caracterizaria a própria Web Semântica, assim, as coisas na web seriam definidas em relação a outras coisas consideradas úteis e estáveis. Sendo impossível a criação de uma ordem global entre especificações mais básicas e menos básicas da Web Semântica, pois isso interromperia o dimensionamento da web.

Assim, embora o conceito de documentos fundamentados em um determinado conjunto de bases seja importante para a interoperabilidade, não deve ser visto como um objetivo forçar a Web Semântica a uma estrutura acíclica. Não haverá um único sistema decimal Dewey para a Web Semântica. Os conceitos de especificações estáveis bem definidas ainda serão essenciais. O mesmo acontecerá com o respeito pelas definições dos termos. A diferença será que qualquer um escolherá seu próprio conjunto de idiomas que considera "básico" e encontrará maneiras de definir outros idiomas que encontrar em esses termos. Uma rica teia de conversões, traduções crescerá para suportar isso. A rede de confiança fornecerá ferramentas para navegar e selecionar dessa rede de maneira segura. (Berners-Lee, 1999, tradução nossa).

Em 1992, em outra publicação de Tim Berners-Lee, chamado "*Generic Resources*", define os princípios arquitetônicos acerca dos recursos informacionais. Dessa forma, define que uma URI representa um recurso e um "recurso" é uma entidade conceitual. Um recurso específico representado eletronicamente pode corresponder a apenas uma possível representação de fluxo de bits, como um objeto digital representando um texto é codificado num fluxo de bits único para cada item. Por outro lado, um recurso genérico pode ser especificado em um único fluxo de bits, mas também em outras representações específicas, criando uma relação genérica e outras específicas.

Nesse caso, podem existir outros URIs que identificam um recurso mais especificamente. Essas outras URIs também identificam recursos, e há uma relação de genericidade entre o recurso genérico e o relativamente específico. Por exemplo, recursos sucessivamente específicos podem ser

1. A Bíblia
2. A Bíblia, Versão King James
3. A Bíblia, KJV, em inglês
4. Uma renderização ASCII particular da Bíblia KJV em inglês (Berners-Lee, 1992, tradução nossa).

Assim, o agente que criou o URI é quem determinará em que medida aquele recurso é genérico ou específico. O que se configura como uma genericidade multinível existente nos livros e documentos eletrônicos. As dimensões de genericidade foram divididos em Tempo, Linguagem, Tipo de Conteúdo e Target Medium.

Com a evolução dos protocolos da web, surgiu a demanda de estabelecer uma abordagem sobre o que as URIs HTTP identificavam, esse era um assunto que foi pouco abordado nas especificações anteriores. Esse fato gerou um documento produzido por Berners-Lee (2002) que hoje tem a posição de visualização pessoal no site do W3C, não sendo uma recomendação e nem uma nota técnica, intitulado “O que os URIs HTTP identificam?” que mostra a abordagem do autor e sua visão naquele momento, sobre problemas específicos no domínio da filosofia da linguagem. A questão arquitetônica tratado no texto foi decidida quase dois anos após essa publicação, pelo W3C TAG (Technical Architecture Group) e descrito em uma breve nota posterior. O uso dos URIs em sistemas formais e a utilização das propriedades do RDF para representar e descrever “coisas” na web, gerou essa demanda de problematizar o que os URIs identificam: documentos, páginas ou recursos?

O funcionamento da web é devido ao fato que um HTTP URI pode obter uma representação do documento. A codificação em fluxo de bits gera representações, por exemplo, em telas legíveis por humanos, esse fluxo de bits representa espaços de informação. A Web Semântica vai além desses espaços de informações e passou a enfatizar os conteúdos dos documentos eletrônicos e no universo da linguagem humana e das máquinas.

1. As páginas da web podem comparar as obras de um coral renascentista com sucessos pop do jazz e discutir se os porcos têm asas. Documentos processáveis por máquinas podem codificar informações sobre sapatos, navios e lacres. Até recentemente, os padrões de protocolo da Internet com os quais a web é construída tinham pouco a dizer sobre essas coisas. Eles estavam preocupados apenas com o lado legível por humanos, então eram pessoas, lendo linguagem natural (não especificações de internet) que formou e

comunicou os conceitos neste nível. Hoje em dia, no entanto, as linguagens da Web Semântica permitem que informações sejam expressas não apenas sobre URIs, portas TCP e documentos, mas também sobre conceitos arbitrários - os sapatos, os navios e o lacre, e se os porcos têm asas (Berners-Lee, 2002, tradução nossa).

Como já dito, foi o aumento do número de sistemas formais da Web Semântica que usavam URIs para identificar entidades e propriedades RDF, principalmente o surgimento do Dublin Core e do FOAF, que impulsionou a demanda de definir o que e como os URIs identificam: documentos na forma de páginas web ou recursos de informação identificando objetos do mundo real. O que os URI identificam passou por mudanças ao longo do tempo nas especificações, o RDF Working Group produziu em 2014 uma nova versão do RDF, atualizando o que os recursos identificam. Como essa recomendação adiciona recursos a antiga versão RDF, a pesquisa considera importante apresentar as especificações de 2004.

De acordo as especificações técnicas do “RDF *Working Group*”, na primeira ⁵⁴recomendação do W3C (antiga versão RDF), o objetivo do RDF era promover um modo simples de declarar propriedades de recursos, ou seja, afirmações sobre um recurso, por exemplo, páginas da Web. A documentação técnica traz o exemplo para a declaração que uma determinada pessoa criou uma página na web (Manola; Mille, 2004).

<http://www.example.org/index.html> tem um *criador* cujo valor é *John Smith*

Dessa forma, o *assunto* é <http://www.example.org/index.html>, o *predicado* é a palavra "criador" e o *objeto* são as palavras "John Smith".

O RDF é focado em instruções *processáveis por máquina* e para tornar esses tipos de declarações adequados para processamento por máquinas, é necessário:

- um sistema de identificadores processáveis por máquina que permita identificar um assunto, predicado ou objeto em uma declaração sem qualquer possibilidade de confusão com um identificador de aparência semelhante que pode ser usado por outra pessoa na web.
- um formato processável por máquina para representar essas instruções e trocá-las entre máquinas. (RDF Primer, 2001, tradução nossa) (Manola; Mille, 2004, tradução nossa).

É possível criar um identificador URI para referenciar qualquer coisa, incluindo:

- coisas acessíveis pela rede, como um documento eletrônico, uma imagem, um serviço (por exemplo, "o boletim meteorológico de hoje para Los Angeles") ou uma coleção de outros recursos.

⁵⁴ <https://www.w3.org/TR/rdf-primer/>

- coisas que não são acessíveis pela rede, como seres humanos, corporações e livros encadernados em uma biblioteca.
- conceitos abstratos que não existem fisicamente, como o conceito de um "criador" (Manola; Mille, 2004, tradução nossa).

A documentação técnica traz o conceito de URI reference (URIref) que se difere das URIs convencionais. As URIs fornecem um mecanismo de identificação geral, apropriados para identificar qualquer coisa. O RDF emprega os URIs para identificar os assuntos, predicados e objetos nas instruções. De maneira que, o RDF usa URI references para definir seus assuntos, predicados e objetos. Uma URI reference (URIref) é um URI, juntamente com um identificador de fragmento opcional no final. O “RDF define um recurso como qualquer coisa que seja identificável por uma referência de URI e, portanto, o uso de URIrefs permite que o RDF descreva praticamente qualquer coisa e declare relacionamentos entre essas coisas” (Manola; Mille, 2004, tradução nossa). Os autores enfatizam que por uma questão de ambiguidade, são duas coisas diferentes: identificadores para documentos da web (páginas web) e identificadores para outros recursos. Cada URIs destina-se a identificar apenas um deles, portanto, um URI não pode representar um documento da web e um objeto do mundo real.

Na web tradicional as páginas são documentos da web e cada documento da web tem seu próprio URI. Do mesmo modo, os URIs são usados principalmente para vincular e acessar documentos da web em um navegador, conseqüentemente, não existia a demanda de se criar uma noção de identidade dos recursos, a URL simplesmente identificava o documento da web por completo do jeito representado num navegador. Por outro lado, na Web Semântica, todas as informações devem ser expressas como declarações sobre recursos identificados por *Uniform Resource Identifiers* (URIs), essa abordagem de modelagem é central para o RDF. Logo, URIs são usadas para coisas que não são páginas da web, como pessoas, produtos, lugares, ideias e conceitos, como classes de ontologia (Sauermann; Cyganiak, 2008).

Conseqüentemente, existem duas soluções que atendem aos requisitos do W3C para identificar objetos do mundo real: 303 URIs e hash URIs e a aplicação de cada uma depende do contexto. Existem cenários⁵⁵ de implantação nos quais os dados RDF e os dados HTML são servidos separadamente, como um documento RDF/XML independente com um documento HTML (Sauermann; Cyganiak, 2008).

⁵⁵ Os metadados também podem ser embutidos em HTML, utilizando tecnologias como RDFa, microformatos e outros documentos aos quais os mecanismos GRDDL podem ser aplicados. Nesses casos, os dados RDF são extraídos do documento HTML retornado.

Posteriormente, o RDF Working Group produziu uma recomendação do W3C para uma nova versão do RDF que adiciona recursos a versão de 2004, onde foi definido o RDF a partir de uma atualização sobre o que identificam os recursos.

Desse modo, definiram que “o RDF é uma estrutura para expressar informações sobre recursos. Os recursos podem ser qualquer coisa, incluindo documentos, pessoas, objetos físicos e conceitos abstratos” (Schreiber; Raimond, 2014, tradução nossa). O modelo de dados RDF permite realizar declarações sobre recursos em formato simples, sempre com a seguinte estrutura:

<assunto> <predicado> <objeto>

De maneira que, como uma instrução RDF expressa uma relação entre dois recursos, a relação é ordenada de forma direcional (do sujeito para o objeto), em RDF todos os tipos de relações, são chamadas de propriedade e as declarações sobre um recurso tem nome de triplos.

O mesmo recurso é geralmente referenciado em vários triplos, de forma que um recurso pode ser o sujeito de vários triplos, e ao mesmo tempo, estar na posição de objeto de outro triplo diferente. Essa capacidade possibilita encontrar conexões entre as triplas, sendo possível visualizar os triplos como um grafo conexo. Os grafos consistem em nós e arcos. Os sujeitos e objetos das triplas são os nós do grafo; os predicados os arcos (Schreiber; Raimond, 2014).

Na nova recomendação de 2014, o termo URI passou ser chamado por sua generalização o IRI⁵⁶, que é uma abreviação de "*International Resource Identifier*". Um IRI identifica um recurso, sendo as URLs (*Uniform Resource Locators*) usadas como endereços da web uma forma de IRI. IRIs podem aparecer em todas as três posições de um triplo e usadas para identificar recursos como documentos, pessoas, objetos físicos e conceitos abstratos.

Os IRIs são identificadores globais, permitindo a reutilização do IRI para identificar a mesma entidades ou propriedades em outro grafo do conhecimento, ontologia ou vocabulário, significa que o identificador não está restrito ao contexto do bando de dados convencional, mas pode ser reutilizado globalmente na sua organização ou fora dela. Por exemplo, a propriedade `http://xmlns.com/foaf/0.1/knows` do vocabulário FOAF tem esse IRI usado por muitas pessoas como uma propriedade RDF para indicar uma relação de conhecimento entre as pessoas (Schreiber; Raimond, 2014, tradução nossa).

Na nova recomendação de 2014, foi definido que o RDF é agnóstico sobre o que o IRI representa. “No entanto, IRIs podem receber significado por vocabulários ou convenções

⁵⁶ IRI é uma generalização de URI (*Uniform Resource Identifier*), permitindo que caracteres não-ASCII sejam usados na cadeia de caracteres IRI. IRIs são especificados no RFC 3987.

particulares. Por exemplo, a DBpedia usa IRIs da forma `http://dbpedia.org/resource/Name` para denotar a coisa descrita pelo artigo correspondente da Wikipedia” (Schreiber; Raimond, 2014, tradução nossa).

Dessa forma, a discussão se um IRI representa um recurso na forma de documento web ou um objeto do mundo real, deve ser definido pelos domínios e especificações dos produtores de KOS LOD. Ou seja, não fica estabelecido nas especificações RDF quais tipos de conceito um recurso identificado por um IRI pode denotar, podendo representar documentos, seres concretos ou conceitos abstratos.

O termo *Linked Data* refere-se a um conjunto de práticas recomendadas para publicação de dados estruturados na web. Esses princípios foram cunhados por Tim Berners-Lee em nota sobre o projeto *Linked Data*. Os princípios são:

1. Use URIs como nomes para coisas.
2. Use URIs HTTP para que as pessoas possam pesquisar esses nomes.
3. Quando alguém procura um URI, forneça informações úteis.
4. Inclui links para outros URIs. para que eles possam descobrir mais coisas.

O termo LOD refere-se a todos os dados publicados na web de acordo com um conjunto de melhores práticas, os Princípios de Dados Vinculados.

4.5 ONTOLOGIA LEVE (*LIGHTWEIGHT ONTOLOGY*)

A evolução da web e o mecanismo do Google popularizaram a indexação automática, chegando ao ponto de ser anunciado a obsolescência de vocabulários em favor das modernas soluções tecnológicas (Vatant, 2010, *apud* Castro, 2011). Contudo, as tecnologias baseadas em HTML e hipertextos demonstraram limitações e inadequações, como, por exemplo: excesso de informações não estruturadas, a opacidade do conteúdo informativo disponível nos bancos de dados, a utilização de cadeias de caracteres (*strings*), ao invés de conceitos, condições que demonstram a necessidade de informações semânticas na web (Castro, 2011).

Nesse sentido, é importante ressaltar que no ambiente da web o significante é a palavra e a busca é realizada por texto completo ou *tag*⁵⁷, ademais a indexação não faz o uso de

⁵⁷ Definição de TAG no contexto da computação: “an instruction appended to a piece of text in a markup language in order to specify how it is displayed or interpreted. A word, phrase, or name used to identify digital content such as blog and social media posts as belonging to a particular category or concerning a particular person or topic.”you can easily add tags to photos en masse”. Disponível em: <https://languages.oup.com/>

vocabulários controlados intermediários, acrescenta-se que a recuperação de documentos é realizada com as mesmas palavras-chave da busca, em vista disso, apresenta os problemas de polissemia e sinonímia e os resultados desse sistema de busca são dispersos, incoerentes e desvinculados (González, 2013).

De tal modo, existe a demanda por metodologias para organização do conhecimento de forma semântica na web, com ferramentas apoiadas pelo *linked data* e pelos LOD na estruturação de dados e informação. Assim, com objetivo de vincular dados estruturados mediante SOC, nos últimos anos surgiram vários formatos vinculados à publicação de novos padrões de elaboração de tesouros, como no quadro 29:

Quadro 26 - Novos padrões de elaboração de tesouros

BS 8723 XML Schema, vinculados a BS 8723: Vocabulários estruturados para recuperação de informação
ISO 25964 XML Schema, vinculados a ISO 25964: Thesauri e interoperabilidade com outros vocabulários
SKOS (Simple Knowledge Organization System)

Fonte: Adaptado de Castro (2011, tradução nossa).

Sobre as transformações nos tesouros e SOC semânticos, Hjørland (2015) coloca uma questão importante: se um tipo de SOC pode ser reutilizado como a base principal para estabelecer outro tipo. O autor ainda coloca se seria possível, por exemplo, “um sistema de classificação ser transformado em um tesouro? Ou um dicionário de sinônimos pode ser transformado em uma ontologia? A resposta a esta pergunta é de interesse prático e teórico” (Hjørland, 2015, p. 119, tradução nossa).

O SKOS tem uma grande associação com os SOC, oriundos das práticas da área da Biblioteconomia e Ciência da informação (BCI). Além disso, o SKOS também conecta o cenário da BCI com as comunidades da Web Semântica, ao transferir os modelos dos SOC para o contexto de tecnologia da Web Semântica. Ademais, SKOS é tido como um modo econômico, do ponto de vista dos recursos, para migrar SOC já existentes para RDF.

O modelo de dados SKOS aborda os SOC como um esquema que compreende um conjunto de conceitos. Esses esquemas de conceito SKOS e os conceitos SKOS são identificados por URIs. Dentre as iniciativas *linked data* para publicação de SOC, como tesouros, taxonomias, classificações ou listas de autoridade, o SKOS está se tornando o modelo mais comum e vem sendo desenvolvido desde 2013, a partir da iniciativa de bibliotecas e da comunidade da Web Semântica (Castro, 2011). O quadro 30 apresenta uma visão geral do SKOS de acordo sua especificação normativa:

Quadro 30 - SKOS (Simple knowledge organization system)

Visão geral do SKOS	SKOS é um modelo de dados comum para sistemas de organização de conhecimento, como tesouros, esquemas de classificação, sistemas de cabeçalho de assunto e taxonomias. Com o SKOS, um sistema de organização do conhecimento pode ser expresso como dados legíveis por máquina.
Espaço de nome e vocabulário SKOS	O URI do namespace SKOS é: http://www.w3.org/2004/02/skos/core# O vocabulário SKOS é um conjunto de URIs.
skos:Concept Class	A classe skos:Concept é a classe de conceitos SKOS. Pode ser visto como uma ideia ou noção; uma unidade de pensamento. É útil para descrever a estrutura conceitual ou intelectual de um sistema de organização do conhecimento e quando se refere a ideias ou significados específicos estabelecidos dentro de um SOC.
Concept Schemes	Um esquema de conceito SKOS pode ser visto como uma agregação de um ou mais conceitos SKOS. As relações semânticas (ligações) entre esses conceitos também podem ser vistas como parte de um esquema de conceito. É útil ao lidar com dados de uma fonte desconhecida e ao lidar com dados que descrevem dois ou mais sistemas diferentes de organização do conhecimento.
Etiquetas lexicais	Um rótulo léxico é uma string de caracteres UNICODE, como "amor romântico" ou "れんあい", em um determinado idioma natural, como inglês ou japonês (escrito aqui em hiragana). SKOS permite que seja feita uma distinção entre os rótulos lexicais preferidos, alternativos e "ocultos" para qualquer recurso dado. Os rótulos preferidos e alternativos são úteis ao gerar ou criar representações legíveis por humanos de um sistema de organização do conhecimento. Esses rótulos fornecem as pistas mais fortes sobre o significado de um conceito SKOS.
Notações	Uma notação é uma sequência de caracteres como "T58.5" ou "303.4833" usada para identificar exclusivamente um conceito dentro do escopo de um determinado esquema de conceito.
Propriedades da documentação	As notas são usadas para fornecer informações relacionadas aos conceitos SKOS.
Relações Semânticas	As relações semânticas do SKOS são elos entre os conceitos do SKOS, em que o elo é inerente ao significado dos conceitos vinculados. O SKOS distingue entre duas categorias básicas de relação semântica: hierárquica e associativa
Concept Collections	As coleções de conceitos SKOS são grupos rotulados e / ou ordenados de conceitos SKOS. As coleções são úteis onde um grupo de conceitos compartilha algo em comum e é conveniente agrupá-los sob um rótulo comum ou onde alguns conceitos podem ser colocados em uma ordem significativa
Propriedades de mapeamento	skos:closeMatch, skos:exactMatch, skos:broadMatch, skos:narrowMatch, skos:relatedMatch. Essas propriedades são usadas para estabelecer links de mapeamento (alinhamento) entre conceitos SKOS em diferentes esquemas de conceito, onde os links são inerentes ao significado dos conceitos vinculados.

Fonte: Adaptado de Miles e Bechhofer (2009, tradução nossa).

Ainda no quadro 7, a classe de conceitos SKOS (skos:Concept), introduz a noção que um conceito SKOS pode ser visto como uma ideia ou uma unidade de pensamento, o que conecta esse modelo as metodologias usadas nos SOC. No caso, esse conceito ou ideia é nomeado com um URI e pode ser representado por diferentes rótulos de forma multilíngue e

se relacionar, por hierarquia ou associação, com outros conceitos do próprio vocabulário ou de outro vocabulário controlado (Castro, 2011). Ao mesmo tempo, o SKOS interpreta os SOC como um esquema de conceitos, chamado “*Concept Schemes*”, nesse sentido, apesar da possibilidade de os conceitos serem criados e usados como entidades independentes, é comum que os termos da indexação provenham de vocabulários estruturados, como tesouros ou esquemas de classificação, a classe `skos:conceptScheme` permite representar esses SOC (vocabulários, tesouros, esquemas de classificação) (Miles; Bechhofer, 2009).

Segundo Pastor-Sanchez; Mendez, Rodríguez-Munõz (2009), em relação ao SKOS, existem duas abordagens principais para considerar como central: um tesouro baseado em termos e um tesouro baseado em conceitos. O primeiro é um conjunto de termos interligados por uma rede de diferentes tipos de relacionamento semântico, enquanto o segundo é um conjunto de conceitos relacionados por relações semânticas semelhantes às de um tesouro baseado em termos. O SKOS é uma ontologia definida com OWL Full que estrutura os dados na forma de triplos que podem ser codificados em qualquer sintaxe válida para RDF. A conversão de um tesouro em conhecimento formal pode ser um processo custoso, pois os tesouros são desenvolvidos para fornecer um mapa intuitivo de como os temas são organizados nos processos de classificação e busca de objetos, e não fornecem um modelo de dados que possa ser facilmente aplicado (Pastor-Sanchez; Mendez, Rodríguez-Munõz, 2009).

Tesouros baseados em termos envolvem problemas de aplicação na indexação ao realizar operações para alterar termos preferidos e não preferidos. Assim, se um descritor que foi usado para indexar um documento se tornar um não-descritor, os índices que ligam esse descritor aos documentos devem ser reorganizados. Isso não ocorre no caso dos tesouros conceituais, pois os processos de indexação independem das mudanças terminológicas do tesouro. Nesse caso, os documentos são associados a conceitos, não a termos, alterações envolvendo termos preferenciais e não preferenciais não impactam na indexação (Pastor-Sanchez; Mendez, Rodríguez-Munõz, 2009, tradução nossa).

As práticas e recomendações formais para a publicação de dados estruturados de forma semântica na web, como o *linked data* e LOD, operam na organização do conhecimento a partir de esquemas semânticos visando vincular e publicar dados estruturados por meio de SOC. Na web, os SOC se estenderam no contexto dos dados abertos e conectados, com a publicação de conjuntos de dados nos padrões LOD (W3C), disponibilizados para download nas páginas das instituições na forma *datasets* LOD. O conceito de KOS LOD (Zeng; Mayr, 2019), engloba a coleta, modelagem, enriquecimento e publicação no modelo de dados da Web Semântica RDF,

para que se tornem legíveis por máquinas, o conjunto dessas etapas forma o enriquecimento semântico dos SOC.

De acordo a Zeng e Mayr (2018) um vocabulário KOS LOD deve seguir os princípios do *Linked Data* e ter acesso aberto, por serem usados como **vocabulários de valor** que especificam recursos que representam conceitos usados como valores dos metadados. Os exemplos incluem: tesouros, listas de códigos, listas de termos, esquemas de classificação, listas de títulos de assuntos, arquivos de autoridade, dicionários geográficos digitais, esquemas de conceito e outros tipos de SOC. São exemplos de sistemas KOS LOD, provedores de serviços (PS)⁵⁸ de vocabulários como o EuroVoc e instituições individuais fornecedoras de todos os serviços ligados aos vocabulários produzidos na instituição, como a Library of Congress Linked Data Services e o Getty Vocabularies LOD.

No processo denominado de enriquecimento semântico, o KOS LOD é usado como **vocabulário de valor**, visto que, nos modelos baseados em RDF, os triplos estruturam a sintaxe: **recurso** (*resource*), **tipo de propriedade** (*property-type*) e **valor da propriedade** (*property-value*) e para isso utilizam “uma lista controlada de valores permitidos para um elemento em dados estruturados” (Zeng; Mayr, 2018, p. 210, tradução nossa). Segundo os autores, um vocabulário de valor identifica os valores para registros de metadados. Os quais são recursos que representam, por exemplo, entidades de tópicos, estilos de arte ou entidades nomeadas, dessa forma, os SOC atuam como roteiros semânticos de domínios com dados estruturados num modelo semântico que norteiam indexadores e usuários, humanos ou máquinas.

O conceito de KOS LOD identifica todos os vocabulários de valor e “ontologias leves” que fazem parte do ambiente da Web Semântica. Onde o termo "ontologias leves" refere-se àquelas que usam classes e propriedades provenientes de ontologias para estruturar o SOC tradicional e fornecer estrutura terminológica mínima para representar uma conceituação de domínio.

No contexto dos KOS LOD, o SKOS é fortemente usado para modelar diferentes tipos de SOC na Web Semântica. O SKOS é um modelo de dados comum para SOC, como tesouros, esquemas de classificação, sistemas de cabeçalho de assunto e taxonomias. Dessa forma, a pesquisa se aprofundou na documentação⁵⁹ com as especificações normativas e a guia dos

⁵⁸ Provedores de serviço (SP) desenvolvem estratégias e tecnologias para garantir a disponibilidade, a interoperabilidade, estabilidade e escalabilidade dos SOC e seus respectivos serviços.

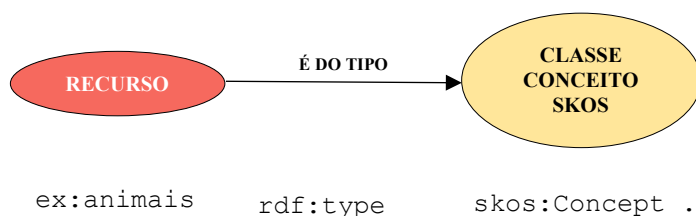
⁵⁹ <https://www.w3.org/TR/skos-reference/#collections> (especificação normativa) e <https://www.w3.org/TR/2009/NOTE-skos-primer-20090818/#seccollections> (guia do usuário).

usuários, publicada pelo W3C, para entender como o modelo de dados SKOS é usado para compartilhar e vincular SOC na web e pode ser expresso como dados legíveis por máquina. Assim, investigar os níveis conceituais, terminológicos e lexicais do modelo SKOS.

O SKOS espelha as relações usadas em tesauros (ISO25964), a partir de três propriedades padrões: *skos:broader*, *skos:narrower* e *skos:related* e pode ser expresso como dados legíveis por máquina. A abordagem central no SKOS é o conceito e suas relações semânticas, a partir da aplicação de rótulos para definir as relações hierárquicas entre dois conceitos (*skos:broader*, *skos:narrower*) e as relações associativas entre dois conceitos (*skos:related*), os quais são inerentemente associadas, mas não hierárquicas. As propriedades de mapeamento *skos:closeMatch*, *skos:exactMatch*, *skos:broadMatch*, *skos:narrowMatch* e *skos:relatedMatch*, estabelecem links de mapeamento (alinhamento) entre conceitos SKOS em diferentes esquemas de conceito.

O elemento essencial do vocabulário SKOS é o *conceito*. Os conceitos são as *unidades de pensamento*, ideias, significados ou categorias de objetos e eventos, designados por termos que representam os conceitos num domínio do conhecimento dos SOC. Desse modo, os conceitos como entidades abstratas são independentes dos termos usados para rotulá-los. Ainda, no SKOS os conceitos são representados pela classe *skos:Concept*, que afirma que um determinado recurso é um conceito, ao criar ou reutilizar um Identificador Uniforme de Recursos (URI) para identificar exclusivamente o conceito e ao afirmar em RDF, usando a propriedade *rdf:type*, que o recurso identificado por este URI é do tipo *skos:Concept* (Issac; Summer, 2009). Como na figura 32, num triplo RDF sempre o link aponta para declarar o objeto da sentença:

Figura 32 - Triplo RDF



Fonte: Elaborado pela autora.

Após a afirmação que um recurso é do tipo classe conceito SKOS (*skos:Concept*), a caracterização do conceito é realizada via rótulos em linguagem natural. O SKOS fornece três propriedades para anexar rótulos a recursos

conceituais: *skos:prefLabel*, *skos:altLabel* e *skos:hiddenLabel*, que fornecem rótulos lexicais. Um rótulo lexical é uma *string* (cadeia de caracteres) UNICODE, como "amor romântico" ou "れんあい", em um idioma natural específico e são úteis ao gerar ou criar representações legíveis por humanos de um sistema de organização do conhecimento (Miles; Bechhofer, 2009). Importante ressaltar que rótulos lexicais não pertencem à classe de conceitos SKOS, ou seja, não se caracterizam como recursos e não possuem URI.

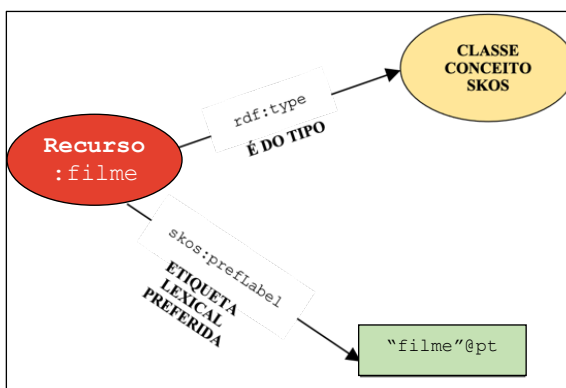
Desse modo, cria-se um relacionamento de denotação unívoco a uma *string* (cadeia de caracteres) que auxilia na pesquisa do termo, ademais, as propriedades lexicais *skos:prefLabel*, *skos:altLabel* e *skos:hiddenLabel*, são propriedades consideradas disjuntas, ou seja, é um erro se um conceito tiver um mesmo literal (*string*) tanto como rótulo preferencial quanto como rótulo alternativo. Assim, a três propriedades fornecem rótulos simples usados para vincular a classe *skos:concept* a um literal simples RDF, sendo uma cadeia de caracteres (por exemplo, "love") combinada com uma tag de idioma opcional ("love"@en) (Issac; Summer, 2009).

A propriedade *skos:prefLabel* atribui um rótulo léxico preferencial a um recurso e representa os termos usados como *descritores* em sistemas de indexação. O SKOS permite uma forma simples de rotulagem multilíngue, usando a marca de idioma de um rótulo léxico para restringir seu escopo a um idioma específico. O quadro 31 e a figura 33, a seguir, mostram o rótulo preferencial em inglês e outro em francês:

Quadro 27 - Rótulo lexical preferencial

```
ex:animals rdf:type skos:Concept;
           skos:prefLabel "filme"@pt;
           skos:prefLabel "movie"@en.
```

Figura 33 - Rótulo lexical preferencial



Fonte: Elaborado pela autora.

Fonte: Elaborado pela autora.

A noção de rótulo *preferencial* implica que um recurso pode ter apenas um desses rótulos por tag de idioma, visto que, mesmo o modelo de dados SKOS não restringindo formalmente, não é recomendável que dois conceitos recebam o mesmo rótulo lexical preferencial para qualquer tag de idioma.

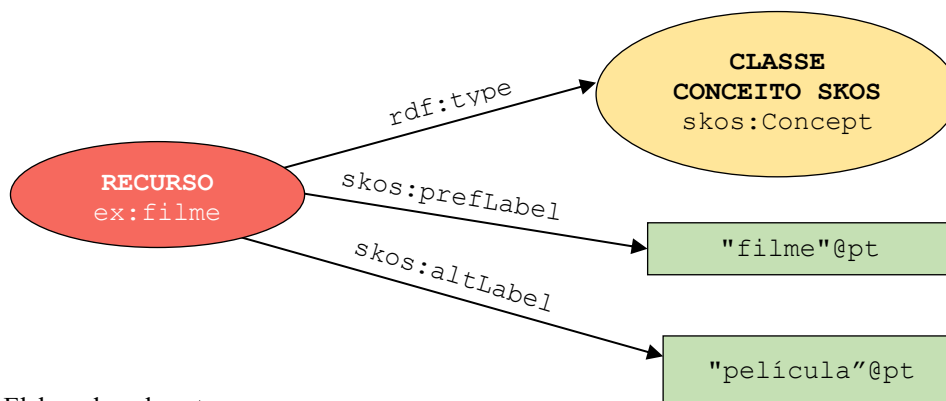
A propriedade `skos:altLabel` atribui um rótulo lexical alternativo a um conceito, sendo aplicável, por exemplo, no caso da representação de sinônimos (Issac; Summer, 2009), mostrado no quadro 32 e figura 34:

Quadro 28 - Rótulo lexical alternativo SKOS

```
ex:filme rdf:type skos:Concept;
          skos:prefLabel "filme"@pt ;
          skos:altLabel "película"@pt ;
          skos:prefLabel "movie"@en ;
          skos:altLabel "film"@en .
```

Fonte: Elaborado pela autora.

Figura 34 - Rótulo lexical alternativo SKOS



Fonte: Elaborado pela autora.

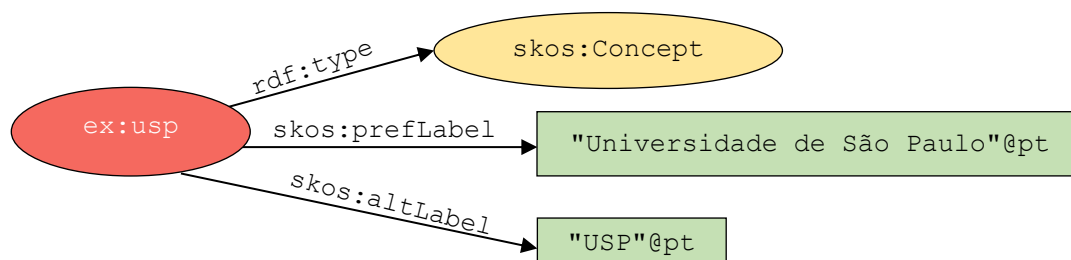
Além disso, os *quase-sinônimos*, abreviaturas e acrônimos podem ser representados da mesma forma do quadro 33 e figura 35:

Quadro 29 - Rótulo lexical alternativo usado para abreviaturas

```
ex:usp rdf:type skos:Concept ;
        skos:prefLabel "Universidade de São Paulo"@pt ;
        skos:altLabel "USP"@pt .
```

Fonte: Elaborado pela autora.

Figura 35 - Rótulo lexical alternativo usado para abreviaturas



Fonte: Elaborado pela autora.

Os rótulos ocultos são úteis para pesquisas baseadas em texto num sistema de organização do conhecimento, pois as palavras incorretas são comparadas com um rótulo oculto, apontando o conceito relevante, mas o rótulo oculto não ficará visível para o usuário (Miles; Bechhofer, 2009). Rótulos ocultos podem ser usados para incluir variantes com erros ortográficos de outros rótulos lexicais (Issac; Summer, 2009). Por exemplo:

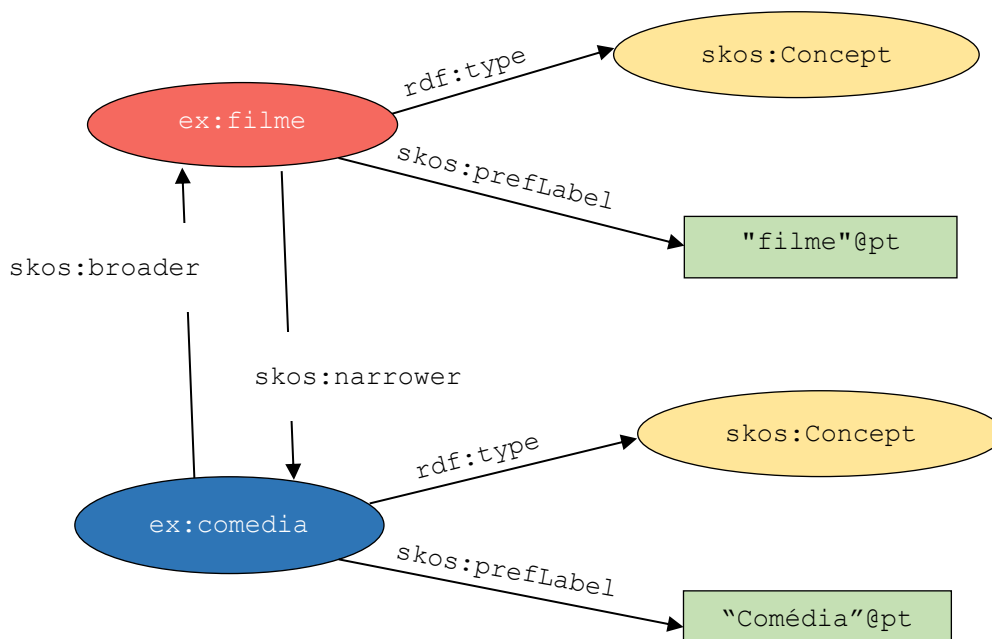
```
ex:alcoollista rdf:type skos:Concept ;
  skos:prefLabel "alcoollista"@pt ;
  skos:altLabel "alcoólatra"@pt ;
  skos:hiddenLabel "bêbado"@pt .
```

O SKOS espelha as relações semânticas usadas em tesouros (ISO25964), a partir de três propriedades padrões: *skos:broader*, *skos:narrower* e *skos:related*. As duas primeiras possibilitam a representação das relações hierárquicas: *skos:broader* é a propriedade cujo valor é o conceito mais amplo e a *skos:narrower* é usado para afirmar o conceito mais específico:

```
ex:genero rdf:type skos:Concept ;
  skos:prefLabel "genêro"@pt ;
  skos:narrower ex:comedia.

ex:comedia rdf:type skos:Concept ;
  skos:prefLabel "comédia"@pt ;
  skos:broader ex:filme .
```

Figura 36 - Relações semânticas SKOS



Fonte: Elaborado pela autora.

E a última propriedade *skos:related*, possibilita a representação das relações associativas e não hierárquicas, estabelece ligações de simetria entre os conceitos (Issac; Summer, 2009):

```
ex:birds rdf:type skos:Concept ;
  skos:prefLabel "birds"@en;
  skos:related ex:ornithology.
ex:ornithology rdf:type skos:Concept;
  skos:prefLabel "ornithology"@en.
```

Entretanto, *skos:related* não é uma propriedade transitiva, não sendo possível com SKOS fazer a vinculação, como mostrado a seguir (Miles; Bechhofer, 2009):

```
<A> skos:related <B> .
<B> skos:related <C> .
```

não implica

```
<A> skos:related <C> .
```

Além das propriedades que representam as relações entre conceitos, existem as propriedades de mapeamentos, como *skos:closeMatch* e o *skos:exactMatch*, usadas para estabelecer links de mapeamento (alinhamento) entre conceitos SKOS em diferentes esquemas de conceito. Ademais, o *skos:closeMatch* e o *skos:exactMatch*, podem inicialmente parecerem iguais, pelo fato de ambas serem usadas para vincular conceitos SKOS em esquemas diferentes (Miles; Bechhofer, 2009). Assim, o *skos:closeMatch*, indica que dois conceitos, ou seja, dois

recursos e duas URI, são suficientemente semelhantes para serem intercambiáveis em alguns aplicativos de recuperação de informação:

```
ex1:animal rdf:type skos:concept ;
skos:prefLabel "animal" @en .

ex2:animaux rdf:type skos:concept ;
skos:prefLabel "animaux" @fr .

ex4:animals rdf:type skos:concept ;
      skos:closeMatch ex1:animal ;
      skos:closeMatch ex2:animaux ;
```

Porém, `skos:closeMatch` não é uma propriedade transitiva.

```
<A> skos:closeMatch <B>
<B> skos:closeMatch <C>
```

não implica

```
<A> skos:closeMatch <C>
```

Por outro lado, uma ligação `skos:exactMatch` possui um alto grau de confiança para que dois conceitos, ou seja, dois recursos e duas URI, sejam intercambiáveis através de uma extensa gama de aplicativos de recuperação de informações (Miles; Bechhofer, 2009).

```
ex3:animals rdf:type skos:concept ;
skos:prefLabels "animals" @en .

ex4:animals rdf:type skos:concept ;
      skos:exactMatch ex3:animals .
```

Também, `skos:exactMatch` é uma propriedade transitiva:

```
<A> skos:exactMatch <B>
<B> skos:exactMatch <C>
```

Logo implica

```
<A> skos:exactMatch <C>
```

O SKOS possibilita representar diferentes SOC a partir de vocabulários, como os tesauros e esquemas de classificação, usando a classe `skos:ConceptScheme`. O exemplo a seguir

em Turtle⁶⁰, mostra como definir um recurso de esquema de conceito (representando relações semânticas de um tesouro) e descrever esse recurso usando as propriedades `dct:title` e `dct:creator` do Dublin Core:

```
ex:filmThesaurus rdf:type skos:ConceptScheme ;
    dct:title "Simple film thesaurus" ;
    dct:creator ex:cavalcanteSantos .
```

Em seguida a criação do recurso de esquema de conceito (*Concept Scheme*), é possível vincular conceitos ao esquema usando a propriedade `skos:inScheme`:

```
ex:experimentalFilm rdf:type skos:Concept ;
    skos:inScheme ex:filmThesaurus .

ex:surrealistFilm rdf:type skos:Concept ;
    skos:broader ex:experimentalFilm ;
    skos:inScheme ex:filmThesaurus .

ex:educationFilm rdf:type skos:Concept;
    skos:inScheme ex:filmThesaurus .
```

No caso, no grafo RDF acima, os conceitos `ex:experimentalFilm`, `ex:surrealistFilm` e `ex:educationFilm` são do tipo (instâncias de classe) `skos:Concept` e são conceitos do esquema `ex:filmThesaurus`. Ademais, o segundo grafo RDF tem uma afirmação indicando que `ex:experimentalFilm` tem uma relação semântica mais ampla (`broader`), é o termo mais amplo, para `ex:surrealistFilm`.

Para o acesso aos pontos de entrada de hierarquias de conceito mais amplos/específicos num esquema de conceitos, o SKOS define uma propriedade `skos:hasTopConcept`. Esta propriedade permite vincular um esquema de conceito aos seus conceitos mais gerais, como no exemplo do tesouro filmico a seguir:

```
ex:filmThesaurus rdf:type skos:ConceptScheme ;
    skos:hasTopConcept ex:experimentalFilm ;
    skos:hasTopConcept ex:educationFilm .
```

Devido ao uso de URI é possível referenciar inequivocamente cada conceito SKOS em qualquer aplicação SKOS. Desse modo, criar relações semânticas entre conceitos a partir de mapeamentos essenciais para aplicações e ferramentas de recuperação de informação que

⁶⁰ O Terse RDF Triple Language (Turtle) é um formato de arquivo e sintaxe para expressar dados no modelo de dados do Resource Description Framework (RDF).

utilizam mais de um SOC ao mesmo tempo, e que possuem domínios sobrepostos (Issac; Summer, 2009). Uma característica importante do mapeamento é a interoperabilidade entre vocabulários de contextos diferentes. Com o SKOS é possível afirmar que dois conceitos de esquemas diferentes “têm significados comparáveis e especificar como esses significados se comparam, mesmo que venham de contextos diferentes e possivelmente seguem princípios de modelagem diferentes”, os mapeamentos conceituais são uma das principais vantagens de usar o SKOS para publicar SOC na Web Semântica (Issac; Summer, 2009).

Devido a isso, dois conceitos de diferentes esquemas de conceito também podem ser mapeados usando as propriedades: *skos:broadMatch*, *skos:narrowMatch* e *skos:relatedMatch*. No exemplo a seguir estão representados conceitos em dois esquemas de conceitos diferentes, além disso, também está representado o mapeamento entre conceitos de um esquema e conceitos de outro esquema diferente, servindo para mapear conceitos em vocabulários externos.

```
ex1:GenreScheme rdf:type skos:ConceptScheme ;
    dct:title "Scheme Film Genre"@en .

ex1:surrealist rdf:type skos:Concept ;
    skos:prefLabel "Surrealist"@en ;
    skos:inScheme ex1:GenreScheme .

ex1:art rdf:type skos:Concept;
    skos:prefLabel "Art"@en ;
    skos:inScheme ex1:GenreScheme .

ex2:experimentalFilmScheme rdf:type skos:ConceptScheme ;
    dct:title "Experimental Film Scheme vocabulary"@en .

ex1:surrealist rdf:type skos:Concept ;
    skos:inScheme ex2:experimentalFilmScheme .

ex2:film rdf:type skos:Concept ;
    skos:prefLabel "Film"@en ;
    skos:inScheme ex2:experimentalFilmScheme .

ex2:art rdf:type skos:Concept ;
    skos:prefLabel "Art"@en ;
    skos:inScheme ex2:experimentalFilmScheme .
```

A seguir está o mapeamento entre os conceitos do esquema "*Extensive list of film genre*"@en e conceitos do esquema "*Experimental Film Scheme vocabulary*"@en, ambos representados no exemplo acima:

```
ex1:surrealist skos:broadMatch ex2:experimentalFilm.
```

```
ex1:surrealist skos:relatedMatch ex2:film .
ex1:art skos:exactMatch ex2:art.
```

As propriedades `skos:broadMatch` e `skos:narrowMatch` são usadas para estabelecer um link de mapeamento hierárquico entre dois conceitos. A propriedade `skos:relatedMatch` é usada para indicar um link de mapeamento associativo entre dois conceitos. Ou seja, diferentemente de `skos:closeMatch` e `skos:exactMatch`, as propriedades *skos:broadMatch*, *skos:narrowMatch* e *skos:relatedMatch* **não** indicam que dois conceitos possuem semelhança suficiente para serem intercambiáveis, ou seja, usados alternadamente no sistema, mas estabelecem links de mapeamento hierárquico entre dois conceitos de esquemas diferentes.

Além das propriedades de mapeamento entre diferentes esquemas de conceitos, a partir das URI, como já demonstrado acima, é possível reutilizar e compartilhar de forma distribuída os recursos (conceitos), usando a propriedade *skos:conceptscheme*, que cria um esquema de conceitos usando as relações semânticas entre os conceitos para formar uma **rede semântica** em SKOS. Assim, um conceito pode pertencer a mais de um esquema ao mesmo tempo, ao estender um esquema de conceitos e afirmar novos conceitos *vinculados* a conceitos que já foram definidos anteriormente no esquema existente (Issac; Summer, 2009).

A extensão de um SOC pode ser especialmente útil para obter uma melhor cobertura de um domínio ou subdomínio, fazendo o reuso de alguns de seus conceitos. A extensão e a reutilização explícita de SOC também é projetada para cobrir vários domínios e operar em determinados subconjuntos de conceitos. A propriedade `skos:inScheme` é a que permite que um novo esquema de conceito reutilize conceitos existentes (Issac; Summer, 2009).

```
Ex1:GenreScheme rdf:type skos:ConceptScheme ;
  dct:title "Scheme Film Genre"@en .

ex1:comedy rdf:type skos:Concept ;
  skos:prefLabel "Comedy"@en ;
  skos:inScheme ex1:referenceGenreScheme .
```

O criador de outro esquema de conceito dedicado a descrições de filmes pode incluir livremente o conceito `ex1:experimentalFilm` de referência em seu esquema e, em seguida, referenciá-lo da seguinte forma:

```
ex2:subGenreThes rdf:type skos:ConceptScheme ;
  dct:title "The Film Subgenre Thesaurus"@en .

ex1:comedy skos:inScheme ex2:subGenreThes .

ex2:romanticComedy rdf:type skos:Concept ;
```

```

skos:prefLabel "Romantic Comedy"@en ;
skos:broader ex1:comedy;
skos:inScheme ex2:subGenreThes .

ex2:parody rdf:type skos:Concept ;
skos:prefLabel "Parody"@en ;
skos:broader ex1:comedy ;
skos:inScheme ex2:subGenreThes .

```

A classe *skos:Collection* é usada para criar coleções quando um grupo conceitos compartilha algo em comum, e é conveniente agrupá-los sob um rótulo comum (Miles; Bechhofer, 2009). Contudo, a classe *skos:collection* e *skos:concept* são disjuntas, isso significa que o domínio e o alcance das propriedades das relações semânticas (*broader*, *narrower* e *related*) serão sempre *skos:concept*, como mostram os exemplos abaixo. Os primeiros triplos do exemplo não são consistentes com o modelo de dados SKOS:

```

<A> skos:narrower <B> .
<B> rdf:type skos:Collection .

<A> skos:broader <B> .
<B> rdf:type skos:Collection .

<A> skos:related <B> .
<B> rdf:type skos:Collection .

```

O exemplo abaixo é consistente com o modelo de dados SKOS:

```

<A> skos:narrower <B>, <C>, <D> .

<ResourceCollection> rdfs:label "Resource Collection"@en ;
skos:member <B> , <C> , <D> .

```

Uma coleção é definida como um nó em branco, ou seja, não foi atribuído nenhum URI. Esses agrupamentos são normalmente renderizados em tesauros como no exemplo a seguir, a partir de rótulo de nó⁶¹:

```

Action Genre
  <Sub-genres>
    War
    Spy
    Martial Arts

```

⁶¹ Nos tesauros tradicionais um “rótulo de nó” (*Node Label*) agrupa um subconjunto de termos especificando uma característica de divisão na hierarquia e não devem ser usados para fins de recuperação, pois tem uma função organizacional.

Para modelar corretamente a coleção de conceitos, o SKOS apresenta uma `skos:Collection` classe. As instâncias desta classe agrupam conceitos específicos por meio da `skos:member` propriedade, como no exemplo a seguir:

```
ex:film rdf:type skos:Concept ;
        skos:prefLabel "Film"@en .

ex:surrealist rdf:type skos:Concept ;
              skos:prefLabel "Surrealist"@en ;
              skos:broader ex:film .

ex:art rdf:type skos:Concept;
        skos:prefLabel "Art"@en ;
        skos:broader ex:film .

ex:romanticComedy rdf:type skos:Concept ;
                  skos:prefLabel "Romantic Comedy"@en ;
                  skos:broader ex:film .

ex:parody rdf:type skos:Concept ;
           skos:prefLabel "Parody"@en ;
           skos:broader ex:film .

_:b0 rdf:type skos:Collection;
     skos:prefLabel "List's Film Genre"@en;
     skos:member ex:surrealist ;
     skos:member ex:art ;
     skos:member ex:romanticComedy ;
     skos:member ex:parody .
```

A coleção "List's Film Genre" foi definida como um nó em branco, ou seja, nenhum URI definido foi atribuído. Apesar da possibilidade de atribuir URIs para coleções, normalmente não é necessário. A propriedade `skos:prefLabel` pode ser usada com recursos não conceituais, por isso foi atribuído um rótulo léxico à coleção. O `skos:prefLabel` pode ser aplicado para atribuir um rótulo léxico a coleção, pois essa propriedade pode ser usada com recursos não conceituais.

Desse modo, para capturar a ordem dos conceitos em uma coleção, como, por exemplo, listar os conceitos em ordem alfabética ou cronológica, é necessário definir uma coleção ordenada de conceitos, com a classe `skos:OrderedCollection` com a propriedade `skos:memberList`. Esta propriedade vincula uma instância de `skos:OrderedCollection` a um nó (possivelmente em branco) do tipo `rdf:List`. Por exemplo:

```
ex:Adventure rdf:type skos:Concept ;
             skos:prefLabel "Adventure"@en .
```



```

ex:action rdf:type skos:Concept;
  skos:prefLabel "Action"@en .

ex:drama rdf:type skos:Concept ;
  skos:prefLabel "Drama"@en .

_:b0 rdf:type skos:OrderedCollection;
  skos:prefLabel "Most Viewed Film Genres 1995-2022"@en;
  skos:memberList _:b1.
_:b1 rdf:first ex:adventure;
  rdf:rest _:b2.
_:b2 rdf:first ex:action;
  rdf:rest _:b3.
_:b3 rdf:first ex:drama;
  rdf:rest rdf:nil.

```

Como já citado, no modelo de dados SKOS, as coleções (`skos:Collection`) são disjuntas dos conceitos (`skos:Concept`), não sendo possível usar as relações semânticas do SKOS (*Broader*, *narrower* e *Related*), ou seja, não é possível representar relações semânticas entre uma coleção e outros conceitos do vocabulário.

Portanto, agrupar conceitos em coleções, não substitui a necessidade de definir o significado de um conceito pelas relações que estabelece com outros conceitos do esquema de conceitos. Por exemplo, no exemplo acima da coleção "List's Film Genre", todos os conceitos foram explicitamente anexados a um conceito mais genérico usando a propriedade `skos:broader`.

```

ex:surrealist skos:broader ex:film .
ex:art skos:broader ex:film .
ex:romanticComedy skos:broader ex:film .
ex:parody skos:broader ex:film .

```

Uma exibição sistemática (hierárquica) pode então ser gerada incluindo o conceito de agrupamento "List's Film Genre", usando a propriedade hierárquica `skos:broader` e as associações da coleção, mas o processo demanda o desenvolvimento de um algoritmo e a implementação de aplicativos específicos, o que está fora do escopo do modelo SKOS, ou seja, são ações para além do modelo de dados.

Conseqüentemente, isso demonstra que coleções adicionam complexidade às representações das aplicações. Portanto, para alguns casos, por exemplo, quando os SOC são destinados a hierarquias de navegação, “parece mais intuitivo representar ‘rótulos de nó’ ou ‘termos de guia’ como instâncias de `skos:Concept`, e usar relacionamentos semânticos

normais para vinculá-los a outros conceitos” (Issac; Summer, 2009, tradução nossa). Um exemplo disso pode ser visto a seguir:

```
ex:filmList rdf:type skos:Concept;
            skos:prefLabel "List's Film Genre"@en;
            skos:broader ex:film;
            skos:narrower ex:surrealist;
            skos:narrower ex:art;
            skos:narrower ex:romanticComedy .
```

As recomendações SKOS deixam em aberto a escolha entre usar uma coleção para agrupar conceitos ou usar conceitos e relacionamentos semânticos com outros conceitos para representar “rótulos de nós”, o que vai depender da aplicação. A documentação técnica afirma que o não uso de coleções pode resultar em uma perda prejudicial de precisão semântica, visto que, existem aplicações nas quais os "rótulos de nós" são entidades de natureza realmente específica e representá-los como meros conceitos, nesses casos, não é uma prática recomendada.

4.5.1 KOS LOD

O termo SOC abrange todos os tipos de esquemas para organizar informações e incluem: esquemas de classificação e categorização, cabeçalhos de assuntos e arquivos de autoridade, também incluem vocabulários altamente estruturados, como tesouros, e esquemas menos tradicionais, como redes semânticas e ontologias (Zeng; Mayr, 2018).

Os SOC na web, nos últimos anos, se desenvolveram no sentido dos dados abertos e conectados, com abordagens definidas pelo W3C, no que diz respeito à “Web Semântica” ou “dados conectados”. Nos KOS LOD (*Knowledge Organization System Linked Open Data*), termo cunhado por Zeng e Mayr (2018), os dados podem ser coletados, modelados, enriquecidos e publicados na linguagem da Web Semântica RDF, assim estruturam-se recursos para torná-los compreensíveis pelas máquinas.

No enriquecimento semântico o SOC é usado como vocabulário de valor, visto que nos modelos baseados em RDF, os triplos que estruturam a sintaxe <recurso><tipo de propriedade><valor da propriedade>, utilizam “uma lista controlada de valores permitidos para um elemento em dados estruturados” (Zeng; Mayr, 2018, p. 210, tradução nossa).

Um vocabulário de valores define recursos (como instâncias de tópicos, estilos de arte ou entidades nomeadas) que são usados como valores para elementos em registros de metadados (Zeng; Mayr, 2018, p. 210, tradução nossa).

Os exemplos incluem: tesouros, listas de códigos, listas de termos, esquemas de classificação, listas de cabeçalhos de assuntos, taxonomias, arquivos de autoridade, dicionários geográficos digitais, esquemas de conceito e outros tipos de SOC (Zeng; Mayr, 2018).

É importante lembrar, no entanto, que um vocabulário SOC é mais do que apenas a fonte de valores a serem usados em descrições de metadados: modelando as estruturas semânticas subjacentes de domínios, SOC agem como roteiros semânticos e tornam possível uma orientação comum por indexadores e futuros usuários, sejam humanos ou máquinas (Zeng; Mayr, 2018, p. 210, tradução nossa).

O KOS LOD é um termo abrangente para se referir a todos os vocabulários de valor e ontologias leves dentro da estrutura da Web Semântica (Zeng; Mayr, 2018).

O termo "ontologias leves" refere-se àquelas que usam classes e propriedades ontológicas para expressar o SOC convencional. Isso é popular entre aqueles que publicam um dicionário de sinônimos com um modelo de ontologia ao lado do SKOS. Normalmente, eles não são considerados como "ontologias de referência" que têm teorias ricas e axiomáticas com o foco em esclarecer os significados pretendidos de termos usados em domínios específicos. Nesse contexto, as ontologias leves são consideradas "ontologias de aplicação" que fornecem uma estrutura terminológica mínima para atender às necessidades de uma comunidade específica. No entanto, o termo "ontologias" tem sido aplicado a vários tipos de vocabulários, enquanto as abordagens, como ontologias superiores e ontologias híbridas, têm sido amplamente aplicadas na geração de novos SOC (Zeng; Mayr, 2018, p. 210, tradução nossa).

Segundo Zeng e Mayr (2018), os KO LOD tem como características seguir os princípios do *linked data*: ter um identificador único (URI) para representar uma entidade ou recurso para fornecer dados não ambíguos e processáveis por máquina, usar URIs HTTP para liberar conjuntos de dados como LOD, ser expressos como triplos RDF e codificados usando alguma sintaxe RDF concreta (serialização), como RDF/XML, Turtle, TriG, NQuads e JSON-LD. Permitindo a interoperabilidade dos dados, possibilitando que o SOC seja usado em aplicativos de metadados descentralizados e distribuídos. Um produto final KOS LOD pode estar disponível como um *dump* de dados RDF ou consultados em um terminal SPARQL. Também, como relacionamentos visualizados, serviços de mapeamento/correspondência instantâneos e outros métodos de entrega inovadores também podem enriquecer a presença de KOS LOD na web (Zeng; Mayr, 2018).

Os vocabulários KOS LOD são oferecidos por serviços de instituições, como bibliotecas, sendo necessário a sincronia entre os bancos de dados originais e os armazenados

em RDF, quando a fonte original sofre alterações, de modo que seja mantido a qualidade do serviço. A seguir estão representantes de provedores de serviços (PS) que desenvolveram estratégias e tecnologias para garantir a interoperabilidade, estabilidade e escalabilidade dos conteúdos e aplicativos que fornecem (Zeng; Mayr, 2018, p. 217, tradução nossa). São exemplos de sistemas KOS LOD, provedores de vocabulário individual, como o EuroVoc, além de instituições individuais fornecedoras de todos os vocabulários produzidos na instituição, como a *Library of Congress Linked Data Services* e o *Getty Vocabularies LOD*.

O *Library of Congress Linked Data Services* fornece acesso interativo e legível por máquina, como ontologias, vocabulários controlados e outras listas para descrição bibliográfica. A extensa lista de vocabulários da LC, o LCSH, por exemplo, o serviço obedece ao conjunto de práticas para publicação de dados estruturados na web, o conceito está nomeado com URI e pode ser pesquisado por meio de um URI HTTP. A figura 37 mostra a codificação legível por máquina, do conceito *Animals* modelado em SKOS serializado em RDF/XML⁶² e as ligações para vocabulários SOC externos, sendo destacado os conceitos/termos variantes, que no SKOS são representados pelo `skos:altLabel`. Ainda na figura 36, o conceito é uma instância das classes: *MADS/RDF Topic*⁶³, *MADS/RDF Authority* e *SKOS Concept*, além disso, é um conceito do esquema de conceito *Library of Congress Subject Headings*. Desse modo, o esquema formado pela classe *Concept Scheme* representa os sistemas de organização do conhecimento, ou seja, no caso desse conceito, o cabeçalho de assuntos da *Library of Congress*.

⁶² Fizemos o download do arquivo RDF/XML (MADS and SKOS) da página da LCSH. Disponível em: <https://id.loc.gov/authorities/subjects/sh85005249.html>

⁶³ MADS/RDF is a knowledge organization system (KOS) designed for use with controlled values for names (personal, corporate, geographic, etc.), thesauri, taxonomies, subject heading systems, and other controlled value lists. It is closely related to SKOS, the Simple Knowledge Organization System and a widely supported and adopted RDF vocabulary. Disponível em: <https://id.loc.gov/ontologies/madsrdf/v1.html>

Figura 37 - Conceito e rótulo lexical skos:altLabel

The screenshot shows the Library of Congress website interface. On the left, there is a sidebar with the following sections:

- Animals**
- URI(s)**
 - <http://id.loc.gov/authorities/subjects/sh850052>
 - info:lc/authorities/sh85005249
 - <http://id.loc.gov/authorities/sh85005249#concept>
- Instance Of**
 - MADS/RDF Topic
 - MADS/RDF Authority
 - SKOS Concept
- Scheme Membership(s)**
 - Library of Congress Subject Headings
- Collection Membership(s)**
 - LCSH Collection - Authorized Headings
 - LCSH Collection - General Collection
 - LCSH Collection - May Subdivide Geographically
- Variants**
- Animal kingdom
- Beasts
- Fauna
- Native animals
- Native fauna
- Wild animals
- Wildlife
- Broader Terms**

The main content area displays the RDF/XML code for the 'Animals' concept. A purple box highlights the following lines of code:

```

2023 <skos:altLabel xml:lang="en" xmlns:skos="http://www.w3.org/2004/02/skos/core#">Animal Kingdom</skos:altLabel>
2024 <skos:altLabel xml:lang="en" xmlns:skos="http://www.w3.org/2004/02/skos/core#">Beasts</skos:altLabel>
2025 <skos:altLabel xml:lang="en" xmlns:skos="http://www.w3.org/2004/02/skos/core#">Fauna</skos:altLabel>
2026 <skos:altLabel xml:lang="en" xmlns:skos="http://www.w3.org/2004/02/skos/core#">Native animals</skos:altLabel>
2027 <skos:altLabel xml:lang="en" xmlns:skos="http://www.w3.org/2004/02/skos/core#">Native fauna</skos:altLabel>
2028 <skos:altLabel xml:lang="en" xmlns:skos="http://www.w3.org/2004/02/skos/core#">Wild animals</skos:altLabel>
2029 <skos:altLabel xml:lang="en" xmlns:skos="http://www.w3.org/2004/02/skos/core#">Wildlife</skos:altLabel>
2030 </skos:changeNote>
2031 </cs:changeSet>
2032 <cs:subjectOfChange rdf:resource="http://id.loc.gov/authorities/subjects/sh85005249"/>
2033 <cs:creatorName rdf:resource="http://id.loc.gov/vocabulary/organizations/dlc"/>
2034 <cs:createdDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2003-08-21T00:00:00</cs:createdDate>
2035 <cs:changeReason rdf:datatype="http://www.w3.org/2001/XMLSchema#string">new</cs:changeReason>
2036 </cs:changeSet>
2037 </skos:changeNote>
2038 <skos:changeNote xmlns:skos="http://www.w3.org/2004/02/skos/core#">
2039 <cs:changeSet xmlns:cs="http://purl.org/vocab/changeset/schema#">
2040 <cs:subjectOfChange rdf:resource="http://id.loc.gov/authorities/subjects/sh85005249"/>
2041 <cs:creatorName rdf:resource="http://id.loc.gov/vocabulary/organizations/dlc"/>
2042 <cs:createdDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2003-10-06T12:58:04</cs:createdDate>
2043 <cs:changeReason rdf:datatype="http://www.w3.org/2001/XMLSchema#string">revised</cs:changeReason>
2044 </cs:changeSet>
2045 </skos:changeNote>
2046 </mads:rdfs:Topic>
2047 </rdf:RDF>
    
```

Fonte: Adaptado de Library of Congress (2021).

A figura 38 mostra o conceito *Animals* com as relações semânticas, hierárquicas e associativas, skos:broader, skos:narrower, que fazem parte do LCHS:

Figura 38 - Conceito e propriedades de relação semântica entre conceitos

Broader Terms	1289	<skos:broader xmlns:skos="http://www.w3.org/2004/02/skos/core#">
- Organisms	1290	<rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh2003007697">
	1291	<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
	1292	<skos:prefLabel xml:lang="en">Organisms</skos:prefLabel>
	1293	</rdf:Description>
	1294	</skos:broader>
Narrower Terms	1295	<skos:narrower xmlns:skos="http://www.w3.org/2004/02/skos/core#">
- Animals with disabilities	1296	<rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh85046565">
- Aquatic animals	1297	<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
- Arid regions animals	1298	<skos:prefLabel xml:lang="en">Extinct animals</skos:prefLabel>
- Armored animals	1299	</rdf:Description>
- Bloodsucking animals	1300	</skos:narrower>
- Burrowing animals	1301	<skos:narrower xmlns:skos="http://www.w3.org/2004/02/skos/core#">
- Canyon animals	1302	<rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh96012026">
- Captive wild animals	1303	<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
- Carnivorous animals	1304	<skos:prefLabel xml:lang="en">Tundra animals</skos:prefLabel>
- Cave animals	1305	</rdf:Description>
- Cerrado animals	1306	</skos:narrower>
- Coastal animals	1307	<skos:narrower xmlns:skos="http://www.w3.org/2004/02/skos/core#">
- Cold-blooded animals	1308	<rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh88001495">
- Compost animals	1309	<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
- Dangerous animals	1310	<skos:prefLabel xml:lang="en">Piscivores</skos:prefLabel>
- Dead animals	1311	</rdf:Description>
- Domestic animals	1312	</skos:narrower>
- Endemic animals	1313	<skos:narrower xmlns:skos="http://www.w3.org/2004/02/skos/core#">
- Euthanasia of animals	1314	<rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh2001008847">
- Exotic animals	1315	<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
- Extinct animals	1316	<skos:prefLabel xml:lang="en">Orphaned animals</skos:prefLabel>
- Famous animals	1317	</rdf:Description>
- Feral animals	1318	</skos:narrower>
- Floodplain animals	1319	<skos:narrower xmlns:skos="http://www.w3.org/2004/02/skos/core#">
- Food animals	1320	<rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh85068494">
- Forest animals	1321	<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
- Fortune-telling by animals	1322	<skos:prefLabel xml:lang="en">Island animals</skos:prefLabel>
- Fungivores	1323	</rdf:Description>
- Game and game-birds	1324	</skos:narrower>
- Garden animals	1325	<skos:narrower xmlns:skos="http://www.w3.org/2004/02/skos/core#">
- Germfree animals	1326	<rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh2007001591">
- Grassland animals	1327	<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
- Halophilic animals	1328	<skos:prefLabel xml:lang="en">Floodplain animals</skos:prefLabel>
- Heathland animals	1329	</rdf:Description>
	1330	</skos:narrower>
	1331	<skos:narrower xmlns:skos="http://www.w3.org/2004/02/skos/core#">
	1332	<rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh2008020036">
	1333	<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
	1334	<skos:prefLabel xml:lang="en">Migratory animals</skos:prefLabel>
	1335	</rdf:Description>
	1336	</skos:narrower>

Fonte: Adaptado de Library of Congress (2021).

Na figura 38, o conceito modelado em SKOS, fornece propriedades que mapeiam conceitos entre diferentes esquemas, através dos links das propriedades `skos:closeMatch` e `skos:exactMatch`, que conecta conceitos intercambiáveis. Importante ressaltar que aqui trata-se de dois recursos diferentes com nomes diferentes (dois URI), indicando serem semelhantes. Ademais, a figura 39, também mostra os conceitos relacionados (`skos:related`), usada para afirmar um link associativo entre dois conceitos SKOS.

Figura 39 - Conceito LD service Library of Congress e propriedades de mapeamento

Related Terms

- Human-animal relationships
- Zoology

Exact Matching Concepts from Other Schemes

- animals

Closely Matching Concepts from Other Schemes

- animal Label from public data source Wikidata
- Animals
- Animaux
- Animaux sauvages
- eläimet
- fauna
- Tiere
- wild animals
- wildlife

Narrower Concepts from Other Schemes

- Animals--Audio-visual aids
- Animals--Biblical teaching
- Animals--Classification
- Animals--Computer network resources
- Animals--Computer-assisted instruction
- Animals--Control
- Animals--Economic aspects
- Animals--Effect of dams on
- Animals--Effect of habitat modification on
- Animals--Effect of human beings on
- Animals--Experiments
- Animals--Fiction
- Animals--Folklore
- Animals--Growth
- Animals--Identification
- Animals--Information services
- Animals--Juvenile fiction
- Animals--Miscellanea
- Animals--Models

```

1721 <skos:related xmlns:skos="http://www.w3.org/2004/02/skos/core#"
1722 <rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh85062838">
1723 <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
1724 <skos:prefLabel xml:lang="en">Human-animal relationships</skos:prefLabel>
1725 </rdf:Description>
1726 </skos:related>
1727 <skos:related xmlns:skos="http://www.w3.org/2004/02/skos/core#"
1728 <rdf:Description rdf:about="http://id.loc.gov/authorities/subjects/sh85149983">
1729 <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
1730 <skos:prefLabel xml:lang="en">Zoology</skos:prefLabel>
1731 </rdf:Description>
1732 </skos:related>
1733 <skos:exactMatch xmlns:skos="http://www.w3.org/2004/02/skos/core#"
1734 <rdf:Description rdf:about="http://lod.nal.usda.gov/nalt/65">
1735 <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
1736 <skos:prefLabel xml:lang="en">animals</skos:prefLabel>
1737 </rdf:Description>
1738 </skos:exactMatch>
1739 <skos:closeMatch xmlns:skos="http://www.w3.org/2004/02/skos/core#"
1740 <rdf:Description rdf:about="http://data.bnf.fr/ark:/12148/cb11930908q">
1741 <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
1742 <skos:prefLabel xml:lang="fr">Animaux sauvages</skos:prefLabel>
1743 </rdf:Description>
1744 </skos:closeMatch>
1745 <skos:closeMatch xmlns:skos="http://www.w3.org/2004/02/skos/core#"
1746 <rdf:Description rdf:about="http://d-nb.info/gnd/4060007-7">
1747 <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
1748 <skos:prefLabel xml:lang="de">Tiere</skos:prefLabel>
1749 </rdf:Description>
1750 </skos:closeMatch>
1751 <skos:closeMatch xmlns:skos="http://www.w3.org/2004/02/skos/core#"
1752 <rdf:Description rdf:about="http://lod.nal.usda.gov/nalt/9588">
1753 <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
1754 <skos:prefLabel xml:lang="en">fauna</skos:prefLabel>
1755 </rdf:Description>
1756 </skos:closeMatch>
1757 <skos:closeMatch xmlns:skos="http://www.w3.org/2004/02/skos/core#"
1758 <rdf:Description rdf:about="http://id.worldcat.org/fast/809468">
1759 <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
1760 <skos:prefLabel>Animals</skos:prefLabel>
1761 </rdf:Description>
1762 </skos:closeMatch>
1763 <skos:closeMatch xmlns:skos="http://www.w3.org/2004/02/skos/core#"
1764 <rdf:Description rdf:about="http://www.wikidata.org/entity/Q729">
1765 <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
1766 <skos:prefLabel>animal</skos:prefLabel>
1767 </rdf:Description>
1768 </skos:closeMatch>

```

Fonte: Adaptado de Library of Congress (2021).

Os vocabulários KOS LOD são fontes primárias de conjuntos de dados abertos e vinculados, que em referência ao sistema de classificação e esquema de dados abertos 5 estrelas para LOD (Berners-Lee, 2006), são classificados como 4 a 5 estrelas. O sistema de classificação 5 estrelas efetua uma divisão de acordo a qualidade dos dados. Iniciado por 1 estrela onde o critério é disponibilizar informação na web em qualquer formato, mas com uma licença aberta. No nível 2 estrelas a informação precisa estar no formato de dados estruturado e legível por máquina. No nível 3 os dados são como o nível 2, mas em formatos não proprietários e o nível 4 segue todos os níveis anteriores e mais os padrões abertos do W3C (RDF e SPARQL). E o por fim o nível 5 estrelas, que comporta todos os anteriores, adicionado da vinculação dos dados (*linked data*). Desse modo, uma instituição que resolve serializar seus dados em RDF, seguindo os padrões abertos do W3C, se eleva nível 4 ou 5 estrelas LOD, o mesmo acontece quando uma instituição que publica seu vocabulário como um KOS LOD, ela irá ter uma classificação 4 ou 5 estrelas, como indicado no quadro 34.

Quadro 30 - Linked Open Data 5 star

- ★ Disponível na web (em qualquer formato), mas com uma licença aberta, para ser Open Data.
- ★★ Disponível como dados estruturados legíveis por máquina (por exemplo, Excel em vez de digitalização de imagem de uma tabela).
- ★★★ como (2) mais formato não proprietário (por exemplo, CSV em vez de Excel).
- ★★★★ Além disso, use os padrões abertos do W3C (RDF e SPARQL) para identificar as coisas, para que as pessoas possam apontar para suas coisas.
- ★★★★★ Todos os itens acima, mais: Vincule seus dados aos dados de outras pessoas para fornecer contexto.

Fonte: Berners-Lee (2006).

Nesta parte, as conclusões preliminares são apresentadas para produtores de conjuntos de dados LOD (DPs) que enfrentam diferentes níveis de situações ao produzir produtos LOD: (1) criar conjuntos de dados LOD do zero e lidar com dados que não têm valores controlados para as entidades e tópicos nomeados; (2) alcançar os conjuntos de dados que podem ter ou não usado vocabulários padrão da comunidade em seus dados estruturados; e (3) transformar os conjuntos de dados existentes que têm usado vocabulários de valores em LOD de 4 e 5 estrelas (Zeng; Mayr, 2018).

Os produtores de conjuntos de dados LOD são dedicados a explorar os dados existentes e fornecer dados estruturados no formato RDF. Os dados podem já estar estruturados, como registros bibliográficos, arquivos de documentação de museus, bancos de dados de ensaios clínicos. Entretanto, é frequente a necessidade de estruturar dados a partir de dados brutos não estruturados, como transcrições de história oral (Zeng; Mayr, 2018).

Para quebrar os silos e se conectar com as ricas informações fora de seus limites, muitos deles adotaram a abordagem Linked Data e se abriram. Os pontos de ligação são principalmente os conceitos e entidades nomeadas, ou seja, as coisas identificáveis, incluindo pessoas, organizações, lugares, eventos, objetos, conceitos e virtualmente qualquer coisa que pode ser representada em dados estruturado. Nos triplos RDF (sujeito-predicado-objeto), eles ocupam as posições de sujeito e objeto (Zeng; Mayr, 2018, p. 214, tradução nossa).

Quadro 31 - Produtores de conjuntos de dados LOD usando vocabulários SOC LOD

PRODUTOR LOD DATASET 1			
OBJETIVOS	EXEMPLOS	PROBLEMAS/NECESSIDADES	TAREFAS
Tratar os dados semiestruturados e não estruturados que não têm valores controlados para as entidades e tópicos nomeados, a fim de criar conjuntos de dados LOD do zero.	<ul style="list-style-type: none"> • Materiais digitalizados (textuais ou não textuais) hospedados em silos; • auxiliares de localização de arquivos; • Transcrições de história oral; • arquivos locais mesclados e outros. 	<ul style="list-style-type: none"> • Nomes dos lugares mudam com o tempo; • Existem nomes alternativos; • Mesmo nome está associado a vários locais; • Lugares não identificáveis; • Locais sem nome; • Localização cartográfica versus localização geográfica; • Os resultados de digitação / categorização de recursos estão incorretos ou inconsistentes. 	<ol style="list-style-type: none"> 1. Identificar as entidades; 2. Coloque as entidades em dados estruturados; 3. Limpar os dados recém-estruturados, com controle local; 4. Codifique as entidades com vocabulários SOC padronizados (como strings); 5. Obtenha URIs para entidades fornecidas pelos conjuntos de dados KOS LOD; e 6. Use URIs http para nomes de quaisquer entidades.
PRODUTOR LOD DATASET 2			
OBJETIVOS	EXEMPLOS	PROBLEMAS/NECESSIDADES	TAREFAS
Alcançar os conjuntos de dados que podem ter ou não usado vocabulários padrão da comunidade em seus dados estruturados.	The Thesaurus for French Local Archives and the Semantic Web.	<ul style="list-style-type: none"> • Representar o dicionário de sinônimos de uma forma compreensível por máquina; • Facilitar sua integração em ferramentas de recuperação; • Garantir a consistência da indexação mesmo que o dicionário de sinônimos evolua; • Expressar todos os conceitos já representados no tesauro (conceitos e termos, relações entre esses conceitos, anotações, etc.); e • Usar padrões e modelos relacionados a tesouros e vocabulários controlados para fins de interoperabilidade. 	Conversão do SOC existente em LOD antes de aplicá-los como vocabulários de valor padrão em todos os conjuntos de dados a serem integrados.
PRODUTOR LOD DATASET 3			
OBJETIVOS	EXEMPLOS	PROBLEMAS/NECESSIDADES	TAREFAS
Ter conjuntos de dados que têm usado vocabulários de valores em dados estruturados, transformando-os em LOD de 4 e 5 estrelas	Exemplos de tais dados incluem bibliografias nacionais, catálogos, portais de coleção especial, repositórios de metadados e muitos produtos de LD baseados em temas feitos em projetos.	O objetivo do produtor de conjunto de dados é transformar os conjuntos de dados existentes em LOD de 4 e 5 estrelas. Esses conjuntos de dados têm usado vocabulários de valores (nascidos com ou mapeados para) em seus dados estruturados.	<ol style="list-style-type: none"> 1. Usar protocolos padronizados para estrutura de metadados; 2. Enriquecer os metadados originais, especialm; 3. Controle os espaços de valor para todas as entidades; 4. Codifique as entidades com vocabulários SOC padronizados (como strings); 5. Use URIs para nomes de entidades; e 6. Use URIs http para nomes de quaisquer entidades

Fonte: Adaptado de Zeng e Mayr (2018).

Segundo Zeng e Mayr (2018), os processos que envolvem os produtores de vocabulário de valor no contexto LOD, são considerados diferentes dos provedores de vocabulário usuais,

como tesouros ou sistemas de classificação. Como exemplo, esses provedores apresentam objetivos como:

(1) criar novos vocabulários de valor para produtos de projetos específicos, extraindo os componentes de um vocabulário SOC abrangente; (2) criar um esquema unificado para um domínio baseado em vários vocabulários SOC; (3) criar um meta-vocabulário heterogêneo; (4) enriquecer o SOC manualmente e conectando-se com coisas reais; e (5) aumentar a consistência semântica dos dados por meio da mistura de atividades compartilhadas e não convencionais da organização do conhecimento. Apesar de os casos apresentados se assemelharem às abordagens utilizadas na comunidade SOC há muito tempo, novos métodos, funções e resultados são observados nas abordagens atuais (Zeng; Mayr, 2018, p. 217).

Quadro 32 - Produtores de vocabulários (pvs) envolvidos no enriquecimento de soc

INICIATIVAS DE VOCABULÁRIOS	ABORDAGEM DOS PRODUTORES DE VOCABULÁRIOS	MÉTODOS DOS PRODUTORES DE VOCABULÁRIOS	EXEMPLOS DE INICIATIVAS E VOCABULÁRIOS
PV 1-MICRO TESAURO	Criação de novos vocabulários de valor para produtos de projetos específicos, extraindo os componentes de um vocabulário SOC abrangente.	Enriquecimento de vocabulários a extraindo componentes de um SOC abrangente para complementar a terminologia.	O projeto “ <i>CHIN Guide to Museum Standards</i> ”, usa componentes do <i>Art and Architecture Thesaurus</i> (AAT).
PV 2-ESQUEMAS UNIFICADOS DE DOMÍNIO PV 2	Criação de um esquema unificado para um domínio com base em vários vocabulários SOC.	Desenvolvimento de novo vocabulário com base em mais de uma fonte. Harmonização e formalização de conceitos para características amplas de domínio.	- <i>Thesaurus of Plant characteristics</i> (TOP). - <i>Global Agricultural Concept Scheme</i> (GACS).
PV 3-META-VOCABULÁRIO	Criação de um metavocabulário heterogêneo	Meta-vocabulário heterogêneo para apoiar a representação de mudanças e de opiniões divergentes de conceitos de domínios específicos.	TaxMeOn
PV 4-DTESAURO SKOSIFICADO/ TESAURO CONCEITUAL	Enriquecendo o SOC disponíveis e conectando a objetos reais.	Tesouro skosificado que vincula um conceito aos URIs fornecidos por autoridades de nomes e/ou pela Wikipedia para se beneficiar totalmente do LOD e enriquecer um SOC existente e disponível.	FAST (Faceted Application of Subject Terminology),

Fonte: Baseado em Zeng e Mayr (2018).

De tal modo, é possível definir os aspectos dos KOS LOD, com a síntese abaixo (Clarke, 2018, tradução nossa):

- a) **LOD:** *Linked Open Data* - dados abertos estruturados como triplos RDF (*subject-predicate-object*) em que entidades e predicados são identificados exclusivamente por URIs HTTP.
- b) **SOC:** *Sistemas de organização do conhecimento* - terminologias conceituais e nomeadas de entidades organizadas dentro de um esquema estruturado usando hierárquico, relacionamentos associativos e de equivalência.

- c) **Estrutura semântica:** família de especificações W3C para SOC e descrição de recursos semânticos, incluindo RDF, SKOS, SKOS-XL, OWL, Linked Data Platform.
- d) **Vocabulários de valor:** terminologias descritivas de conceitos e entidades nomeadas, por exemplo léxicos, tesouros, taxonomias, esquemas de classificação e autoridades de nomes.
- e) **Vocabulários de propriedade:** conjuntos de elementos de metadados ou tipos de propriedade, incluindo as propriedades de dados e propriedades de objeto (ou seja, predicados) de uma ontologia.
- f) **Ontologias leves:** SOC que usam seleções de classes ontológicas e propriedades derivadas de um ou mais vocabulário de propriedade para atender às necessidades de uma comunidade específica (em contraste com ontologias de referência contendo regras semânticas e axiomáticas).

Os SOC acompanharam a evolução da web e estão presentes nos sistemas informatizados a bastante tempo, sendo responsáveis pelos processos gerais de classificação e categorização de documentos nesses ambientes, na estruturação taxonômica de informações digitais e nas linguagens de indexação dos sistemas de recuperação. Assim como, na organização terminológica de termos, conceitos e descritores nos cabeçalhos de assuntos, tesouros e demais vocabulários controlados nos portais das instituições.

Nesse sentido, nos últimos anos “cresceu a aplicação de dados vinculados na elaboração de SOC no ambiente web, visto que as instituições estão cada vez mais explorando os dados vinculados, o que gerou o aumento significativo do número de funcionários dedicados e envolvidos nessas atividades” (Dobreski *et al.*, 2020, p. 2, tradução nossa). Dessa forma, bibliotecas e arquivos que exploram os dados vinculados, possuem modelos de dados enriquecidos por “partes de informação” de diferentes fontes, como vocabulários e conjuntos de dados LOD. Porém, para as bibliotecas, arquivos e centros de documentação, esse processo de fragmentar a informação não é recente.

Para o mundo das bibliotecas, arquivos e centros de documentação, a ideia de atomizar a informação desta forma não é nova. Os índices temáticos vinculados aos tesouros e, principalmente, os catálogos de autoridade partem de um princípio semelhante: diferentes elementos do registro bibliográfico, que descreve o documento como um todo, vinculam-se aos registros do catálogo de autoridade. Este catálogo é onde as formas autorizadas e alternativas para representar um conceito, pessoa, lugar, instituição, etc. são registrados. Além disso, qualquer uma dessas entidades é contextualizada por meio de uma rede de referências semânticas e/ou funcionais que especificam seu significado, permitindo a exploração de documentos relacionados (Castro, 2011, tradução nossa).

Embora os profissionais de biblioteconomia e ciência da informação trabalhem com modelos de dados, ferramentas e padrões tradicionais, possuem menos familiaridade com

modelos e conceitos pertencentes a Web Semântica, como ontologias, RDF, URI, padrões, ferramentas e tecnologias utilizadas nos dados vinculados (Dobreski *et al.*, 2020).

Em 2017, a ALCTS lançou suas competências essenciais para profissionais de catalogação e metadados, que incluem dados vinculados como uma área importante do conhecimento e faz referência a ferramentas e tecnologias de dados vinculados, como RDF, Turtle e OpenRefine (Evans *et al.*, 2018). Essas competências refletem a crescente importância das atividades da Web Semântica nas instituições de informação, mas também indicam que os profissionais de metadados novos e atuais sem habilidades de dados vinculados correm o risco de ficar para trás (Dobreski *et al.*, 2020, p. 2, tradução nossa).

Dessa forma, de acordo com Tennis e Calzado-Prado (2007), as possíveis contribuições da área de Biblioteconomia e Ciência da Informação (BCI) para o desenvolvimento de vocabulários e ontologias para a Web Semântica, apresentaram alguns desafios. Como a falta de consenso terminológico entre e dentro das disciplinas (ciência da computação, ciência da informação, linguística, biblioteconomia), além da resistência dos profissionais em incorporar os avanços, as práticas e ainda a falta de compreensão conceitual com relação às formações e disciplinas acadêmicas.

Conseqüentemente, devido às mudanças nos perfis dos profissionais de Ciência da computação e Biblioteconomia/Ciência da Informação, existe uma demanda inicial pela educação mista (Tennis; Calzado-Prado, 2007). Assim, ocorreu a evolução de catalogadores para ontologistas em anúncios de emprego publicados em periódicos profissionais, existindo pontos de vistas diferentes sobre as funções da ontologia (Rupp; Burke, 2004 *apud* Tennis; Calzado-Prado, 2007).

Uma posição comum na Ciência da computação é que, ao contrário dos ambientes de informação tradicionais e limitados, onde a representação e a recuperação baseadas em vocabulários controlados eram viáveis, a natureza aberta do ambiente da web os torna inadequados. Por outro lado, os profissionais de Biblioteconomia/Ciência da Informação argumentam que vocabulários controlados são essenciais para dar qualidade à criação de ontologias e, conseqüentemente, aos serviços da Web Semântica (Tennis; Calzado-Prado, 2007, p. 309, tradução nossa).

Nos últimos anos, os cursos de graduação na Europa e nos Estados Unidos, e seus currículos se adaptaram para dar maior ênfase aos conceitos e práticas de dados conectados, embora essa seja apenas uma das muitas áreas da organização da informação (Dobreski *et al.*, 2020).

Uma revisão de 2014 dos currículos de pós-graduação em organização da informação concluiu que os futuros profissionais de metadados devem estar preparados para ver seu trabalho ocorrendo em um ambiente cada vez mais conectado, baseado em RDF e outras tecnologias da Web Semântica (Dobreski *et al.*, 2020, p. 2, tradução nossa).

A Web Semântica não é o único tema da área de BCI, mas é urgente e necessário o desenvolvimento de habilidades e projetos com as metodologias do *linked data* e ontologias, para os estudantes e profissionais que se interessarem no campo terem onde buscar formação acadêmica e atuação profissional. Em uma pesquisa com profissionais de metadados, Tosaka e Park (*apud* Dobreski *et al.*, 2020) descobriram que os principais desafios no treinamento e educação continuada de dados vinculados, inclui a dificuldade em encontrar recursos de aprendizagem apropriados e dúvidas sobre quais habilidades serão mais importantes em ambientes futuros.

Os profissionais que não estão ativamente envolvidos em um projeto em suas instituições podem não ter certeza sobre quais habilidades se concentrar e correm o risco de ficar para trás profissionalmente sem a capacidade de aprimorar essas habilidades (Dobreski *et al.*, 2020, p.2, tradução nossa).

Nesse contexto, Tennis e Calzado-Prado (2007, p. 310, tradução nossa) demonstram recomendações feitas para profissionais de BCI, como o avanço na tradução de vocabulários específicos de domínio para a sintaxe da Web Semântica, com o desenvolvimento de vocabulários e a construção de referências disponíveis e confiáveis sobre vocabulários. Ainda, melhorar o enriquecimento e expansão das relações semânticas dos tesouros. Outro ponto que os autores tratam, é a visualização de informações baseadas em ontologia, como a navegação conceitual, ademais, enfatizam a colaboração, participação e promoção de grupos de pesquisa multidisciplinares. Sobre isso, Dobreski *et al.* (2020) concluem:

Os futuros profissionais de metadados precisarão produzir e manter dados vinculados, encontrar e aproveitar fontes de dados externas e trabalhar em plataformas de produção em linha com modelos de publicação de dados baseados na Web Semântica. A preparação de profissionais para essas responsabilidades teve muito progresso, mas também existem desafios importantes (Dobreski *et al.*, 2020, p. 2, tradução nossa).

A demanda narrada por formação e projetos multidisciplinares na perspectiva da Web Semântica surge do contexto atual, pois os SOC no ambiente web estão associados a diferentes e variados conjuntos de dados. Em projetos de agregação de grande escala, como o Europeana, existem diversos tesouros com conceitos em muitos idiomas diferentes. Em tal medida, “a

publicação de tesouros e outros vocabulários em representações padrões e o mapeamento entre diferentes vocabulários está se tornando um componente-chave da interoperabilidade” (Binding; Tudhope, 2016, p. 5, tradução nossa).

4.5.2 Enriquecimento semântico de dados estruturados e humanidades digitais

De tal modo, as Humanidades Digitais estão no encontro das “disciplinas de humanidades e tecnologia da informação digital, é nesta junção que a tecnologia digital irá gerar uma mudança de paradigma em breve, permitindo que os estudiosos identifiquem os principais padrões na história, literatura e nas artes” (Zeng, 2019, p. 2, tradução nossa).

Novos campos de pesquisa estão em ascensão no domínio das Humanidades Digitais estimulados pelo surgimento do *Big Data*. Por tais razões são aplicados e desenvolvidos métodos computacionais para resolver problemas de pesquisa em humanidades e ciências sociais (Hyvönen, 2020, tradução nossa). “Neste contexto, *Big Data* significa dados que são muito grandes ou complexos para serem analisados manualmente por leitura atenta” (Hyvönen, 2020, p. 187).

A digitalização produziu diferentes objetos digitais e informações de fontes textuais, não-textuais e imagéticas, que foram escaneadas digitalmente ou nasceram nativas digitais, dessa maneira surgiu um grande volume de dados não estruturados provenientes de documentos, itens, objetos analógicos e recursos multimídias. Esse fato criou a demanda pela estruturação desses dados, visto a necessidade do processamento no ambiente de pesquisa, assim como nas bibliotecas, arquivos e museus. Nesse sentido, a etapa seguinte a digitalização é “*datafying*”, desse modo, “*datafy*” os dados não estruturados, existindo uma demanda pela “*datafication*”, que é modelar e estruturar os dados para que se tornem legíveis e processáveis por máquinas (Zeng, 2019).

Na perspectiva da pesquisa em Web Semântica, os dados do patrimônio cultural criam desafios para a pesquisa em humanidades digitais, pois a) os dados são sintaticamente heterogêneos (texto, imagens, som, vídeos e dados estruturados em diferentes formatos, como XML, JSON, CSV e RDF), b) são semanticamente ricos, cobrindo inúmeros domínios e aspectos da sociedade, c) os dados são incompletos, imprecisos, incertos devido à natureza do patrimônio histórico e cultural e d) os dados estão interligados em diferentes fontes de dados, distribuídos em diferentes países e bancos de dados (Hyvönen, 2020). O autor conclui que dessa forma a computação ajuda nos problemas metodológicos das humanidades digitais ao lidar com tais dados em problemas semanticamente complexos.

Os tipos de enriquecimento semântico, que pertencem aos diversos modelos de metadados, correspondem ao objetivo da representação de dados, por exemplo, se planeja compartilhar os metadados descritivos de um documento. Ou ainda, o objetivo pode ser a modelagem de um domínio ou o intercâmbio de entidades conceituais por esquemas de metadados.

Desse modo, Hyvönen (2020) afirma que os modelos de metadados centrados em documentos fazem parte da primeira geração conceitual de portais semânticos para pesquisa e navegação na Web Semântica. Além disso, o autor cita que inicialmente a pesquisa no domínio da Web Semântica enfocou as questões acerca da interoperabilidade sintática e semântica e agregação de dados, assim o modelo centrado em documento contempla a descrição compartilhada das propriedades e o mapeamento para promover a interoperabilidade.

Nesse cenário, os padrões de metadados e modelos de dados para harmonizar os dados, inclui os padrões W3C (RDF, OWL, SKOS, etc.), os modelos centrados em documentos, como Dublin Core e modelos centrados em eventos como CIDOC CRM para museus e suas extensões, o FRBRoo e IFLA Library Reference Model (LRM) em bibliotecas (Hyvönen, 2020).

A primeira geração conceitual de portais semânticos na Web Semântica também aborda os modelos centrados em eventos, que “são mais focados em desenvolver modelos ontológicos fundamentais do mundo real nos quais diferentes dados e metadados podem ser transformados para interoperabilidade” (Hyvönen, 2020, p. 188, tradução nossa). Assim, após serem harmonizados são publicados em um terminal SPARQL, possibilitando a criação de portais semânticos por meio de APIs (Hyvönen, 2020).

Uma análise geral e específica das abordagens centradas em documentos e centradas em eventos demonstra que ambas obtiveram sucesso, visto que “o Dublin Core e suas extensões se tornaram a norma de metadados para representar documentos na web, e muitos casos de uso e aplicativos do CIDOC CRM e outros sistemas centrados em eventos foram publicados” (Hyvönen, 2020, p. 188).

Segundo Chen (2019, p.608, tradução nossa) o enriquecimento baseado em instância, ou entidade, no “contexto do LOD é um processo para dotar os dados originais com outro conhecimento suplementar ou suportado, que está além do conteúdo original dos dados, de recursos externos”.

O enriquecimento semântico melhora a qualidade e a capacidade de descoberta dos dados das bibliotecas, arquivos e museus e se tornou uma estratégia cada vez mais usada nos últimos anos (Zeng, 2019).

Desse modo, o enriquecimento semântico estende o panorama dos dados e integra os recursos heterogêneos em um conhecimento mais completo e sofisticado, estendendo a base de conhecimento (Chen, 2019).

Chen (2019) identificou três abordagens para obter o enriquecimento da entidade, como exposto a seguir:

- 1) Enriquecer as entidades vinculando-a a diversos recursos externos de domínio cruzado para fornecer dados adicionais de múltiplas perspectivas, 2) Enriquecer as entidades com dados ausentes que são necessários para satisfazer as consultas semânticas e 3) Fornecer às entidades uma extensão para uma base de conhecimento (Chen, 2019, p. 607, tradução nossa).

De tal modo, Prasad, Giunchiglia e Devika (2017 *apud* Zeng, 2019, p. 5, tradução nossa) desenvolveram o modelo DERA (Domínio, Entidade, Relações, Atributos), onde os autores formalizaram a transição do enriquecimento semântico centrado em documento para modelagem de conhecimento centrada em entidades através da estrutura DERA, como no quadro 37:

Quadro 33 - Modelo DERA (Domínio, Entidade, Relações, Atributos)

Domínio (D)	Entidade (E)	Relação (R)	Atributo (A)
Faz referência a uma área particular do conhecimento ou campo de interesse em estudo.	consiste em um conjunto de facetas, onde cada uma representa um grupo de termos que significam as classes e as entidades do mundo real (instâncias).	consiste em um conjunto de facetas, onde cada uma representa um grupo de termos que significam as relações entre entidades e estabelece uma relação semântica entre duas entidades.	No atributo cada faceta representa um grupo de termos que denotam as propriedades qualitativas e/ou quantitativas das entidades

Fonte: Prasad, Giunchiglia e Devika (2017 *apud* Zeng, 2019, p. 5).

O enriquecimento baseado em esquema ou propriedade em LOD é um processo que comporta o significado hierárquico ou associativo em pares de propriedades, ou seja, institui uma ligação entre entidades relacionadas e também permitir uma consulta de dados em uma estrutura semântica hierárquica (Chen, 2019, p. 610, tradução nossa), ao enriquecer as propriedades definindo as relações hierárquicas e associativas entre as propriedades e especificar o domínio e intervalo das propriedades para raciocínio de dados.

4.6 ONTOLOGIA FORMAL (*HEAVYWEIGHT ONTOLOGY*)

Na filosofia o termo ontologia tem origem na filosofia metafísica, estuda e explica a natureza do ser e as coisas do mundo via conceitos. Na computação, a ontologia visa um entendimento comum, um conhecimento genérico e compartilhado de um domínio, e estabelecer uma representação formal desse conhecimento colaborativamente entre humanos e máquinas (Isotani; Bittencourt, 2015).

Na Ciência da Computação uma ontologia pode ser definida como um conjunto de conceitos fundamentais e suas relações, que capta como as pessoas entendem (ou interpretam) o domínio em questão e permite a representação de tal entendimento de maneira formal, compreensível por humanos e computadores (Mizoguch, 2004 *apud* Isotani; Bittencourt, 2015).

Além disso, uma ontologia se refere a um “artefato de engenharia, constituído por um vocabulário específico usado para descrever uma determinada realidade” (Guarino, 1998), com especificações explícitas sobre o significado das palavras do vocabulário. Ou ainda, “uma ontologia é a estrutura básica ou couraça em torno da qual uma base de conhecimento pode ser construída” (Swartout *et al.*, 1999 *apud* Isotani; Bittencourt, 2015).

A ontologia, como artefato computacional, apresenta diferenças e semelhanças com taxonomia e tesauro, termos da Ciência da informação, desse modo, a convergência ou característica comum entres os três tipos de estruturas é a capacidade de restringir a linguagem natural (Gilchrist, 2003 *apud* Almeida, 2013). As estruturas têm objetivos diferentes: “Enquanto os vocabulários controlados visam recuperar documentos, as ontologias visam modelar um domínio para produzir inferências automáticas” (Almeida, 2013, p. 1689). O tesauro é um tipo de vocabulário controlado e ambos são sistema de organização do conhecimento.

Uma ontologia pode ser modelada para permitir o compartilhamento de conhecimento e a sua reutilização em diferentes aplicações (Guarino, 1997 *apud* Isotani; Bittencourt, 2015).

No caso mais simples, uma ontologia descreve uma hierarquia de conceitos relacionados por relacionamentos de subsunção; em casos mais sofisticados, axiomas adequados são adicionados a fim de expressar outras relações entre conceitos e restringir sua interpretação pretendida (Guarino, 1998).

Desse modo, as ontologias estabelecem uma estrutura conceitual comum para o desenvolvimento de bases de conhecimento compartilháveis e reutilizáveis, ainda possibilitam a interoperabilidade e a fusão das informações de aplicações computacionais mais inteligentes (Isotani; Bittencourt, 2015).

Segundo Gruber (1993) uma ontologia é uma especificação formal explícita de uma conceituação compartilhada, na afirmação, o termo especificação formal (*formal specification*) provem da lógica matemática, com relação com a semântica formal legível por máquinas. A definição de lógica nesse contexto é o estudo de como fazer deduções corretas formais e inferências. Ademais, formal também expressa que uma ontologia é legível por máquinas e o termo “explícita” expressa que todos os conceitos devem ser definidos.

O processo que defini o domínio e suas relações por uma representação formal e explícita, chama-se conceitualização com relação com os modelos abstratos (domínio, identificação dos conceitos e relacionamentos). Assim, a conceitualização refere-se ao significado de conceitos e suas relações, dado o contexto do domínio. E “especificação” significa uma representação formal, declarativa e explícita dos mesmos conceitos e relações (Isotani; Bittencourt, 2015).

Além das duas formas de caracterizar ontologias, ontologias pesadas e as ontologias leves; existem outras formas de diferenciar ontologias. Guarino (1998) classifica as ontologias em: (i) ontologias fundacionais (ou de nível superior), que descrevem conceitos bem gerais independentemente de um domínio particular, (ii) ontologias de domínio, que descrevem a conceituação relacionada a um domínio genérico (por exemplo, medicina ou matemática), (iii) ontologias de tarefa, que descrevem a conceitualização relacionada a uma tarefa genérica (como diagnosticar ou vender) e (iv) ontologias de aplicação que descrevem conceitos dependentes em um determinado domínio e uma tarefa.

A conceituação também pode ser referida como modelagem conceitual, sendo um processo intrínseco ao desenvolvimento de ontologias e no alcance de seus objetivos, sendo uma das mais importantes no processo de desenvolvimento de ontologias (Fernandez; Gomez-Perez; Juristo, 1997 *apud* Souza; Silva, 2016) A criação de modelos conceituais está diretamente relacionada com a melhoria na representação em sistemas de informação (Mendonça; Felipe, 2020). “A modelagem conceitual descreve formalmente aspectos do mundo físico, para fins de compreensão e comunicação” (Mylopoulos, 1992 *apud* Mendonça; Felipe, 2020).

De acordo com Guizzardi, Herre e Wagner (2002) a modelagem conceitual é fruto da necessidade de formalizar modelos intermediários entre o constructo mental e o artefato computacional. A ontologia se presta a tais melhorias, por ser “[...] a conceptual specification that describe knowledge about a domain in a manner that is independent of epistemic states and state of affairs” (Guizzardi, 2005, p. 83 *apud* Mendonça; Felipe, 2020).

As ontologias, como modelos formais, existem em diferentes níveis de especificidade, desde o muito genérico até o específico. O que significa que as ontologias variam em abstração, no caso, em relação à cobertura do assunto que englobam.

A abordagem bottom-up parte da definição das classes mais específicas, para o agrupamento dessas classes em conceitos mais gerais. Ao contrário disso, a abordagem top-down, começa com a definição dos conceitos mais gerais em um domínio e continua com conceitos mais especializados.

Uma ontologia de domínio especifica os conceitos e as relações entre conceitos em uma área de assunto específica, em vez de conceitos genéricos, como ocorre em uma ontologia funcional (ou superior); de forma que, modela a informação conhecida sobre um determinado assunto (Boyce; Pahl, 2007).

Os processos que permitem essa estruturação da linguagem natural para a representação formal do conhecimento ocorrem através da extração de informações e do processamento da linguagem natural, tarefas que auxiliam na identificação dos conceitos e das relações entre conceitos. Dessa forma, todo conteúdo não estruturado na web e não legível pelas máquinas, podem ter seu conhecimento representado por triplos RDF (sujeito-predicado-objeto). Em vista disso, a representação necessita que seja nomeando e feita a descrição do recurso, no caso, recursos podem ser qualquer coisa, identificada e referenciada por um URI. Todos esses processos fazem parte dos princípios da Web Semântica.

A Web Semântica trata de uma rede de dados e de links para que pessoas e máquinas possam explorar essa estrutura conectada, o diferencial é que com um conjunto de dados vinculados é possível encontrar outros dados relacionados (Berners-Lee, 2006). Os URIs identificam qualquer tipo de objeto ou conceito:

- a) Use URIs como nomes para coisas;
- b) Use URIs HTTP para que as pessoas possam pesquisar esses nomes;
- c) Quando alguém procura um URI, forneça informações úteis, usando os padrões (RDF *, SPARQL);
- d) Inclui links para outros URIs . para que eles possam descobrir mais coisas.

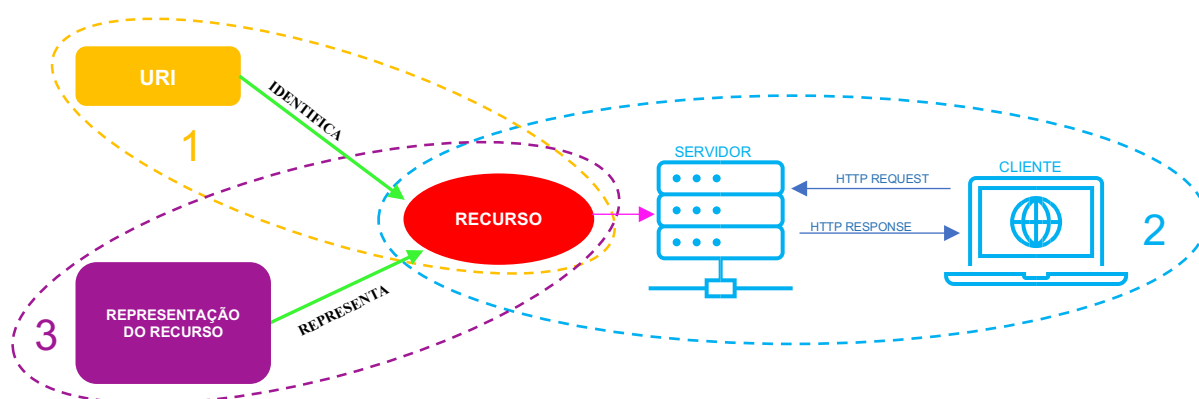
O modelo de dados RDF no contexto do *linked data* define indivíduos e as relações entre eles no nível dos dados, além disso, a estrutura de tripla rdf <Sujeito> <Predicado> <Objeto>, forma uma afirmação sobre o recurso.

O modelo RDF foi desenvolvido para descrever recursos na web e necessita de um identificador único e global (URI) para que aquele espaço de informação seja identificado como uma espécie de nome. “O RDF equivale a uma linguagem de representação de informação na

web, permitindo que recursos possam ser descritos formalmente e sejam acessíveis por máquinas” (Isotani; Bittencourt, 2015).

A arquitetura da web é composta por três bases fundamentais, como descritas a seguir (Jacobs *et al.*, 2004 *apud* Isotani; Bittencourt, 2015), conforme a figura 40, mostra o recurso único (URI), usado para identificar o recurso, seguido do processo de Interação, visto que, a web possui uma arquitetura cliente-servidor e por último, a imagem mostra os formatos: na comunicação cliente-servidor, o servidor irá retornar para o cliente (navegador) uma representação em um determinado formato. No modelo RDF existe diferentes formas de representação/serialização.

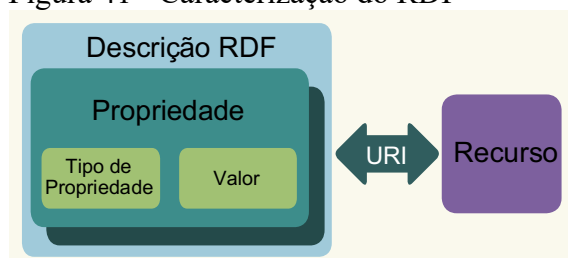
Figura 40 - As 3 bases fundamentais da arquitetura da web



Fonte: Baseado em Jacobs *et al.* (2004 *apud* Isotani; Bittencourt, 2015).

O RDF possui como características um recurso que pode ser descrito por uma coleção de propriedades denominada descrição RDF, sendo que cada propriedade na descrição RDF possui um tipo (*property type*) e um valor (*value*). Dessa forma é possível observar na Figura 41 as relações estabelecidas.

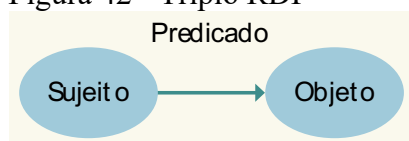
Figura 41 - Caracterização do RDF



Fonte: Isotani e Bittencourt (2015).

Ademais, o RDF auxilia na descrição dos recursos disponíveis na web como uma função sintática (descrição sintática), na construção de afirmações, igualmente, o RDF também descreve a relação e significados entre diversos recursos (descrição semântica) (Isotani; Bittencourt, 2015). Conforme a figura 42, o modelo RDF para descrever a relação entre recursos, oferece uma estrutura de triplas do tipo <sujeito> <predicado> <objeto> (Cyganiak *et al.*, 2014 *apud* Isotani; Bittencourt, 2015) chamada de Grafo RDF.

Figura 42 - Triplo RDF



Fonte: Isotani e Bittencourt (2015).

No contexto da Web Semântica, inferência sempre significa adicionar triplas, mas especificamente significa adicionar novas triplas em um grafo RDF, nas bases das triplas que já estão formadas. Uma regra de inferência pode ter vários números de premissas, mas apenas uma conclusão.

O RDFS adiciona o conceito de classe que são tipos ou conjuntos de recursos. O RDFS é um vocabulário para modelagem de dados que amplia a expressividade do RDF para prover mecanismos de descrição de taxonomias entre recursos e suas propriedades. Ou seja, permite descrever grupos de recursos (também conhecidos como classes) e suas relações utilizando o conceito de triplas (Isotani; Bittencourt, 2015). Um vocabulário predefinido permite afirmações sobre as classes. As definições de classes e propriedades são mostradas no quadro 38:

Quadro 34 - Definição de classes e propriedades RDFS

Definição de Classes	
rdfs: resource	Todas as coisas descritas por RDF são chamadas de recursos e são instâncias da classe rdfs:Resource. Esta é a classe de tudo, pois todas as outras classes são subclasses desta classe. O rdfs:Resource é uma instância de rdfs:Class.
rdfs: class	Esta é a classe de recursos que são classes RDF. O rdfs:Class é uma instância de rdfs:Class.
rdf: property	O rdf:Property é a classe de propriedades RDF. O rdf:Property é uma instância de rdfs:Class.
Definição de Propriedades	
rdf: type:	Relaciona os recursos às classes das quais eles são membros.
rdfs: domain	Domínio de uma relação
rdfs: range	Intervalo de uma relação.
rdfs: subClassOf	Inclusão de classe.

Fonte: Brickley e Guha (2014, tradução nossa).

Conforme Dilvan de Abreu Moreira⁶⁴, o RDFS suporta três tipos principais de padrão de raciocínio (Informação verbal).

I. Propagação de tipo: “O 2CV é um carro, e todos os carros são veículos motorizados, então...”

II. Herança de propriedade: “Steve dá palestras na Ifi, e qualquer pessoa que o faça é contratada pela Ifi, então...”

III. Domínio e raciocínio de intervalo: “Tudo o que alguém escreveu é um documento. Alan escreveu ‘Computing Machinery’, portanto...”

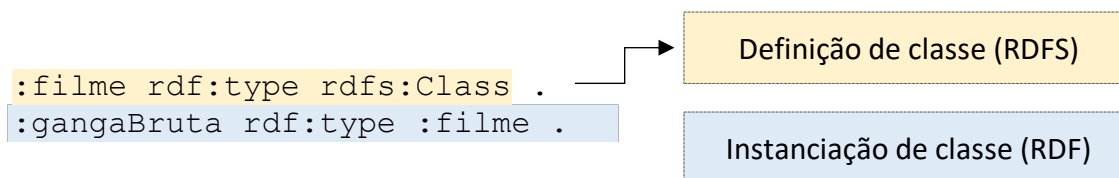
O RDFS tem uma Herança múltipla: uma filha tem mais de uma mãe, uma classe pode ser subclasse de mais de uma classe. Os problemas encontrados são:

- Não permite expressões complexas, como negação.
- Não existe fronteira rígida entre o que é ontologia e o que é dado.
- Não existe restrições no uso de construções de afirmações, é possível relacionar classes e relações.

Um modelo RDF tem apenas bordas com nomes e nós (recursos e propriedades), que são duas “coisas conectados por alguma outra coisa”, mas não apresenta uma expressividade semântica, não apresenta semântica para significar as relações entre as classes e as classes e seus indivíduos. Algumas indagações auxiliam na identificação nesse aspecto: O que é o recurso que estou fazendo declarações? Pertence a qual tipo de classe? Quais são os atributos que descrevem os indivíduos da classe? Desse modo, o RDFS permite:

- Definição de classe via `rdfs:Class`
- Instanciação de classes em RDF via `rdf:type`

Exemplo de definição e instanciação de classe em Turtle:



Definição de propriedades via `rdf:Property`

Definição de restrições de propriedade no domínio e no intervalo via `rdfs:domain` e `rdfs:range`. Exemplo em Turtle:

⁶⁴ Docente da disciplina Introdução à Web Semântica da pós-graduação em Ciências de Computação e Matemática Computacional do ICMC da Universidade de São Paulo, informação coletada na aula.

```

:artesEspetaculo rdf:type :Class .
:temContinuidade :rdf:type rdf:Property .
:temContinuidade rdfs:domain :artesEspetaculo .
:temContinuidade rdfs:range :artesEspetaculo .

```

Definição de relacionamento hierárquicos

- Subclasses e superclasse via `rdfs:subClassOf`

```
:filme rdfs:subClassOf :artesEspetaculo .
```

- Subpropriedades e superpropriedades via `rdfs:subPropertyOf`

```
:temTipo rdfs:subPropertyOf :temArte
```

Exemplo de mais algumas propriedades:

<code>rdfs:seeAlso</code>	Define uma relação de um recurso com outro, o que o explica.
<code>rdfs:isDefinedBy</code>	Define a relação de um recurso com sua definição.
<code>rdfs:comment</code>	Comentário, geralmente como texto.
<code>rdfs:label</code>	Usado para fornecer uma versão legível por humanos do nome de um recurso.
<code>rdfs:member</code>	Superpropriedade de todas as propriedades de associação do contêiner.

Definição de classe em RDFS

```

@prefix rdfs: http://www.w3.org/2000/01/rdf-schemal .
@prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns# .
@prefix : <http://example.org/filmeAnotacao> .

```

```

:categorias rdf:type rdfs:Class ;
             rdfs:subclassof :anotacaoFilme.
:filmes rdf:type rdfs:Class ;
         rdfs:subclassof :anotacaoFilme .
:digitalizados rdf:type rdfs:Class ;
               rdfs:subclassof :filmes .

```

Definição de Classe

```

:temAnotacao rdf:type rdf:Property .
:temAnotacao rdfs:domain :anotacaoFilme .
:temAnotacao rdfs:range :anotacaoFilme .

```

Definição de Propriedade

```

:campo rdt:type :categoria .
:gangabruta.35mm rdf:type :filmes ;
                 :temAnotacao :campo .
:gangabruta.mp4 rdf:type :digitalizados ;
                :temAnotacao :campo ;
                rdfs:label "Ganga Bruta mp4"@pt ;
                rdfs:comment "Anotação da versão digital"@pt .

```

Definição de Instância

A linguagem OWL⁶⁵ é projetada para uso por aplicativos que precisam processar o conteúdo da informação, em vez de apenas apresentar as informações legíveis a humanos, o que facilita que o conteúdo da web seja mais interpretado por máquina do que o XML, RDF e RDF Schema (RDF-S), fornecendo vocabulário adicional com uma semântica formal e possui três sub linguagens: OWL Lite, OWL DL e OWL Full (Mcguinness; Harmelen, 2004).

Nesse sentido, tornou-se uma recomendação em 2003, de forma que combina a expressividade da lógica descritiva com as tecnologias RDF (URI, namespaces). A lógica descritiva, que o OWL se baseia, possui dois níveis, um referente a camada dos modelos de dados (*Data level*), referente aos indivíduos e as relações entre eles e o outro com a camada de ontologias (*Ontology level*), relacionado as propriedade e classes.

A linguagem possui um vocabulário fixo para construir relações entre classes e propriedades, e seus membros. Uma interpretação dos aspectos de OWL possui as Classes, como conjuntos de indivíduos e as Propriedades, as quais são relações entre indivíduos, conjunto de pares.

As propriedades em OWL são de três tipos mutuamente disjuntas em OWL e estão explicadas no quadro 39:

Quadro 35 - Propriedades OWL

owl: DatatypeProperty	owl: ObjectProperty	owl: AnnotationProperty
Vincular indivíduos a valores de dados, por exemplo, xsd: string.	Vincular indivíduos a indivíduos. As propriedades, que são relações binárias, usadas para estabelecer relacionamentos entre indivíduos ou entre indivíduos e valores de dados.	Não tem implicação lógica, é ignorado pelos raciocinador (<i>reasoner</i>), mas aparece nas queries (busca) e qualquer coisa pode ser anotada.
Exemplos		
:hasAge :tem sobrenome.	:hasFather :driveAxle.	rdfs:label dc:creator.

Fonte: Baseado em Miles e Bechhofer (2009, tradução nossa).

Os axiomas OWL consistem nos três blocos de construção, as classes são comparáveis com as classes em RDFS. Os indivíduos são comparáveis com as instâncias de classe no RDFS. As propriedades são comparáveis com as propriedades do RDFS. Existem duas classes pré-definidas:

- OWL: *Thing* (classe que contém todos os indivíduos)
- OWL: *Nothing* (classe vazia)

⁶⁵ O OWL namespace é @prefix OWL: <http://www.w3.org/2002/07/owl#>65

A definição de uma classe OWL em Turtle serialização, pode ser realizada como, por exemplo:

```
:filme a owl:Class .
```

A definição de indivíduos OWL em Turtle pode ser realizada como, por exemplo uma definição de indivíduos através da associação em classe

```
:gangaBruta a :filme .
```

Os indivíduos também podem ser definidos sem associação de classe como indivíduos nomeados

```
:gangaBruta a owl:NamedIndividual .
```

A definição de propriedades OWL pode ser realizada como, por exemplo, existem duas variantes de propriedade:

- Propriedades do objeto
- Propriedades do tipo *datatype*

As propriedades dos objetos têm classes como domínio (*rdfs:domain*) e intervalo (*rdfs:range*).

```
:diretor a owl:ObjectProperty.
```

O Domínio e o intervalo de propriedades do objeto (*owl:ObjectProperty*) em Turtle tem exemplo a seguir:

```
:autor a owl:ObjectProperty;
  rdfs:domain :filme ;
  rdfs:range :diretor .
```

As propriedades do tipo OWL Datatype têm os tipos de dados como intervalo:

```
:cenaID a owl:DatatypeProperty.
```

O Domínio e alcance de propriedades tipo *datatype* em Turtle tem exemplo a seguir:

```
:cenaID a owl:DatatypeProperty;
  rdfs:domain :media ;
  rdfs:range xsd:integer.
```

As propriedades OWL e indivíduos, na serialização Turtle, segue o exemplo:

```
:filme a owl:Class.
:criador a owl:Class.
```

Classes

```
:diretor a owl:ObjectProperty;
  rdfs:domain :filmes ;
  rdfs:range :criador .
```

Propriedades de objeto

```
:anoLançamento a owl:DatatypeProperty;
rdfs:domain :filme ;
rdfs:range xsd:integer.
```

Propriedades de datatype

```
:humbertoMauto a criador .
:gangaBruta a :filme ;
:diretor :humbertoMauro ;
:anoLaçamento 1933 .
```

Instanciação de classe

Hierarquias de Classe OWL

Não é preciso definir uma nova propriedade subclassof para OWL, pois é reutilizada a propriedade `rdfs:subclassof`.

```
:Longa a owl:Class;
rdfs:subClassOf :Filme.
```

```
:Filme a owl:Class ;
rdfs:subClassOf :Obra.
```

```
:Obra a owl:Class.
```

De modo que, via inferência pode ser implicado que `:Longa` é também uma subclasse de `:Obra`.

Classes OWL Disjunct

Tudo pode ser potencialmente idêntico, se não for declarado explicitamente a diferença.

```
:filme a owl:Class.
```

```
:criador a owl:Class.
```

```
:Longa a owl:Class;
rdfs:subClassOf :filme .
```

```
:cineasta a owl:Class; rdfs:subClassOf :criador.
:filme owl:disjointwith :criador .
```

De maneira que, via inferência pode ser implicado que :Longa e :realizador também são *disjoint* (disjuntas).

OWL Class Equivalence

```
:Criador a owl:Class .
:Realizador a owl:Class .
:Cineasta a owl:Class ;
  rdfs:subClassOf :Criador .
:Criador owl:equivalentClass :Realizador .
```

Assim, via inferência, pode ser implicado que :Cineasta também é um :Realizador .

OWL Individuals - Disjunction and Equivalence

```
:GangaBruta a :Longa;
  :diretor :GeorgeOrwell;
  :anoLancamento 1933 ;
  owl:sameAs :LMB012345 .
:Longa a owl:Class;
  rdfs:subClassOf :Filme.
:Filme a owl:Class.
```

Portanto, por inferência pode-se concluir que :LMB012345 é um :Filme .

Para declarar a diferença de indivíduos via owl:differentFrom.

```
:FCN221385 a :Longa;
owl:differentFrom :FCN221189 .
```

Para indivíduos idênticos: owl:sameAs e para classes idênticas: owl: equivalentClass.

OWL Classes complexas

Nominais

Também chamadas de classes fechadas, são classes onde são definidos explicitamente quais são os indivíduos que eles podem hospedar ou classificar, não sendo possível adicionar novos indivíduos após a definição.

```
:Longa a owl:Class.
```

```

:BrasaDormida a :Longa.
:GangaBruta a :Novel.

:filmesColeção a owl:Class ;
    owl:oneOf
    (:BrasaDormida
     :GangaBruta ).

```

Logo, ficou definido que há apenas dois filmes pertencentes a coleção de filmes.

OWL Construtores de Classe Lógicas

logical AND (conjunction):	owl:intersectionof	\sqcap
logical OR (disjunction):	owl:unionOf	\sqcup
logical negation:	owl:complementof	\neg

Os construtores lógicos são aplicados para criar descrições de classes complexas a partir de classes atômicas.

OWL Construtores de Classe Lógica - INTERSECÇÃO

```

:Filme a owl:Class.
:Coleção a owl:Class.
:FilmesColeção a owl:Class;
    owl:intersectionof (:Filme :Coleção).

```

A classe :FilmesColeção resulta da intersecção de todos os indivíduos das classes :Filme e :Coleção.

OWL Construtores de Classe Lógica – UNION

```

:Filme a owl:Class;
    owl:equivalentClass [
        owl:unionof (:Longa
                     :Curta
                     :Media )
    ] .

```

Neste caso acima, :Longa, :Curta e :Media também são filmes.

OWL Construtores de Classe Lógica – NEGAÇÃO

```

:Filme a owl:Class;
    rdfs:subclassof

```

```

    [ owl:complementof :Criador
  ] .

```

A afirmação é equivalente semanticamente a declaração:

```

:Filmes a owl:Class;
  owl:disjointwith :criador .

```

Restrições de propriedade OWL

As restrições de propriedade OWL são usadas para descrever classes complexas através de propriedades:

Restrições de valores

- owl:hasValue
- owl:allValuesFrom
- owl:someValuesFrom

Restrições de cardinalidade

- owl:cardinality
- owl:minCardinality
- owl:maxCardinality

OWL Restrições de propriedade com CONSTANTS

```

:HumbertoMauroFilmes a owl:Class;
  rdfs:subclassof
    [a owl:Restriction;
      owl:onProperty :diretor;
      owl:hasValue :HumbertoMauro
    ] .

```

Assim, a classe :HumbertoMauroFilmes é descrita através da atribuição de valor fixo (=constante) do indivíduo :HumbertoMauro a propriedade :Diretor.

OWL Restrições de propriedade com STRICT BINDING

```

:Curta a owl:Class;
  rdfs:subclassof
    [a owl:Restriction;
      owl:onProperty :diretor ;
      owl:allValuesFrom :Criador
    ] .

```

Neste caso, o `owl:allValuesFrom` fixa todas as instâncias de uma classe específica como faixa permitida para uma propriedade, no caso apenas instâncias da classe `:Criador`.

OWL Restrições de propriedade com LOOSE BINDING

```
:cineasta a owl:Class;
  rdfs:subclassof

    [ a owl:Restriction;
      owl:onProperty :director ;
      owl:someValues From :Filme
    ] .
```

OWL Restrições de propriedade com CARDINALIDADES

```
:trilogia a owl:Class ;
  rdfs:subclassof

    [ a owl:Restriction;
      owl:onProperty :temVolumes ;
      owl:cardinality 3
    ] .
```

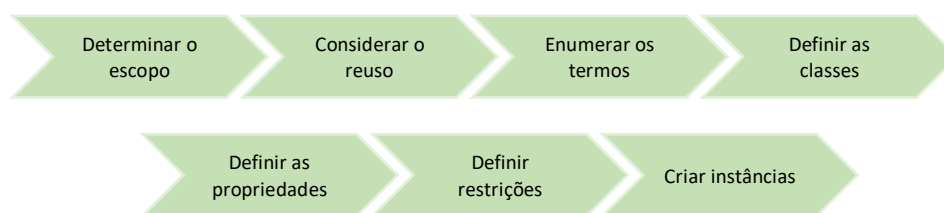
O artigo “*Ontology Development 101: A Guide to Creating Your First Ontology*” aborda as questões de como e quando construir uma ontologia e apresenta uma metodologia para criar ontologias baseadas em sistemas de representação de conhecimento, tratando o desenvolvimento e manutenção de ontologias em vários ambientes de ontologia, incluindo o Protégé. O artigo utiliza um exemplo de base de conhecimento de vinhos para apresentar a metodologia.

A engenharia de ontologia é uma disciplina da Ciência da computação que estuda o desenvolvimento de ontologias em geral, ou seja, não apenas no contexto da web e da linguagem OWL, existindo outras linguagens para construção de ontologias. As ontologias são SOC, portanto atuam na organização do conhecimento. Além disso, tem conexão com a gestão do conhecimento, com a transformação de dados em conhecimento e com a aplicação coletiva de conhecimento em um contexto. Dessa forma, as ontologias podem atuar na organização de informação audiovisual na Web Semântica, na representação e recuperação de conceitos em domínios compartilhados da imagem em movimento, em projetos de indexação semântica, anotação semântica, interoperabilidade de dados, no reuso de ontologias e no compartilhamento de vocabulários.

De acordo a metodologia 101, os principais passos para construir uma ontologia, são definir os termos no domínio do conhecimento, quais as “coisas” trabalhadas no domínio e as relações entre elas. Ao mesmo tempo, definir os conceitos do domínio, as classes do domínio, as hierarquias, atributos das classes e suas restrições de valores, o que pode ou não pertencer a uma classe (Noy; McGuinness, 2001).

Em relação aos motivos para desenvolver ontologias, entende-se que o principal motivo, é para compartilhar um entendimento comum de uma estrutura da informação entre pessoas e agentes de software, chamada de conceituação. O reuso de conhecimento de domínio e introdução de padrões para interoperabilidade são aplicados para fazer proposições explícitas sobre um domínio. As fases para construção de uma ontologia de acordo a metodologia 101, incluem as seguintes etapas (Noy; McGuinness, 2001):

Figura 43 - Fase de construção da metodologia 101



Fonte: Noy e McGuinness, (2001).

a) Determinar o escopo

A Ontologia deve refletir uma estrutura do mundo e a estrutura de conceitos e não é necessária a representação física do dado durante a modelagem, ou seja, o tipo de dado, se é um tipo integral, *float*, *string*, etc.

Ademais, Noy e McGuinness (2001), afirmam serem perguntas essenciais para o desenvolvimento da metodologia 101: qual domínio sua ontologia irá cobrir, se é amplo ou estreito, caso seja estreito, se relaciona com a modelagem de um domínio específico. Por outro lado, um exemplo de uso amplo: schema.org, por permitir cobrir os variados escopos da web social. Um exemplo de motivo para usar a ontologia: um modelo de dados de um sistema de informação.

Além do que, os autores colocam que também são questionamentos pertinentes: que tipos de questões a informação presente na ontologia responde e qual a competência da ontologia. Essas questões e respostas podem mudar ao longo do ciclo de vida da ontologia.

Um fato, muito importante, é considerar o reuso, para obter economia de esforço, diminuir o custo, interagir com ferramentas que usam outras ontologias, além de, validação de

conceitos que já foram usados em outras aplicações e interoperabilidade. Dessa forma, é possível reusar: a) Bibliotecas de ontologia, b) ontologias gerais (conceitos básicos de ontologias, ex: worldnet), c) ontologias de domínios específicos (CIDOC-CRM-Cultural heritage) (Noy; Mcguinness, 2001).

b) Enumerar os termos

Para a enumeração dos termos é necessário, determinar quais são os termos importantes de usar na ontologia e os termos necessários de pontuar, ainda, quais são as propriedades que esses termos devem ter, igualmente é imprescindível o acompanhamento de um profissional ou pesquisador “*domain expert*” para a definição das informações. Por fim, é importante, definir o que se planeja afirmar sobre os termos (Noy; Mcguinness, 2001).

Nesse estágio ainda não é necessário definir se é uma classe, propriedade ou indivíduo (Noy; Mcguinness, 2001).

c) Definir as classes

Uma classe é um conceito num domínio (Noy; Mcguinness, 2001), desse modo, uma classe define um conjunto de recursos e instâncias. Como, por exemplo, uma classe de vinho, uma classe de vinícolas e uma classe de vinho tinto. Uma classe é uma coleção de instâncias com a mesma propriedade, já uma instância é um exemplo da coleção. Por exemplo, São Paulo é uma instância da classe Cidade, ou da classe “lugar habitado”, ou também da classe “cidade do Estado de São Paulo”.

Uma base de conhecimento de domínio, para representar o conhecimento, é dividida em duas partes, o conhecimento intensional (*Intensional Knowledge*), sendo o conhecimento geral sobre o domínio do problema, sobre os conjuntos de indivíduos que apresentam as mesmas propriedades, por outro lado, o conhecimento Extensional (*Extensional Knowledge*), especifica um problema particular e representa o conhecimento sobre os indivíduos do conjunto (Vieira *et al.*, 2005). Em lógica descritiva esses conhecimentos são chamados de TBox e ABox respectivamente, e usadas para descrever dois tipos diferentes de afirmação em ontologias.

A Tbox (*terminological component*) é um conjunto de classes e propriedades para conceitos. É a ontologia propriamente dita, contém o conhecimento intencional na forma de terminologia, representa as propriedades gerais das classes (conceitos) e só é permitido uma definição para cada conceito (Vieira *et al.*, 2005). A seguir estão dois exemplos:

```
class Man and Woman are disjointed
Student and Lecture are two types of Person
```


A ABox (*assertion component*) especifica os indivíduos ou instâncias que pertencem aos conceitos da TBox e contém o conhecimento extensional. As afirmações são associadas às instâncias da classe, como nos exemplos a seguir:

```
Bob is a student
a is a instance of class A
```

Desse modo, a Tbox define a ontologia e a terminologia, por outro lado, a Abox define as instâncias e os dados.

d) Definindo propriedades de classes

As propriedades de classe são os atributos que definem a classe. Como, por exemplo, a classe “mamífero” o atributo que define é mamar. Além disso, as propriedades descrevem relações com as outras instâncias. As propriedades podem ser dos tipos: Extrínsecas e intrínsecas. Ou ainda, serem simples ou complexas (Noy; McGuinness, 2001):

- Propriedades simples (atributos): valores primitivos (*strings, numbers*).
- Propriedades complexas: apontam para outros objetos e não dados.

Na Herança de propriedade e classe, uma subclasse herda todos os atributos da super-classe, se uma classe tem múltiplas super-classes ele irá herdar todas elas.

e) Definir as restrições (*constraints*)

O quadro 40 descreve as descrições de limites para as propriedades a partir de valores.

Quadro 40 - Descrições de limites para as propriedades a partir de valores

Limites para as propriedades a partir de valores	O nome do vinho só pode ser uma <i>string</i> .
	Um produtor x será sempre uma instância da classe vinícola.
	A vinícola tem exatamente um local como endereço.
Restrições comuns	
Cardinalidade	Os números de valores que uma propriedade pode ter.
Tipo de valor	Dados tipados como, número inteiro, <i>string</i> , etc.
Mínimo e máximo	O range de valores para propriedade numérica.
Domain e Range	
Domínio	O domínio de propriedade limita os indivíduos que a propriedade pode ser aplicada.
Range (alcance)	O alcance de uma propriedade limita os indivíduos que a propriedade pode ter como valor.

Fonte: Elaborado pela autora.

4.6.1 SPARQL 1.1

A SPARQL (*Protocol and RDF Query Language*), trata-se de uma linguagem padronizada para a consulta de grafos RDF, padrão pelo *RDF Data Access Working Group*

(DAWG) do *World Wide Web Consortium* (W3C). É uma tecnologia básica no desenvolvimento da Web Semântica que se constituiu como recomendação oficial do W3C em janeiro de 2008, sendo atualizada à versão 1.1 em 2013 (Wikipédia, 2023).

O SPARQL contém recursos para consultar padrões de grafo obrigatórios e opcionais, com conjunções e disjunções. Os resultados das consultas SPARQL podem ser conjuntos de resultados ou grafos RDF. A maioria das formas de consulta SPARQL contém um conjunto de padrões triplos chamado *padrão gráfico básico*.

O exemplo abaixo mostra uma consulta SPARQL para localizar o título de um filme no grafo de dados fornecido. O padrão gráfico básico neste exemplo consiste em um único padrão triplo com uma única variável (?title) na posição do objeto. O dado completo a ser consultado é o seguinte:

```
<http://example.com/CRIM/recurso/filme#filme1> rdfs:label "Ganga Bruta"@pt .
```

A consulta (*Query*) no dado:

```
SELECT ?label
WHERE
{
<http://example.com/CRIM/recurso/filme#filme1> rdfs:label ?label .
}
```

Esta consulta, nos dados acima, tem uma solução, como resultado da consulta:

Label
"Ganga Bruta"

Além de, também é possível realizar consultar a partir de múltiplas correspondências.

Os dados a serem consultados:

```
<http://example.com/CRIM/recurso/filme#filme1> a tfo:Work ;
  crim:mediaURI      mcrim:1231 ;
  rdfs:label         "Ganga Bruta"@pt ;
  dbo:filmVersion   "Digitalizado"@pt ;
  dbo:releaseDate   "1933-00-00"@pt ;
  dbo:year           "1933"^^xsd:integer .
```

A consulta (*Query*) nos dados:

```
@prefix mcrim: <http://example.com/CRIM/recurso/media#> .
@prefix fcrim: <http://example.com/CRIM/recurso/filme#> .
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dbo:   <http://dbpedia.org/ontology/> .
```

```
SELECT ?filmVersion ?year
WHERE
{ crim:filme1 rdfs:label ?label .
```

```
crim:filme1 dbo:year ?year }
```

O resultado da consulta SPARQL nos dados:

Versão do filme (?filmVersion)	Ano (?year)
Digitalizado	1933

Outra forma de consulta, é através da correspondência de literais com tags de idioma. As tags de idioma no SPARQL são expressas usando @ e o idioma da tag. Como o exemplo a seguir:

```
SELECT ?uri WHERE { ?uri ?p "Ganga Bruta"@pt }
```

URI
<http://example.com/CRIM/recurso/filme#filme1>

Ou ainda, é realizar consultas usando a correspondência de literais com tipos numéricos, como a seguir:

```
SELECT ?uri WHERE { ?uri ?p 1933 }
```

O resultado da consulta nos dados:

URI
<http://example.com/CRIM/recurso/filme#filme1>

Com as Restrições de Termo RDF, o uso de filtros (SPARQL FILTERS) restringem os resultados das consultas para os quais a expressão do filtro é avaliada como TRUE. A função SPARQL `regex` corresponde apenas a literais simples sem marca de idioma. Para os seguintes dados consultados, tidos como exemplo:

```
fcrim:filme1 rdfs:label "Ganga Bruta" .
fcrim:filme1 :mediaNumero 1231 .
fcrim:filme2 rdfs:label "Brasa dormida" .
fcrim:filme2 :mediaNumero 1232 .
```

A consulta (*Query*) nos dados:

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix fcrim: <http://example.com/ CRIM /recurso/filme#> .
@prefix : <http://example.com/ CRIM /recurso/mediaNumero#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

SELECT ?label
WHERE { ?x rdfs:label ?label
        FILTER regex(?label, "^Ganga")
```

}

O resultado da consulta nos dados:

Label
"Ganga Bruta"

Além de que, SPARQL FILTERS podem restringir expressões aritméticas:

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix fcrim: <http://example.com/ CRIM /recurso/filme#> .
@prefix : <http://example.com/ CRIM /recurso/mediaNumero#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

SELECT ?label ?mediaNumero
WHERE { ?x :mediaNumero ?mediaNumero .
        FILTRO (?mediaNumero = 1232)
        ?x rdfs:label ?label . }
```

O resultado da consulta:

Label	N° da mídia
Brasa dormida	1232

4.7 SKOS e OWL

Formalmente o modelo de dados SKOS é definido como uma ontologia OWL. Os elementos do modelo de dados SKOS são classes e propriedades OWL, e um URI é fornecido para cada uma dessas classes e propriedades para poderem ser usadas sem ambiguidade na web. Portanto, este conjunto de URIs é o vocabulário SKOS. As características lógicas e de interdependências entre as classes e propriedades do modelo de dados SKOS definem a estrutura e a integridade do modelo.

Este é talvez um dos aspectos mais poderosos e potencialmente confusos do SKOS, porque o SKOS pode, em aplicações mais avançadas, também ser usado lado a lado com o OWL para expressar e trocar conhecimento sobre um domínio. No entanto, SKOS não é uma linguagem formal de representação de conhecimento (Miles; Bechhofer, 2009, tradução nossa).

A documentação sobre a relação entre SKOS, RDF e OWL, aponta que existe uma distinção entre ontologias formais e esquemas de classificação convencionais. O conhecimento em ontologias é expresso como conjuntos de axiomas e fatos, enquanto esquemas de classificação não afirmam nenhum axioma ou fato. De tal forma, esquemas de classificação e tesouros identificam e descrevem unidades de conhecimento, os "conceitos", usando linguagem natural e relacionamentos hierarquias genéricos/partitivos e relacionamentos associativos. No

entanto, essas estruturas não possuem semântica formal e não podem ser interpretadas como axiomas ou fatos sobre o mundo, pois nunca foram destinadas a ter essa função.

O SKOS permite modelar apenas indivíduos e descrições informais sobre os relacionamentos entre conceitos, mas não como axiomas de classe e propriedade. Isso significa que a rotulagem de um conceito e seu mapeamento em um tesauro não são fatos explícitos sobre a conceitualização de domínio específico, como ocorre em uma ontologia formal.

De forma que, existe o problema específico de articular conceitos SKOS com classes definidas pela linguagem de ontologia OWL, em algum contexto que exige que os recursos do SKOS sejam usados em coordenação com outras abordagens de modelagem.

A referência SKOS define `skos:Concept` como uma classe OWL:

```
skos:Concept rdf:type owl:Class .
```

Assim, instâncias de `skos:Concept` (por exemplo, `ex:Painting` em um vocabulário de arte) são, em termos OWL, indivíduos.

```
ex:Painting rdf:type skos:Concept .
```

Isso levanta a questão de saber se uma instância de conceito SKOS, como `ex:Painting` pode ser tratada como uma classe por si só. Por exemplo, os usuários podem definir propriedades `ex:painting` como `ex:support`:

```
ex:support rdf:type owl:DatatypeProperty.
ex:support rdfs:domain ex:Painting.
```

A necessidade de tratar um conceito SKOS conceitualmente uma classe se dá, pois `skos:Concept` pode ser interpretada como uma metaclasses e suas instâncias são os conceitos que ocorrem em um vocabulário. Portanto, é aceitável que os usuários do SKOS desejem especificar características de nível de classe dos conceitos do SKOS, por exemplo, que os estados tenham cidades ou que a cerveja tenha tipos específicos.

A única afirmação sobre o relacionamento formal entre a classe de conceitos SKOS e a classe de classes OWL é de que `skos:Concept` é uma instância de `owl:Class`. O que permite explorar diferentes padrões de design para trabalhar com SKOS em combinação com OWL. No grafo de exemplo abaixo, `<MyConcept>` é uma instância de `skos:Concept` e uma instância de `owl:Class` e é consistente com o modelo de dados SKOS.

```
<MyConcept> rdf:type skos:Concept , owl:Class .
```

Há quatro situações e possíveis padrões para a implementação de um modelo que combine SKOS e OWL.

- a) Migrar da estrutura de um tesouro semiformal para uma ontologia formal, possibilitando novos tipos de inferência.
- b) Gerar uma estrutura semiformal de um tesouro a partir de uma ontologia formal, como, por exemplo, para usufruir de interface simples de navegação.
- c) Uma conceituação "híbrida" que seja parcialmente formal, parcialmente semiformal. Esta é uma situação muito comum, de um sistema de "navegação facetada".

Por exemplo, o Diretório Ambiental da Web Semântica (SWED) tem um modelo formal descrevendo organizações, projetos e suas respectivas propriedades, como seu tópico de interesse, sua cobertura geográfica etc., e vocabulários semiformais descrevendo, por exemplo, tópicos reais de interesse, como bem-estar animais, conservação etc (Miles; Bechhofer, 2008, tradução nossa).

- d) Adicionar mais informações legíveis por humanos (rótulos e documentação) a uma ontologia formal existente.

A abordagem que vai do menos para mais formal (SKOS para OWL), tem origem numa conceituação semiformal, expressa em SKOS, para uma conceituação formal, expressa usando OWL. Partir de uma conceituação menos formal para mais formal, apresenta dois padrões básicos para expressar o resultado desse processo. Um exemplo, pode ser expresso a partir da seguinte conceituação semiformal em SKOS:

```
ex:biomas rdf:type skos:Concept ;
  skos:prefLabel "Biomas"@pt .

ex:amazonia a skos:Concept ;
  skos:prefLabel "Amazônia"@pt ;
  skos:broader ex:biomas .

ex:FlorestaAmazonica a skos:Concept ;
  skos:prefLabel "Floresta Amazônica"@pt ;
  skos:broader ex:amazonia .
```

Para sobrepor SKOS com OWL, o OWL sobrepõe semânticas adicionais no mesmo vocabulário, por exemplo, adicionando as seguintes triplas:

```
ex:biomas rdf:type owl:Class .

ex:amazonia rdf:type owl:Class ;
  rdfs:subClassOf ex:biomas.

ex:florestaAmazonica rdf:type ex:amazonia.
```

Essa mesclagem de dois conjuntos de triplos leva a uma representação OWL devido à possibilidade de uma instância de `skos:Concept` também ser uma instância de `owl:Class`. Permitir esse padrão de design afeta a semântica de `skos:Concept`, pois ele não pode ser separado de `owl:Class` ou `owl:ObjectProperty`.

Para transformar SKOS em OWL, o processo é uma "transformação" e exige um novo conjunto de URIs para denotar as classes, propriedades e indivíduos no resultado (saída) da transformação.

Ou seja, a antiga URI `@prefix ex: <https://:exemplo.com/ex#>`, foi substituída pela nova URI `@prefix ex2: <https://:exemplo.com/ex2#>`.

Por exemplo, o resultado da transformação pode ser:

```
ex2:ClasseBiomias rdf:type owl:class .
ex2:ClasseAmazonia rdf:type owl:class ;
  rdfs:subClassOf ex2:ClasseBioma.
ex2:IndividuoFlorestaAmazonica rdf:type ex2:ClasseAmazonia .
```

Esse padrão considera SKOS e OWL como universos separados com vocabulários distintos. Permitir esse padrão não afeta a escolha de semântica para `skos:Concept`.

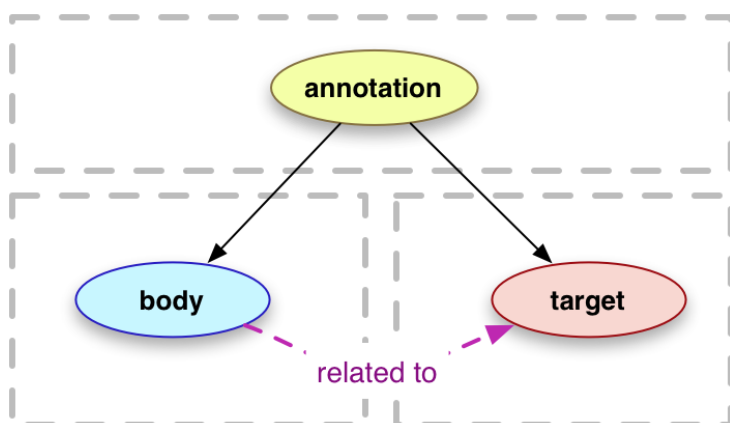
4.8 WEB ANNOTATION DATA MODEL

As anotações são usadas para transmitir informações sobre um recurso ou associações entre recursos, como por exemplos, incluem um comentário ou *tag* numa única página da web ou imagem, ou uma postagem de blog sobre um artigo de notícias (Sanderson; Ciccarese; Young, 2015).

A especificação do “Modelo de Dados de Anotação da Web” descreve um modelo e formato estruturados para permitir que as anotações sejam compartilhadas e reutilizadas em diferentes plataformas de hardware e software (Sanderson; Ciccarese; Young, 2015).

Uma anotação é considerada um conjunto de recursos conectados, geralmente incluindo um *body* (corpo) e um *target* (alvo), onde o corpo está relacionado ao alvo e sempre é "sobre" o alvo. O que forma um modelo básico com três partes, como a seguir na figura 44:

Figura 44 - Web Annotation Data Mode



Fonte: Sanderson, Ciccarese e Young (2015).

O modelo de dados de anotação da web é definido usando os seguintes princípios básicos (Sanderson; Ciccarese; Young, 2015): uma anotação é um gráfico direcionado enraizado que representa um relacionamento entre recursos. Ademais, existem dois tipos principais de recursos que participam desse relacionamento, Corpo (*body*) e Alvo (*target*). As anotações têm 0 ou mais corpos, bem como, as anotações têm 1 ou mais alvos. O conteúdo dos recursos do Corpo está relacionado a, e normalmente "sobre", o conteúdo dos recursos, o Alvo. Outro ponto, as anotações, corpos e alvos podem ter suas próprias propriedades e relacionamentos, geralmente incluindo a criação e informações descritivas. A intenção por trás da criação de uma Anotação ou inclusão de um determinado Corpo ou Alvo é uma propriedade importante e representada por um recurso de Motivação.

Muitas anotações referem-se a parte/região de um recurso como *target*, essa parte chama-se Segmento (de interesse). "Um Seletor é usado para descrever como determinar o segmento de dentro do recurso de origem" (Sanderson; Ciccarese; Young, 2015).

A web é distribuída, com diferentes sistemas trabalhando juntos para fornecer acesso ao conteúdo. As anotações podem ser usadas para vincular esses recursos, sendo referenciadas como Corpo e Destino. O recurso de destino é sempre um recurso da web externo, mas o corpo também pode ser incorporado à anotação. Recursos da web externos podem ser desreferenciados separadamente para recuperar uma representação de seu estado, enquanto o corpo incorporado não precisa ser desreferenciado, pois a representação está incluída na representação da anotação (Sanderson; Ciccarese; Young, 2015).

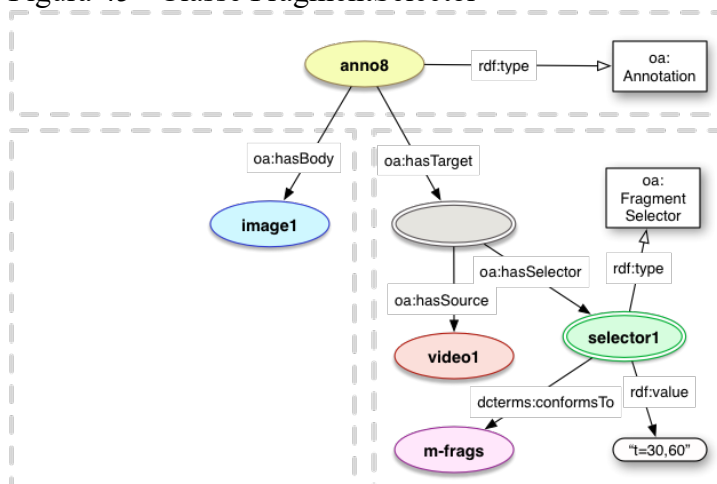
Classe FragmentSelector (Classe do Web Annotation Data Model)

A classe FragmentSelector (Fig.45) é usada para registrar o segmento de uma representação usando a especificação de fragmento IRI definida pelo tipo de mídia da representação.

IRI: <http://www.w3.org/ns/oa#FragmentSelector>

Prefix	Namespace
oa	http://www.w3.org/ns/oa#

Figura 45 - Classe FragmentSelector



Fonte: Sanderson, Ciccarese e Young (2017).

Abaixo um exemplo de `oa:FragmentSelector`:

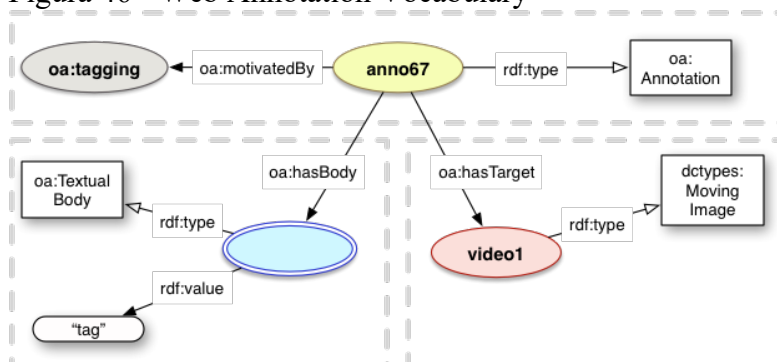
```
<http://example.org/anno7> a oa:Annotation ;
  oa:hasBody <http://example.org/image1> ;
  oa:hasTarget [
    oa:hasSource <http://example.org/video1> ;
    oa:hasSelector [
      a oa:FragmentSelector ;
      dcterms:conformsTo <http://www.w3.org/TR/media-frags/> ;
      rdf:value "t=30,60" ] ] .
```

Classe dctypes:MovingImage (Classe de ontologia recomendada)

IRI: <http://purl.org/dc/dcmitype/MovingImage>

Prefix	Namespace
Dctypes	http://purl.org/dc/dcmitype/

Figura 46 - Web Annotation Vocabulary



Fonte: Sanderson, Ciccarese e Young (2017).

Exemplo de `dctypes:MovingImage`:

```
<http://example.org/anno64> a oa:Annotation ;
  oa:hasBody [
    a oa:TextualBody ;
    rdf:value "tag" ] ;
  oa:hasTarget <http://example.com/video1> ;
  oa:motivatedBy oa:tagging .
```

```
<http://example.org/video1> a dctypes:MovingImage .
```

A classe `dctypes:MovingImage` é definida como um tipo específico de recurso, indicando que o recurso anotado é um conteúdo de imagem em movimento, como um vídeo. A notação `dctypes:MovingImage` é baseada no modelo de metadados Dublin Core (DC). O DCMI Type Vocabulary, uma extensão do Dublin Core, inclui a categoria `MovingImage` para especificar recursos que são imagens em movimento, como vídeos ou animações. A anotação pode conter informações adicionais, como comentários, tags ou descrições.

5 ABORDAGEM METODOLÓGICA: MODELO SEMÂNTICO

Essa pesquisa faz uma abordagem qualitativa, de natureza teórica e aplicada, de caráter exploratório e descritivo, assim os procedimentos adotados são a pesquisa bibliográfica e a modelagem conceitual. A metodologia faz aproximações complementares, sendo a primeira etapa teórica-conceitual e a segunda etapa aplicada e descritiva. A metodologia se divide em duas partes: 1) a revisão bibliográfica para fundamentar e sustentar o referencial teórico e metodológico da tese e 2) parte de aplicação e desenvolvimento: criação da modelagem semântica do vocabulário SKOS e da ontologia OWL.

Nesse sentido, a pesquisa bibliográfica observa os fatos individuais (projetos pilotos, publicações de fontes diversas e ecossistemas científicos-culturais compartilhados) para definir hipóteses gerais de forma indutiva, mapeando e analisando como esses processos estruturam as informações do tema desta pesquisa na perspectiva da Representação e Organização do Conhecimento. Todavia, é necessário realizar o movimento inverso na pesquisa aplicada e descritiva, em direção à fase dedutiva do método científico, justapondo premissas teóricas e hipóteses gerais numa abordagem individual, localizada e delimitada do objeto de pesquisa na forma de modelo conceitual.

Dessa forma, a primeira parte visa realizar uma pesquisa bibliográfica, a partir da revisão de literatura, de modo a auxiliar na identificação do contexto, na explicitação do problema, na seleção de métodos e na delimitação do marco teórico conceitual. Visto que, de modo geral, a revisão de literatura pode ter diferentes finalidades, “a) proporcionar um aprendizado sobre uma determinada área do conhecimento; b) facilitar a identificação e seleção dos métodos e técnicas a serem utilizados pelo pesquisador; c) oferecer subsídios para a redação da introdução e revisão da literatura e redação da discussão do trabalho científico.” (Pizzani *et al.*, 2012, p. 54).

A segunda parte desta pesquisa é aplicada e descritiva, inclui duas etapas, sendo a etapa 1 realizar uma proposta de modelagem de dados abertos conectados (LOD) utilizando SKOS no contexto da representação da imagem em movimento. E a etapa 2 realizar modelagem semântica de uma ontologia para auxiliar a indexação documental da imagem em movimento. Assim, visa a criação de um modelo de tratamento documental baseado na Web Semântica que sirva de base para a representação e recuperação do conhecimento da imagem em movimento no ambiente da web.

Para elaboração da pesquisa bibliográfica recorreu-se à revisão da literatura nas bases de dados da Ciência da Informação. Dessa forma, a partir da análise do tema, foram feitos os processos de análise temática e de conteúdo da literatura da área. Além disso, durante a

estratégia de busca na definição das palavras-chave, foi consultado o tesouro da Ciência da Informação, “para realizar o mapeamento de sinônimos, assim como para traduzir adequadamente os conceitos que integram a questão de revisão para a língua inglesa, visto que as bases de dados bibliográficos internacionais priorizam a língua inglesa como idioma de busca” (Galvão; Ricarte, 2019, p. 66).

Em seguida, realizou-se o mapeamento terminológico, na busca foram inseridos os operadores booleanos para construção das estratégias de busca. Assim, para cada seção foram formuladas estratégias de busca diferentes, com definição de palavras-chave e outras fontes além das bases de dados da área da ciência da informação.

A abordagem metodológica tratou de questões teóricas da ciência da informação, para a representação da imagem em movimento, de modo que, possa ser confrontado com uma abordagem na web e no cenário contemporânea dos estudos de campo. Para isso, também foram selecionados a bibliografia e as referências presentes nas disciplinas do curso de pós-graduação em Ciência da Informação, que auxiliaram na definição do objeto de pesquisa. Bem como, conteúdos dos estágios do Programa de Aperfeiçoamento de Ensino (PAE) do Programa de pós-graduação em Ciência da Informação da Escola de Comunicação e Artes da USP. Conseqüentemente, esse arcabouço teórico condiz com os fundamentos teóricos, metodológicos e conceituais no recorte de abordagens pertinentes a área de organização da informação. A metodologia consistiu em dividir em três tópicos de abordagem, a) organização e representação do conhecimento de imagens em movimento, b) Sistema de organização do conhecimento na web e c) tecnologias da Web Semântica.

Acerca da organização e representação do conhecimento de imagens em movimento, a pesquisa tratou os fundamentos teóricos da representação da imagem, junto com as teorias e metodologias de análise e indexação da imagem. Para isso, recorreu-se em teorias da imagem geral, delimitando a imagem em movimento como um tipo de documento audiovisual e acrescentando os aspectos do tempo e do movimento, transferindo esse processo para o ambiente digital. De tal forma, a literatura tratou a semiótica da imagem e teóricos que desenvolveram métodos para analisar a imagem.

A pesquisa delimitou o arcabouço teórico-epistemológico que analisa o objeto de pesquisa. A relação sujeito-objeto se encontra a) nos questionamentos terminológicos e epistemológicos do desenvolvimento e aplicação de ferramentas semânticas, b) nas metodologias e teorias de análise, representação e indexação da imagem em movimento e c) nos fundamentos teóricos da organização do conhecimento e dos SOC.

Na revisão foram realizadas as seguintes etapas: a elaboração de protocolos de busca; mapeamento e análise sistemática da bibliografia; seleção dos documentos; validação das pesquisas selecionadas a partir dos critérios de relevância; tratamento e coleta de dados para análise de conteúdo. Alguns metadados da literatura científica recuperados nas bases de dados foram corrigidos manualmente concomitantemente à adição no Mendeley.

Do mesmo modo, para a compreensão dos tópicos pesquisados, a proposta de revisão realizou a análise exploratória do tema e o estudo dos artigos encontrados em bases científicas da ciência da informação (LISA, ISTA, BRAPCI e E-lis). Em seguida, elaborou-se o planejamento da estratégia de busca, bem como a formulação da questão de pesquisa e o objetivo da revisão. Dessa forma, o protocolo de pesquisa incluiu as palavras-chave, a estratégia de pesquisa, os critérios de inclusão/exclusão e as bases de dados consultadas. Nas bases de dados consultadas foi feita a recuperação a partir das palavras-chave definidas no protocolo e a exportação dos metadados das bases para um software de gestão de referências. Desse modo, foi possível realizar a codificação do conteúdo dos artigos a partir de códigos e dos descritores, estabelecendo critérios de classificação, representação e a organização dos conceitos, termos e palavras-chave dos documentos selecionados.

Critérios de inclusão

Busca na base de dados:

- a) palavras-chave: para cada base foram selecionadas palavras-chave específicas. Ver os quadros 40 e 41;
- b) bases de dados consultadas: LISA, ISTA, LISTA, E-LIS; pesquisa multilíngue;
- c) lista de assuntos - área do tratamento da informação e SOC.
- d) apenas documentos recuperados de bases de dados nacionais e internacionais.

Mapeamento, análise da literatura e seleção:

- a) mapeamento de literatura científica dos assuntos da tese.
- b) seleção e análise de artigos os assuntos da tese;
- c) seleção de resultados com o termo exato entre aspas em qualquer campo de busca;
- d) seleção de literatura sobre o tema.

Extração de dados:

- a) identificar problemas acerca da temática da tese;
- b) listar os estudos de caso e mapeamentos sobre o tema;
- c) listar as palavras-chave encontradas na literatura.

Cr terios de exclus o

Busca na base de dados:

- a) publica es fora do tema e do escopo de cada se o da tese.

Mapeamento e an lise da literatura:

- a) n o tem como tema principal os assuntos das se es da tese;

Assim, inicialmente, a primeira etapa da metodologia de pesquisa nas bases, englobou o tema da organiza o do conhecimento a partir dos SOC e da representa o tem tica da imagem em movimento, em artigos presentes nas bases de dados da Ci ncia da Informa o. Utilizando uma abordagem de busca distinta para cada base de dados, conforme ilustrado no quadro 41. A abordagem teve car ter explorat rio e descritivo, tendo como finalidade identificar e compreender nos trabalhos referentes    rea da Ci ncia da Informa o, os processos de representa o tem tica da imagem em movimento. Atrav s dos SOC de institui es que possuem bases de dados com documentos de imagem em movimento, como bibliotecas, arquivos, centros de mem ria e institui es do conhecimento. Como citado na pesquisa bibliogr fica, a imagem em movimento   tipo de documento audiovisual, portanto, al m do termo “imagem em movimento”, foram inclusos os termos “documenta o audiovisual” e “audiovisual”, foram inclu dos apenas documentos que tratavam especificamente.

Quadro 36 - Busca estratégica nas bases da ciência da informação

BASES CIENTÍFICAS DA CIÊNCIA DA INFORMAÇÃO	PALAVRAS-CHAVE INSERIDAS NA ESTRATÉGIA DE BUSCA	Publicações incluídas
Base de dados de Periódicos em Ciência da Informação (BRAPCI)	- "documentação audiovisual"; - "análise de conteúdo" AND audiovisual; - "análise de conteúdo" AND "imagem em movimento"; - indexação AND audiovisual; - indexação AND "imagens em movimento"; - indexação AND filmes;	(CORDEIRO; AMÂNCIO, 2005) (CORDEIRO; LA BARRE, 2011); (CORDEIRO, 2013); (GONÇALVES, 2018); (MOURA <i>et al.</i> , 2018); (SANTOS <i>et al.</i> , 2018); (SILVA; CORDEIRO, 2018); (SUNDSTRÖM; MORAES; ALBUQUERQUE, 2019); (SOUZA, 2020);
Library, Information Science & Technology Abstracts with Full Text (LISA)	"content analysis" AND su(audiovisual OR film OR "moving image"); ab("audiovisual documentation"); su("audiovisual documentation"); su(indexing AND (film OR audiovisual OR "moving image")); su(thesauri AND (film OR audiovisual OR "moving image")); ti("audiovisual documentation");	(TURNER, 1990); (TURNER, 1994); (PENA, 2003); (DE FATIMA SOARES; JUVENAL FREIRE, 2005); (CALDERASERRANO; SÁNCHEZ-JIMÉNEZ, 2009); (DOMÍNGUEZ-DELGADO; LÓPEZHERNÁNDEZ, 2016);
Library, Information Science & Technology Abstracts with full text (LISTA)	- SU (audiovisual OR film OR "moving image") AND SU documentation; - SU (audiovisual OR film OR "moving image") AND SU indexing;	(BRAVO, 2005); (TURNER; MATHIEU, 2007)
Information Science & Technology Abstracts (ISTA)	- SU (audiovisual OR film OR "moving image") AND SU documentation; - SU (audiovisual OR film OR "moving image") AND SU indexing;	(LÓPEZ-YEPES; SÁNCHEZJIMÉNEZ; PÉREZ-AGÜERA, 2003); (LICHTENSTEIN; PLANK; NEUMANN, 2014)
Eprints in Library and Information Science (ELIS)	"audiovisual documentation"; audiovisual;	(LÓPEZ-YEPES; SÁNCHEZJIMÉNEZ; PÉREZ-AGÜERA, 2003); (ROCKO, 2009); (SASTRE, 2009); (CASTRO, 2011); (CALDERA-SERRANO; CAROCASTRO, 2013); (GONZÁLEZ, 2013) (ZABALA-VÁZQUEZ; SÁNCHEZGALÁN, 2014); (LÓPEZ-YEPES, 2017); (PASTORSÁNCHEZ <i>et al.</i> , 2019);

Fonte: Elaborado pela autora.

A análise dos resultados (Quadro 41), foi auxiliada com o uso de uma ferramenta computacional, o software de análise de qualitativa WebQDA⁶⁶, o que possibilitou maior agilidade devido a alguns processos automatizados. Assim, na parte referente a seleção das fontes é possível através do recurso Notas, inserir automaticamente os metadados das bases de dados consultadas, para o software fazer a codificação dos metadados inseridos em descritores dos documentos. Porém, como não são todos os documentos que geram automaticamente metadados satisfatórios, foi necessário inserir alguns descritores manualmente, também foi realizada manualmente a eliminação dos descritores pertencentes ao grupo dos artigos excluídos. Os descritores codificados foram divididos nas seguintes categorias: Tipo (página web, artigo, relatório), Ano, Autores, Palavras-chave e Estado (Incluído, Excluído, Duplicado).

Outro processo automatizado da fase de busca, foram os resultados quantitativos da pesquisa, sendo os documentos recuperados separados nos seguintes porcentagens e números: 22 documentos Incluídos (13,3 %), 46 documentos duplicados (27,9 %) e 97 Excluídos (58,8 %). Outra análise automática foram os tipos de documentos: sendo 22 artigos (95,7 %) e 1 relatório (4,3 %), entretanto o resultado dos documentos incluídos fica acima do resultado real, pois na análise automática de tipos, aparecem 3 artigos duplicados nos resultados individuais de duas bases de dados.

Por conseguinte, a segunda etapa da metodologia, realizou a busca na literatura para os conceitos e definições que abordam os dados abertos e LOD. A busca envolveu os conceitos de enriquecimento semântico, SOC na web e os “KOS LOD. A coleta de dados foi realizada na base científica da ciência da informação LISA, na língua inglesa e a estratégia de busca incluiu palavras-chave definidas para elaboração da revisão de literatura. Desse modo, as palavras-chave inseridas na estratégia de busca foram (Quadro 42):

Quadro 37 - Palavras-chave inseridas na estratégia

Palavras-chave inseridas na estratégia
Semantic enrichment AND (Knowledge Organization Systems OR linked Open Data OR semantic web OR controlled vocabularies);
Linked open data AND digital humanities;
Knowledge Organization Systems AND digital humanities.

Fonte: Elaborado pela autora.

Posteriormente, os metadados dos artigos recuperados foram adicionados na plataforma Mendeley, em seguida foi exportado os dados bibliográficos em formato XML do Mendeley e

⁶⁶ webQDA é um software de análise qualitativa de dados, baseado na web. Disponível em: <https://www.webqda.net/>

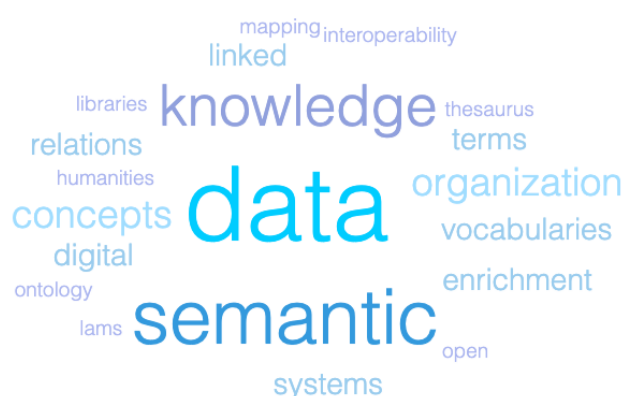
importados no software de análise qualitativa WebQDA, em Fontes->Notas. Logo, foi elaborado a codificação das partes do conteúdo dos trabalhos recuperados, através da representação e a organização dos conceitos, termos e palavras-chave dos documentos selecionados. Os trabalhos foram analisados e estruturados conforme as seguintes informações: Tema do artigo, Objetivos, Metodologia e Resultados. Além disso foi elaborado a classificação de cada documento na totalidade, exemplo:

- Tema principal: SOC semânticos (SKOS) ou KOS LOD
- Origem: nacional ou internacional
- Idioma: inglês ou português ou espanhol
- O documento trata da representação de imagens: SIM ou NÃO

Assim, foi aplicado a funcionalidade *Questionamentos* do WebQDA para estabelecer as 20 palavras-chave mais utilizadas nas publicações, para isso estabeleceu-se uma quantidade de 20 palavras para recuperação, com o mínimo de 4 caracteres, inserindo as *tags* (espaço próprio para isso) das preposições, artigos e palavras que serão excluídos no processamento das palavras-chave, ainda foram excluídos os sinônimos multilíngues, escolhendo os termos em inglês, também foram excluídos as flexões de gênero e número dos sinônimos. Todas as palavras-chave utilizadas foram definidas automaticamente pelo software na parte dedicada aos descritores dos 20 documentos incluídos no software para a revisão da bibliografia. A seguir dois resultados como exemplos das codificações possíveis de serem realizadas na figura 47.

Quadro 38 – Lista e nuvem das 20 palavras-chave mais utilizadas nas publicações

Palavra	Repetição	Carateres
data	313	4
semantic	216	8
knowledge	157	9
organization	105	12
concepts	96	8
vocabularies	83	12
digital	83	7
relations	73	9
linked	67	6
enrichment	64	10
systems	62	7
terms	61	5
humanities	57	10
libraries	56	9
interoperability	49	16
lams	48	4
thesaurus	46	9
mapping	45	7
open	44	4
ontology	42	8



Utilizado software webQDA. Fonte: Elaborado pela autora.

A terceira etapa de busca da metodologia de pesquisa, delimitou como assunto as ontologias OWL do W3C usada no ambiente web, para interoperabilidade, criação de axiomas,

inferências, a partir da lógica descritiva e formal. Sendo usado para definir competências de um domínio do conhecimento a partir da terminologia, classes, subclasses, instâncias e propriedades.

A coleta de dados para o desenvolvimento dessa etapa foi realizada na base científica da Ciência da informação LISA e a estratégia de busca incluiu palavras-chave no idioma inglês, definidas para elaboração da revisão de literatura. Desse modo, as palavras-chave inseridas na estratégia de busca foram:

-Ontologies AND Conceptual Modeling.

-Organization of information AND Documentary languages AND Knowledge organization systems.

As informações foram divididas em tópicos para compreensão da Web Semântica desde as abordagens de lógica descritiva, a partir da definição e demonstração de exemplos. Seguido, da descrição das recomendações do W3C, RDF, RDF Schema, OWL, SPARQL. Finalizando com um tópico sobre metodologia 101 de construção de ontologia utilizando o software *protegé*. A pesquisa delimitou em entender como funciona a estrutura da Web Semântica em relação a suas estruturas e principalmente a construção de ontologias de domínio e modelo de dados vinculados.

Da mesma forma, a quarta etapa da metodologia de pesquisa, abordou as principais iniciativas com acervos de imagem em movimento ou com conteúdo audiovisual no âmbito das humanidades digitais, com recorte principal acerca da representação desse tipo de recurso no ambiente da web. Foram investigadas as iniciativas de Web Semântica no ambiente de dados estruturados de instituições de autoridade, como a FIAF e o BNF. Bem como, investigou-se a estruturação de dados a partir de dados não estruturados, a partir de iniciativas de projetos pilotos de plataformas e players que envolvem organização do conhecimento da imagem em movimento, anotação semântica e humanidade digitais, internacionais. Como o I-Media-Cities e projetos de anotação manual de vídeo, que são pioneiros nas Humanidades digitais.

Foram estudadas pesquisas recentes com perspectivas emergentes, sobre o tratamento da informação audiovisual no domínio da imagem em movimento, como a *deep annotation image* (anotação de imagem profunda) e a anotação de dados abertos para imagem, os estudos demonstram como esses processos em ascensão podem criar e desenvolver metodologias para diversas disciplinas que tratam a imagem e a imagem em movimento. Essas abordagens ajudam a estabelecer caminhos para a que organização e representação da imagem em movimento no ambiente da web possam ser mais bem exploradas, a partir dos dados conectados e abertos, numa perspectiva crítica e pragmática das humanidades digitais.

De forma que, a coleta de dados foi realizada nas bases de dados da Ciência da informação. As palavras-chave:

- SOC AND *semantic web*; *semantic web* AND audiovisual.

Dessa forma, foram os seguintes dados: iniciativas da FIAF a partir do grupo de estudo para *linked data* da federação e o projeto da biblioteca nacional da França, o BNF Data, que é um buscador únicos com vários catálogos da instituição mapeados em RDF e FRBR, também foi adicionado na ontologia o SKOS.

A metodologia buscou iniciativas de instituições nas áreas das bibliotecas, arquivos e museus, com acervos para o audiovisual e com projetos-pilotos sobre acervos na Web Semântica. Além de trazer, as pesquisas recentes sobre organização do conhecimento da imagem, que podem ser relacionadas a imagem em movimento na web, como o *deep annotation image* e os KOS LOD.

A pesquisa envolveu a fase de aplicação e desenvolvimento de propostas de modelagem semântica, ao estruturar a criação da modelagem semântica do vocabulário SKOS e da ontologia OWL. Em vista disso, teve como objetivo construir procedimentos e métodos para a aplicação prática e definir abordagens para problemas específicos. A construção de uma proposta de modelagem semântica baseada em ontologias é pensada para instituições que possuem documentos e informação audiovisual, como bibliotecas, arquivos, museus e instituições do conhecimento. “Muitas vezes, nessa modalidade de pesquisa, os problemas emergem do contexto profissional e podem ser sugeridos pela instituição para que o pesquisador solucione uma situação-problema” (Córdova; Silveira, 2009, p. 35, tradução nossa). Dessa forma, a seção trata-se da pesquisa aplicada e descritiva, dividida em duas etapas intituladas: Modelos de dados LOD para **enriquecimento de SOC de imagem em movimento na web** e modelagem conceitual baseada em ontologias para o **domínio da representação temática da imagem em movimento na web ou como uma linguagem de indexação para o ambiente web**.

Considerando as linguagens de indexação e as especificidades da imagem em movimento, a metodologia compreendeu o estudo e a definição de métodos de análise, síntese e organização de conhecimento documental de imagem em movimento. Para fins de indexação e recuperação em sistemas informacionais, ponto de vista semiótico, iconográfico e da categorização de conceitos.

A partir disso, foi analisado como os processos da organização do conhecimento e representação da informação se integram aos dados abertos e conectados, no contexto da anotação semânticos e o LOD. Logo, a etapa aplicada delineou os passos para o desenvolvimento e criação de micro tesauro e outros tipos de SOC, para sistematizar conceitos

e assuntos temáticos da imagem em movimento, que podem ser usados como KOS LOD e servirem de vocabulário de valor para anotação semântica de imagens em movimento na web.

A anotação semântica da imagem em movimento com KOS LOD, ajuda no desafio de indexar recursos não baseados em texto, uma vez que através da anotação profunda da imagem, ou seja, a utilização de metadados no nível da anotação, os conceitos são anotados como pontos de interesse específicos ou regiões de interesse em uma imagem, não sendo anotada na sua totalidade, ou seja, não se descreve a imagem total (Clarke, 2015). Visto que, a capacidade de descrever a totalidade de uma imagem ou o que ela representa é um dos grandes desafios da indexação de imagens em movimento, devido às dificuldades e limitações de transcodificar a imagem em texto na tentativa de traduzir o conteúdo da imagem. A anotação semântica da imagem em movimento permite estabelecer relações semânticas entre conceitos para a categorização de conjuntos de conceitos no nível da subimagem. A anotação de subimagem pode ser usada como índices de assuntos, através da pesquisa e navegação nas imagens.

A criação de um KOD LOD para ser usado na indexação de imagens em movimento, proporciona o estabelecimento de um tesouro conceitual, proporcionando que conceitos sejam relacionados semanticamente (*broader, narrower e related*). Além das propriedades de mapeamento (*closeMatch, exactMatch, relatedMatch, broaderMatch e narrowerMatch*) entre diferentes esquemas de conceitos, a partir das URI que identificam os conceitos. Ainda, é possível reutilizar e compartilhar de forma distribuída os recursos (conceitos), aplicando a propriedade *skos:conceptscheme* e formar uma rede semântica em SKOS, reutilizando um mesmo conceito em dois ou mais esquemas de conceitos diferentes.

O tesouro baseado em conceitos tem uma estrutura de três níveis:

- (a) nível conceitual: conceitos são identificados e suas inter-relações estabelecidas;
- (b) nível de correspondência terminológica: os termos são associados (preferenciais ou não preferenciais) aos seus respectivos conceitos
- (c) nível lexical: são definidas as relações lexicais para interligar os termos.

A pesquisa estruturou um esquema de conceitos modelado em SKOS, chamado “Esquema de conceito das categorias de representação da imagem em movimento”. Partindo da etapa de categorização do esquema de conceitos: definição dos conceitos de acordo a metodologia QUEM, QUANDO, ONDE, O QUE. Em seguida, realizou-se a delimitação da área, público-alvo, classificação, levantamento das fontes, forma de apresentação (sistemática/alfabética). Em seguida, realizada a construção de parte do esquema de conceitos usando o modelo de dados SKOS, que englobou o desenvolvimento de partes de esquemas de conceitos das categorias QUEM e ONDE, da definição dos conceitos como sendo da classe

skos:concept e atribuí-se os rótulos lexicais. Além da criação de um esquema de conceito para cada categoria (skos:ConceptScheme), utilizando as propriedades skos:inScheme, skos:hasTopConcept e skos:topConceptOf. Em continuação, foram definidas as relações semânticas (*broader*, *narrower* e *related*) e inseridas as propriedades de documentação de cada recurso (conceito) do vocabulário. Assim como, estabeleceu-se links de mapeamento (alinhamento) entre os conceitos SKOS em diferentes esquemas de conceito (LCHS, DBpedia, Wikidata, AAT), para o enriquecimento semântico do esquema com as propriedades de mapeamento para recursos externos de LOD/ skos:closeMatch. Também ocorreu a modelagem SKOS com formato de serialização Turtle (N3) RDF. Para isso foram feitos registro das URI dos principais recursos do esquema no domínio purl.org e hash URI (slash URI e w3id), a validação do modelo de dados SKOS (no SKOS *testing tool version*⁶⁷), a validação da serialização Turtle e dos triplos RDF no IDLab *Turtle Validator*⁶⁸. Por fim, os arquivos serializados em Turtle foram feitos uploads no SKOS Play⁶⁹, um aplicativo gratuito para renderizar e visualizar dicionários de sinônimos, taxonomias ou vocabulários controlados expressos em SKOS.

O enfoque teve como objetivo geral realizar a modelagem semântica de ontologia para representar formalmente a conceitualização do domínio da indexação da imagem em movimento, para auxiliar processos de indexação de imagem em movimento no ambiente web, integrando os processos de anotação semântica.

Dessa forma, os procedimentos e tarefas que foram realizados na etapa de modelagem conceitual iniciaram com a definição das classes, instâncias, propriedades e atributos do domínio de representação temática da imagem em movimento, a partir dos processos teóricos-conceituais e metodológicos de representação temática da imagem em movimento. Em seguida, foi desenvolvido a modelagem da ontologia de domínio temático da imagem em movimento, usando a metodologia 101 no Protegé. De modo, que foi determinado o escopo, considerado o reuso, enumerado os termos, definir as classes, as propriedades e as restrições, além disso, criado as instâncias. Na modelagem da ontologia foi utilizado as recomendações, documentos e padrões do W3C, como RDF, SKOS, OWL, Web Annotation Data Model do W3C, Media Fragments URI e o Web Annotation Data Model. E mapeado e definido as relações semanticamente expressivas para indexar a imagem em movimento, seja a imagem total ou

⁶⁷ SKOS testing tool version. Disponível em: <https://skos-play.sparna.fr/skos-testing-tool/>

⁶⁸ IDLab Turtle Validator. Disponível em: <http://ttl.summerofcode.be/>

⁶⁹ SKOS Play. Disponível em: <https://skos-play.sparna.fr/play/>

anotação de sub-imagem, utilizando a terminologia de um vocabulário controlado por KOS LOD na ontologia. Ademais, as imagens foram anotadas usando vocabulários SOC.

De tal modo, a modelagem conceitual aborda três perspectivas teórico-metodológicas, a representação temática da imagem, os SOC e o desenvolvimento de ontologias de domínio. A abordagem metodológica da pesquisa concentrou-se na realização da modelagem semântica de microtesauros conceituais na forma de esquema de conceitos, para representar o conhecimento de recursos audiovisuais do tipo imagem em movimento e visando publicação no ambiente web.

Para isso, estruturaram-se esquemas de conceitos (skos:concepScheme) modelado em SKOS de quatro categorias de representação da imagem em movimento, para a organização e representação do conhecimento e a modelagem de dados no ambiente web.

A elaboração do esquema de conceitos teve como ponto de partida a representação temática de assuntos da imagem em movimento e os métodos de categorização dos conceitos dos documentos audiovisuais e fotográficos. Partindo do corpus teórico acerca da indexação da imagem em movimento e fotográfica, foram analisadas as formas de organização e representação do conhecimento no âmbito da identificação, definição e tradução de conceitos. Bem como, estudados os modos de conceituação e formação de enunciados sobre conceitos no domínio da indexação de recursos com imagens em movimento.

Dessa forma, a partir dos quadros para a representação genérica e específica da imagem fotográfica de acordo ao quadro 44 (Smit, 1996) foi possível encontrar modos de indexar imagens fotográficas. A partir das categorias de representação da imagem e de seus níveis iconográficos para definir a relação hierárquica gênero-espécie e todo-parte.

Quadro 39 - Quadro das categorias de representação da imagem fotográfica

Categoria	Definição geral	DE genérico	DE específico	SOBRE
QUEM	Animado e inanimado, objetos e seres concretos	Esta imagem é de quem? de que objetos? De que seres?	De quem, especificamente, se trata?	Os seres ou objetos funcionam como símbolos de outros seres ou objetos? Representam a manifestação de uma abstração?
	Exemplo	Ponte	Ponte das Bandeiras	Urbanização
	Exemplo			Arquitetura dos anos 40
ONDE	Onde está a imagem no espaço?	Tipos de lugares geográficos, arquitetônicos ou cosmográficos	Nomes de lugares geográficos, arquitetônicos ou cosmográficos	O lugar simboliza um lugar diferente ou mítico? O lugar representa a manifestação de um pensamento abstrato?
	Exemplo	Selva	Amazonas	Paraíso (supõe um contexto que permita esta interpretação)
	Exemplo	Perfil de cidade	Paris	Monte Olimpo (como o exemplo anterior)
QUANDO	Tempo linear ou cíclico, datas e períodos específicos, tempos recorrentes	Tempo cíclico	Tempo linear	Raramente utilizado, representa o tempo a manifestação de uma idéia abstrata ou símbolo?
	Exemplo	Primavera	1996	Esperança, fertilidade, juventude
O QUE	O que os objetos e seres estão fazendo? Ações, eventos, emoções	Ações, eventos	Eventos individualmente nomeados	Que idéias abstratas (ou emoções) estas ações podem simbolizar?
	Exemplo	Morte	Pietá	Dor (emoção)
	Exemplo	Jogo de futebol (ação)	Copa do Mundo 1996	Esporte

Fonte: Smit (1996).

Também, realizou-se análise em fontes de informação como, vocabulários, tesouros específicos do cinema e relatos de pesquisas que envolveram processos de indexação da imagem em movimento. Para compreender os desafios, temáticas, conceitos e termos que envolvem os processos de indexação e a construção de vocabulários controlados no domínio da imagem em movimento.

Assim, foram desenvolvidos cenários de representação de conceitos e de seus relacionamentos semânticos, que podem ser utilizados em aplicações para a classificação, indexação ou busca de um documento audiovisual com imagem em movimento no ambiente web.

No primeiro cenário, o esquema de conceitos foi estruturado a partir de um vocabulário controlado na forma de lista organizada e pré-definida de conceitos. Visto que o SKOS permite a modelagem de todos os tipos de SOC para publicação no ambiente web. Delimitou-se o tratamento de conceitos pertencentes a “Categoria de representação da imagem em movimento QUEM”, que trata indivíduos “animados e inanimados, objetos e seres concretos” (Smit, 1996). Dessa forma, estabeleceu-se o conceito Personagem como o de maior hierarquia e os tipos de personagens como os conceitos coordenados no nível subordinado.

O vocabulário da USP, figura 47, foi a fonte de informação usada para definir conceitos representativos e obter garantia literária. A macroestrutura do vocabulário da USP apresenta as

relações Lógico-Semânticas com as áreas e terminologia em seus diferentes níveis, sendo que personagem se refere a hierarquias dos descritores Arte/Artes do Espetáculo/Representação/Personagens:

Figura 47 - Macroestrutura do vocabulário USP

DEDALUS	Código da Macroestrutura	Assunto
	CH741	- ARTES
	CH741.2	- ARTES DO ESPETÁCULO
	CH741.2.2	- REPRESENTAÇÃO
	CH741.2.2.6	- PERSONAGENS <==
	CH741.2.2.6.1	- ADÚLTEROS
	CH741.2.2.6.2	- ALCOÓLATRAS
	CH741.2.2.6.3	- ALIENÍGENAS
	CH741.2.2.6.4	- AMIGOS
	CH741.2.2.6.5	- AMNÉSICOS
	CH741.2.2.6.6	- ANDROIDES
	CH741.2.2.6.7	- ANÕES
	CH741.2.2.6.8	- ARISTOCRATAS
	CH741.2.2.6.9	- ASSALTANTES
	CH741.2.2.6.10	- ASSASSINOS
	CH741.2.2.6.11	- ATIVISTAS POLÍTICOS
	CH741.2.2.6.12	- BÉBADOS
	CH741.2.2.6.13	- BICHEIROS
	CH741.2.2.6.14	- BICHO-PAPÃO
	CH741.2.2.6.15	- BOI-TATÁ
	CH741.2.2.6.16	- BURGUESES
	CH741.2.2.6.17	- CAIPIRAS
	CH741.2.2.6.18	- CAMARTESES

Fonte: Vocabulário USP (2023).

O vocabulário controlado da USP é uma lista organizada e pré-definida de termos e não estabelece as relações entre termos e/ou conceitos como num tesouro, mas as hierarquias da macroestrutura auxiliam na coleta dos tipos de personagem que fazem parte do domínio da representação das artes do espetáculo. O termo personagem pode ser definido com um nome para seres presentes em diversos tipos de artes narrativas, de tal forma, existe a associação entre personagem e arquétipo, presentes no universo da diegese e da narrativa. A categorização de personagens em arquétipos decorre de conceituações no âmbito de expressões artísticas e manifestações culturais, desse modo, contém os aspectos culturais e simbólicos.

Os arquétipos de personagens da diegese narrativa, estabelecem classes que englobam o compartilhamento de características, traços e padrões, que auxiliam na identificação do personagem e representa um conjunto específico de comportamentos. Por conseguinte, a personagem (ou arquétipo) do universo diegético é um conceito que tem relação com os conceitos que representam os personagens específicos.

Assim, definiu-se a propriedade skos:hasTopConcept para vincular o esquema de conceito ao único conceito SKOS que está no topo das relações hierárquicas do esquema, o conceito de personagem. Logo, aplicou-se a propriedade skos:hasTopConcept ao conceito Personagem (quem:personagens), além disso, foi aplicado a propriedade skos:broader para indicar a relação semântica dos demais conceitos com o conceito do topo quem:personagens.

No segundo cenário o esquema de conceitos foi mapeado no padrão internacional para tesouros e interoperabilidade com outros vocabulários, a norma ISO 25964⁷⁰. Dessa forma, estruturou-se um micro tesouro intitulado “Categoria de representação da imagem em movimento ONDE”, representando os relacionamentos hierárquicos (gênero/espécie, todo/parte) entre conceitos e indicando no esquema mais de um conceito de topo, dessa forma, aplicando a propriedade `skos:hasTopConcept` para os diversos conceitos mais amplos e a propriedade `skos:broader` nos conceitos mais estritos correspondentes para indicar o conceito mais amplo correspondente.

Além disso, foi elaborado um segundo esquema “Esquema de conceitos das categorias de representação da imagem em movimento (QUEM /QUANDO/ONDE/OQUE)”, que foi estendido e vinculado ao esquema de conceitos “Categoria de representação da imagem em movimento ONDE”. Para isso, foi reutilizado conceitos do micro tesouro, empregando relações semânticas entre os conceitos para formar uma rede semântica em SKOS, visto que, a propriedade `skos:inScheme` permite a reutilização de conceitos. Consequentemente, alguns conceitos pertencem a mais de um esquema ao mesmo tempo, para ter maior abrangência de um domínio ou subdomínio, fazendo o reuso de alguns de seus conceitos.

Ademais, foram aplicadas as propriedades de mapeamento SKOS (`skos:broadMatch`, `skos:narrowMatch`, `skos:relatedMatch`) que foram usadas para estabelecer links de alinhamento entre conceitos SKOS dos diferentes esquemas de conceito: micro tesouro “Categoria de representação da imagem em movimento ONDE” e o “Esquema de conceitos das categorias de representação da imagem em movimento (QUEM /QUANDO/ONDE/OQUE)”. No terceiro cenário foi desenvolvida uma coleção de conceitos usando a classe `skos:Collection`.

A construção do vocabulário aplicou o modelo de dados SKOS, a partir das práticas do linked data para SOC.

- Registro das URI dos principais recursos do esquema no domínio `purl.org`
- Validação da serialização Turtle e dos triplos RDF.
- Validação da modelagem SKOS

Existe uma grande variedade de ferramentas de código aberto e proprietárias para trabalhar com vocabulários SKOS, algumas especificamente projetadas para o formato SKOS. Isso inclui ferramentas para desenvolvimento, validação, pesquisa e navegação e fornecimento de interface web para cada conceito e permitem o uso do SKOS na indexação e pesquisa.

⁷⁰ O experimentou mapeou a norma de forma direta e não foi utilizada a extensão do SKOS ISO-THES

Em vista disso, a validação do modelo de dados SKOS foi utilizado a ferramenta SKOS testing tool version. Para a validação da serialização Turtle e dos triplos RDF foi usado a ferramenta IDLab Turtle Validator. Os SOC desenvolvidos foram testados no SKOS Play, um aplicativo gratuito para renderizar e visualizar dicionários de sinônimos, taxonomias ou vocabulários controlados expressos em SKOS. Igualmente, a modelagem de SOC no formato SKOS foi experimentada através de um conversor de Excel para RDF, o “SKOS Play! Convert”, para gerar arquivos RDF a partir de planilhas Excel estruturadas de forma específica e com recursos específicos para SKOS.

A metodologia da pesquisa explorou a documentação técnica do SKOS, através de modelos para expressar a estrutura básica e o conteúdo de esquemas conceituais como tesouros. E conseqüentemente, aplicando experiências em modelagem SKOS para publicar dados na web. Os procedimentos aplicados englobaram o mapeamento (alinhamento) de recursos conceituais em esquemas conceituais e o agrupamento de coleções rotuladas ou ordenadas (Coleção). Ademais, estendeu o vocabulário SKOS combinando com outros vocabulários de modelagem.

6 RESULTADOS

Os resultados da parte aplicada da tese foram as modelagens dos dados de tesouros conceituais na forma de esquemas de conceitos SKOS, explorando as possibilidades de mapeamento e de formação de redes semânticas entre conceitos de diferentes esquemas de conceitos. Ademais, incluíram a modelagem da ontologia “Categorias de Representação da Imagem em Movimento (CRIM)”, que conceituou o processo de representação temática (indexação da imagem em movimento).

6.1 ESQUEMA DE CONCEITO “QUEM”

Dessa forma, o primeiro resultado⁷¹ foi o **SOC 1**, que é um esquema SKOS (`skos:conceptScheme`) na forma de lista organizada e pré-definida de conceitos, modelado no formato de serialização Turtle, intitulado “Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento”, (fig.49) onde foi escolhido um único conceito de topo para aplicar a propriedade `skos:TopConceptOf`, o conceito Personagem (`quem:personagens`). Além disso, foi aplicado a propriedade `skos:broader` nos demais conceitos para indicar que `quem:personagens` tem relação hierárquica de gênero com os demais conceitos do esquema.

Figura 48 - SOC 1: Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento

```

personagem_concept_scheme.ttl
1 @prefix quem: <https://purl.org/purl/ecmi/categoria-personagens#> .
2 @prefix dct: <http://purl.org/dc/terms/> .
3 @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
4 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
5 @prefix wd: <http://www.wikidata.org/entity/> .
6
7 quem: a skos:ConceptScheme ;
8   dc:title "Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento" ;
9   dc:creator "Denise Cavalcante" ;
10  dct:created "2022-06-15"^^xsd:date ;
11  dct:language "pt" ;
12  skos:hasTopConcept quem:personagens .
13
14 quem:personagens a skos:Concept ;
15   skos:prefLabel "personagens"@pt ;
16   skos:prefLabel "character"@en ;
17   skos:narrower quem:adultero, quem:alcoolatra, quem:alienigenas, quem:amigos, quem:amnesicos, quem:androides,
18   skos:note "conceitos retirados do vocabulário da USP"@pt;
19   skos:topConceptOf quem: .
20

```

SOC 1

Concept Scheme

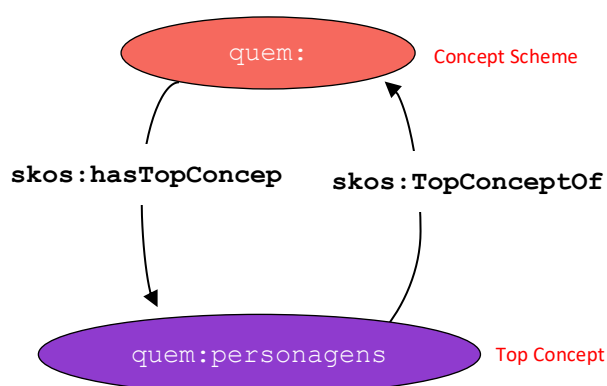
Top Concept

Fonte: Elaborado pela autora.

⁷¹ O arquivo SKOS codificado em um Turtle (ttl) pode ser baixado no repositório github das autoras ou acessado o arquivo .TXT no link <https://sendeyo.com/updownload/file/script/6465961f4388d3547d093102edd633a8.txt>

O esquema de conceito “quem:” tem uma hierarquia de generalização mais ampla/estrita, dessa forma, foi usada a propriedade `skos:hasTopConcept` para afirmar um link entre o esquema de conceito e os conceitos que são os conceitos de nível superior na hierarquia de generalização (Fig.49):

Figura 49 - Propriedade `skos:hasTopConcept` e `skos:TopConceptOf`



Fonte: Elaborado pela autora.

Em seguida foram criados os conceitos de nível inferior do esquema de conceitos, usando a propriedade `skos:inScheme`, como na figura 50:

Figura 50 - Propriedade `skos:inScheme`

```

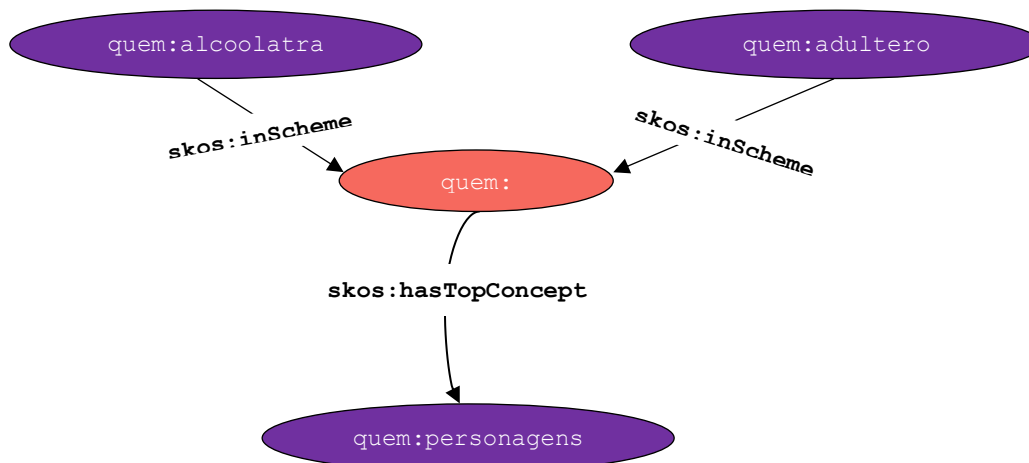
21 quem:adultero a skos:Concept ;
22   skos:prefLabel "adultero"@pt ;
23   skos:prefLabel "adulterer"@en ;
24   skos:altLabel "infiel"@pt;
25   skos:broader quem:personagens ;
26   skos:note "conceitos retirados do vocabulário da USP"@pt;
27   skos:inScheme quem: .
28
29 quem:alcoólatra a skos:Concept ;
30   skos:prefLabel "alcoólatra"@pt ;
31   skos:prefLabel "alcoholic"@en ;
32   skos:altLabel "alcoolista"@pt ;
33   skos:hiddenLabel "bebado"@pt ;
34   skos:broader quem:personagens ;
35   skos:note "conceitos retirados do vocabulário da USP"@pt;
36   skos:inScheme quem: .

```

Fonte: Elaborado pela autora.

No nível conceitual os conceitos são definidos em relação a outros conceitos. Portanto, um 'esquema de conceito' é definido como um conjunto de conceitos e declarações sobre relações semânticas entre esses conceitos (Fig.51).

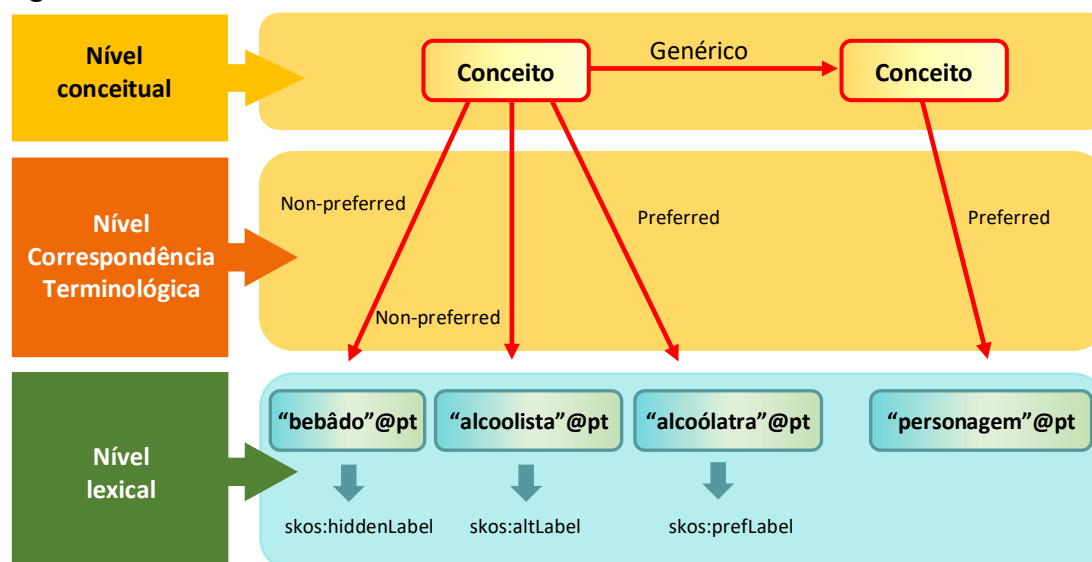
Figura 51 - Grafo com a propriedade skos:inScheme



Fonte: Elaborado pela autora.

Dessa forma, os conceitos do esquema estão no nível conceitual, estabelecendo relações hierárquicas e associativas entre si. No nível de correspondência terminológica estão as propriedades que definem o papel que cada termo designados pelos rótulos (*labels*). Por fim, no nível lexical estão os rótulos com termos da linguagem natural (Fig.52).

Figura 52 - Tesouro conceitual



Fonte: Adaptado de Pastor-Sanchez, Mendez e Muñoz (2009).

Em seguida, os dados SKOS foram validados usando a plataforma SKOS Testing Tool⁷², desenvolvida através do qSKOS uma ferramenta para encontrar problemas de qualidade nos vocabulários SKOS, indicados pelas cores: verde *OK*, amarelo *WARNING* e vermelha *FAIL*.

As regras da ferramenta para testar o arquivo SKOS são:

- **Referências de notação ambígua**
Localiza conceitos com notações múltiplas ou idênticas dentro do mesmo esquema de conceito.
- **Relações Hierárquicas Cíclicas**
Localiza conceitos hierarquicamente relacionados entre si.
- **Clusters de conceito desconectados**
Localiza conjuntos de conceitos isolados do restante do vocabulário.
- **Violação de rótulos separados**
Localiza recursos com entradas idênticas para diferentes tipos de rótulos.
- **Etiquetas Vazias**
Localiza rótulos vazios ou rótulos contendo apenas espaços em branco.
- **Redundância Hierárquica**
Localiza relações mais amplas/estreitas em vários níveis hierárquicos.
- **Violação do Esquema HTTP URI**
Localiza assuntos triplos que não são URIs HTTP.
- **Cobertura de idioma incompleta**
Encontra conceitos sem descrição em linguagens que estão presentes para outros conceitos.
- **Rótulos preferenciais inconsistentes**
Localiza recursos com mais de um prefLabel por idioma.
- **Rótulos ausentes**
Localiza conceitos e conceptSchemes com rótulos ausentes.
- **Link externo ausente**
Localiza conceitos que não estão vinculados a outros vocabulários na Web.
- **Mau uso das relações de mapeamento**
Localiza conceitos dentro do mesmo esquema de conceito que estão relacionados por uma relação de mapeamento.
- **Sem idiomas comuns**
Verifica idiomas comuns em todos os literais de conceito.
- **Conceitos órfãos**
Localiza todos os conceitos órfãos, ou seja, aqueles que não possuem relacionamentos semânticos com outros conceitos.
- **Marcas de idioma omitidas ou inválidas**
Localiza marcas de idioma omitidas ou inválidas de literais de texto.
- **Etiquetas sobrepostas**
Localiza conceitos com rótulos semelhantes (idênticos).
- **Principais Conceitos Omitidos**
Localiza skos:ConceptSchemes que não têm os principais conceitos definidos.
- **Conflitos de relacionamento**
Abrange a condição S27 do documento de referência SKOS (Associative vs. Hierarchical Relation Clashes).
- **Conceitos Reflexivamente Relacionados**
Encontra conceitos que estão relacionados a si mesmos.
- **Conceitos Relacionados Apenas Transitivamente**
Conceitos relacionados apenas por skos:broaderTransitive ou skos:narrowerTransitive.
- **Principais conceitos com conceitos mais amplos**
Localiza os principais conceitos internos à árvore hierárquica do vocabulário.
- **Conceitos não documentados**
Localiza conceitos que não usam nenhuma propriedade de documentação do SKOS.
- **Caracteres não imprimíveis em etiquetas**
Localiza conceitos com rótulos que contêm caracteres não imprimíveis.
- **Conceitos Relacionados Unidirecionalmente**
Conceitos que não incluem relações recíprocas
- **Recursos SKOS indefinidos**
Localiza novos termos 'inventados' dentro do namespace SKOS ou propriedades obsoletas.
- **Relações associativas sem valor**
Localiza pares de conceitos irmãos que também estão conectados por uma relação associativa.

⁷² SKOS Testing Tool é um frontend para qSKOS de Christian Mader. <https://github.com/cmader/qSKOS>

Em continuidade, o **SOC 1** foi renderizado (fig.53) utilizando a plataforma e as ferramentas do **SKOS Play!**, sendo os 12 conceitos processados pela interface. A primeira renderização foi formato de saída de Índice alfabético (lista todos os rótulos preferenciais e alternativos no idioma selecionado).

Figura 53 - Índice alfabético renderizado na ferramenta SKOS Play!
Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento

quem: : A-Z

- **adultero**
- **alcoólatra**
- **alienígenas**
- **amigos**
- **amnésicos**
- **androides**
- **anões**
- **aristocratas**
- **assaltantes**
- **assassinos**
- **ativistas**

Fonte: Elaborado pela autora.

A segunda renderização (fig. 54) foi no formato de Índice alfabético, expandido (índice alfabético com atributos em cada entrada: notação, mais amplo, mais estreito, relacionado, todas as notas).

Figura 54 - Índice alfabético expandido renderizado na ferramenta SKOS Play!

Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento

quem: : A-Z

- **adultero**
BT: personagens
TT: personagens
- **alcoólatra**
BT: personagens
TT: personagens
- **alienígenas**
BT: personagens
TT: personagens
- **amigos**
BT: personagens
TT: personagens
- **amnésicos**
BT: personagens
TT: personagens
- **androides**
BT: personagens
TT: personagens
- **anões**
BT: personagens
TT: personagens
- **aristocratas**
BT: personagens
TT: personagens
- **assaltantes**
BT: personagens
TT: personagens
- **assassinos**
BT: personagens
TT: personagens
- **ativistas**
BT: personagens
TT: personagens

Fonte: Elaborado pela autora.

A terceira renderização (fig.55) foi formato de saída de Árvore hierárquica – estática (cria uma representação em árvore estática dos conceitos usando o rótulo preferido no idioma selecionado.)

Figura 55 - Árvore hierárquica estática renderizada na ferramenta SKOS Play!

Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento

- <https://purl.org/purl/ecmi/categoria-personagens#>
 - personagens
 - [adultero](#)
 - [alcoólatra](#)
 - [alienígenas](#)
 - [amigos](#)
 - [amnésicos](#)
 - [androides](#)
 - [anões](#)
 - [aristocratas](#)
 - [assaltantes](#)
 - [assassinos](#)
 - [ativistas](#)

Fonte: Elaborado pela autora.

A quarta renderização (fig.56) foi formato de saída Índice KWIC (cria um índice 'KeyWord In Context' com base nas palavras dos rótulos no idioma selecionado).

Figura 56 - Índice KWIC renderizado na ferramenta SKOS Play!

Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento

adultero
alcoólatra
alienígenas
amigos
amnésicos
androides
anões
aristocratas
assaltantes
assassinos
ativistas

Fonte: Elaborado pela autora.

A quinta renderização foi em formato de saída Mapeamentos (Liste mapeamentos para outros dicionários de sinônimos, em uma única tabela ordenada alfabeticamente, com uma coluna extra indicando o dicionário de sinônimos de destino). Conseqüentemente, ocorreu o estabelecimento de links de mapeamento (alinhamento) entre os conceitos SKOS com a propriedade de mapeamento `skos:closeMatch` para o recurso externo **Wikidata** (fig.57).

Figura 57 - Mapeamento entre conceitos renderizado na ferramenta SKOS Play!

Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento

Alignment with wde:

source concept		target concept
androides	close	wde:Q181787
assassinos	close	wde:Q931260
ativistas	close	wde:Q15253558

Fonte: Elaborado pela autora.

Além da modelagem SKOS diretamente no editor VS Studio Code em serialização Turtle, também foi testado e aplicado o conversor de Excel para RDF, o “SKOS Play! Convert”, estruturando uma planilha Excel com recursos específicos para SKOS e seguindo a estrutura obrigatória da ferramenta. Esse modelo simplifica a estruturação em SKOS, sem a necessidade de conhecer a sintaxe da codificação Turtle.

Figura 58 - Planilha de conversão de Excel para RDF na estrutura do SKOS Play! Convert

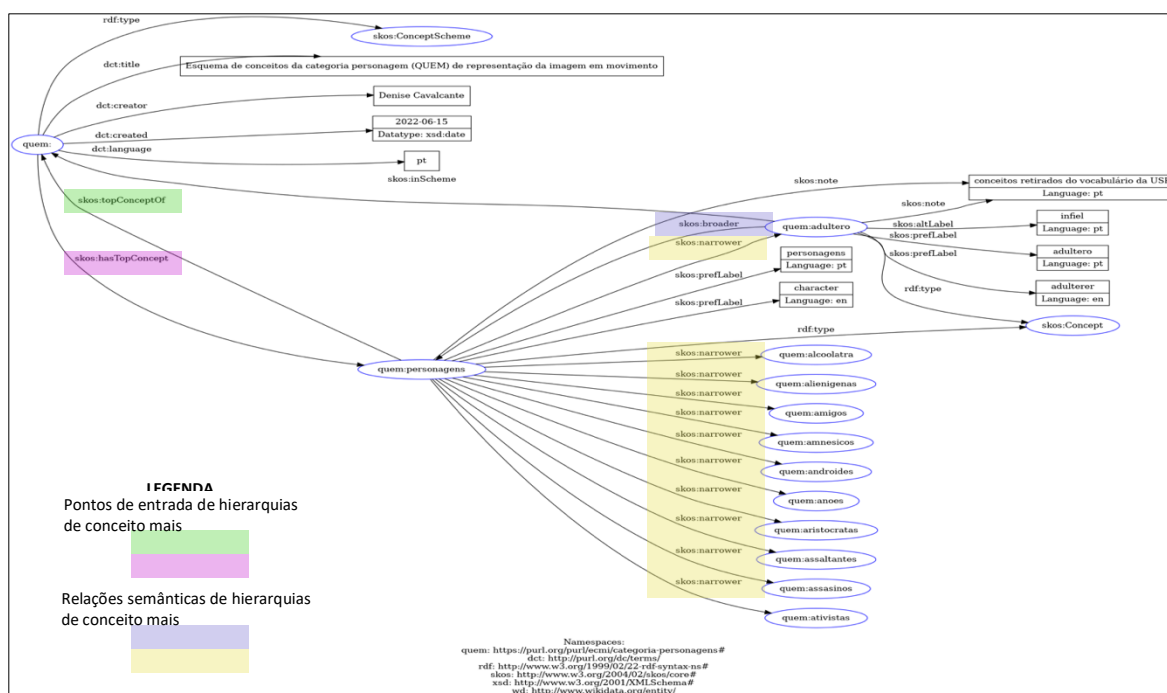
ConceptScheme URI	URI	skos:prefLabel	skos:broader	skos:topConceptOf	skos:inScheme	skos:note
PREFIK	quem	PERSONAGENS				conceito retirado do vocabulário da USP
PREFIK	dict	PERSONAGENS				conceito retirado do vocabulário da USP
PREFIK	skos	PERSONAGENS		categorias:		conceito retirado do vocabulário da USP
PREFIK	sol	PERSONAGENS				conceito retirado do vocabulário da USP
PREFIK	ind	PERSONAGENS				conceito retirado do vocabulário da USP
dict:title	Esquema de conceitos da categoria personagem (QUEM) de representação da imagem em movimento					
dict:description	Criado no âmbito da pesquisa de doutorado do PPGCI, sob supervisão da prof.Dra.Cibele A C M dos Santos					
dict:creator	Denise Cavalcante, Cibele Araújo Camargo Marques dos Santos					
https://url.archive.org/pt/pt/categorias-imagem-movimento/assaltantes	ASSALTANTES	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/aristocratas	ARISTOCRATAS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/atores	PERSONAGENS	PERSONAGENS		categorias:		conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/adulto	ADULTOS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/alcoólatras	ALCOÓLATRAS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/animados	ANIMADOS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/assassinos	ASSASSINOS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/ativistas	ATIVISTAS POLITICOS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/animados	ANIMADOS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/animados	ANIMADOS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/animados	ANIMADOS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/animados	ANIMADOS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/animados	ANIMADOS	PERSONAGENS				conceito retirado do vocabulário da USP
https://url.archive.org/pt/pt/categorias-imagem-movimento/animados	ANIMADOS	PERSONAGENS				conceito retirado do vocabulário da USP

Fonte: Elaborado pela autora.

A análise da serialização do SKOS em um serviço⁷³ web de visualização de grafos, descreve a estrutura do esquema de conceito “quem:”, com as entradas dos conceitos mais amplos e mais específicos através das propriedades `skos:hasTopConcept` e `skos:topConceptOf`. Além de visualizar as relações semânticas hierárquicas `skos:broader` e `skos:narrower` do conceito de topo (*top concept*) “quem:personagens” (Fig.60).

⁷³ Serviço web de visualização de grafos <https://www.ldf.fi/service/rdf-grapher>

Figura 59 - Visualização do grafo do esquema de conceito “quem:”



Renderizado no grapher. Fonte: Elaborado pela autora.

6.2 ISO25964 E SKOS/SKOS-XL

O segundo resultado foi o **SOC 2**, que organizou um micro tesauro intitulado “Esquema de conceitos da categoria locais (ONDE) de representação da imagem em movimento” (fig.42), também serializado em Turtle, estruturando relações hierárquicas, (DeGenérico) e (DeEspecífico), para indicar no esquema um conjunto de conceitos de topo (*top concept*), representando lugares geográficos, arquitetônicos e cosmográficos. Dessa forma, aplicou-se a propriedade `skos:hasTopConcept` para os conceitos que definem os **tipos** de lugares (DeGenérico) e as propriedades `skos:broader/skos:narrower` para as relações semânticas, sendo os conceitos mais estritos os **nomes** dos lugares.

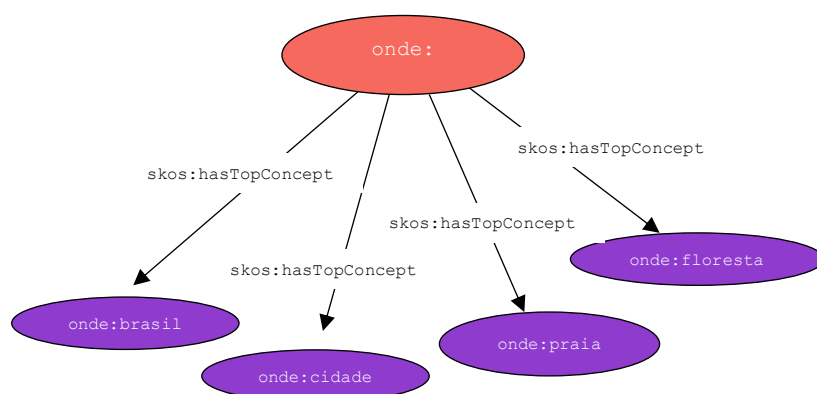
Para isso, foi estabelecido durante a modelagem do esquema de conceitos SKOS, a correspondência entre as principais funções e recursos do modelo de dados da ISO25964 e seu respectivo mapeamento ou extensão para o modelo SKOS/SKOS-XL. A pesquisa explorou diferentes situações em que é necessário estruturar e publicar um mapeamento entre SOC na web.

Apesar de ambos os modelos de dados não serem idênticos, com escopos diferentes, é possível realizar alinhamentos apropriados e seguros, permitindo estruturar um conjunto de

correspondências entre os componentes dos modelos de dados. Os respectivos modelos de dados não são idênticos, porque a ISO 25964 é aplicável a todos os tipos de tesouros (seja para uso na web ou para outras aplicações), enquanto o SKOS é aplicável a todos os tipos de SOC, incluindo os esquemas de classificação e aqueles que não correspondem a norma ISO25964 (Niso, 2011).

A pesquisa definiu como situações exploratórias a construção de um microtesauro modelado em SKOS utilizando a classe *skos:ConceptScheme* e o mapeamento com outros esquemas de conceito. Assim, foi estruturado um esquema para a categoria “Onde” sendo conectados os conceitos usando a propriedade *skos:hasTopConcept*, ou seja, os conceitos superordenados (genéricos e mais amplos) do esquema (Fig.60).

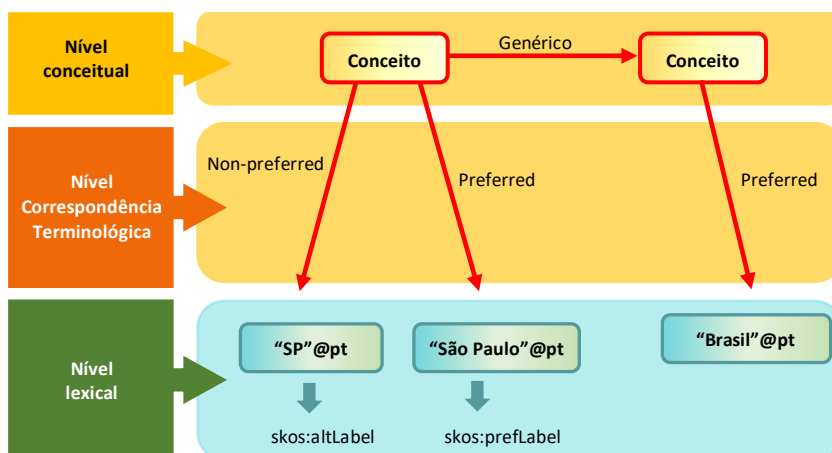
Figura 60 - Propriedade *skos:hasTopConcept*



Fonte: Elaborado pela autora.

Pelo fato de o **SOC 2** ser modelado em SKOS, também apresenta uma estrutura de três níveis: nível conceitual, nível de correspondência terminológica e nível lexical, como mostra a figura:

Figura 61 - Tesouro conceitual SOC 2



Fonte: Adaptado de Pastor-Sanchez, Mendez e Muñoz (2009).

Dessa forma, a correspondência entre a ISO25964 e SKOS/SKOS-XL indica que o recurso **Thesaurus ISO25964**, que tem papel de classe na norma, corresponde a classe `skos:ConceptScheme` no modelo SKOS. A pesquisa estruturou um esquema de conceitos (`skos:ConceptScheme`) modelado em SKOS, chamado "Esquema de conceitos da categoria locais (ONDE) de representação da imagem em movimento". O atributo obrigatório de idioma **lang ISO25964**, foi mapeado para propriedades do Dublin Core `dc:language` ou `dct:language`.

Figura 62 - SOC 2: Esquema de conceitos da categoria locais (ONDE) de representação da imagem em movimento

```

1 @prefix onde: <https://purl.org/purl/ecmi/categorias/> .
2 @prefix dct: <http://purl.org/dc/terms/> .
3 @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
4 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
5
6 onde: a skos:ConceptScheme ;
7   dct:title "Esquema de conceitos da categoria locais (ONDE) de representação da imagem em movimento" ;
8   dct:creator "Denise Cavalcante" ;
9   dct:created "2022-06-15"^^xsd:date ;
10  dct:language "pt" ;
11  skos:hasTopConcept onde:praia, onde:brasil, onde:floresta, onde:cidade .
12

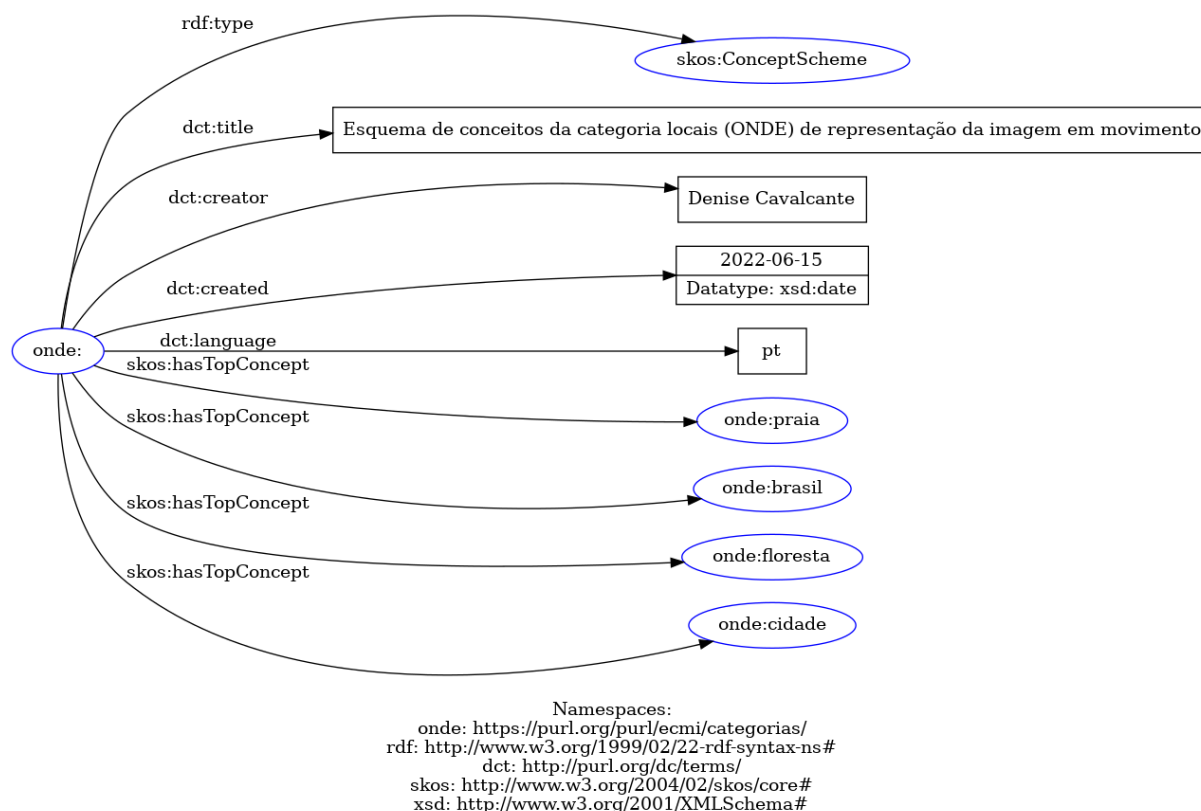
```

SOC 2

Fonte: Elaborado pela autora.

A visualização do esquema de conceito “onde:”, SOC 2, através de um grafo demonstra as entradas dos conceitos mais amplos (Top Concept):

Figura 63 - A visualização do esquema de conceito “onde:”, SOC 2, através de um grafo



Renderizado no grapher. Fonte: Elaborado pela autora.

O recurso do modelo de dados **ThesaurusConcept ISO 25964** pode ser mapeado para a classe `skos:concept`. O recurso “**hasPreferredLabel**” **ISO 25964** que é um tipo de relacionamento associativo, pode ser mapeado para `skos:prefLabel`, como mostra a figura 64 seguir:

Figura 64 - Conceito SKOS onde:brasil

```
onde:brasil a skos:Concept ;
  skos:prefLabel "Brasil"@pt ;
  skos:altLabel "BR"@pt ;
  skos:narrower onde:saopaulo ;
  skos:note "Conceito mais amplo do gênero país Brasil"@pt ;
  skos:topConceptOf onde: .
```

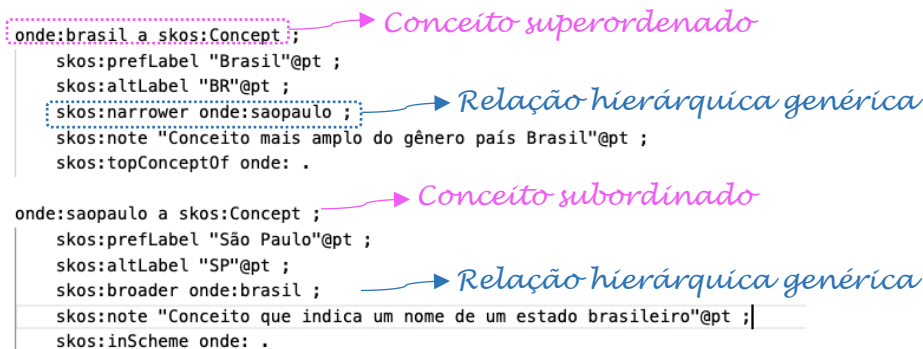
Fonte: Elaborado pela autora.

Os relacionamentos entre conceitos nos tesauros convencionais ocorrem através de associações, o recurso "**isPartOf**" **ISO 25964** pode ser aplicado a qualquer relação que tenha como alvo o **Thesaurus ISO 25964**. O recurso **isPartOf**" **ISO 25964** pode ser mapeado para a propriedade `skos:inScheme`, sendo os assuntos da declaração os seguintes recursos:

ThesaurusConcept, **ConceptGroup** e **ThesaurusArray**. Desse modo, como mostra a figura X, os conceitos `onde:brasil` e `onde:saoPaulo` fazem parte do esquema de conceito `onde: .` Além disto, o recurso **"isPartOf" ISO 25964** também foi mapeado para a propriedade `skos:topConceptOf`.

Ademais o recurso **HierarchicalRelationship ISO 25964** que tem papel de classe, pode ser mapeado para `skos:broader` e `skos:narrower`. No exemplo (fig.65) abaixo o conceito `onde:saoPaulo` possui como relação semântica hierárquica a propriedade `skos:broader` para indicar que o conceito `onde:brasil` é o mais amplo, além de ser um conceito de topo do SOC.

Figura 65 - Relações semânticas SKOS



Fonte: Elaborado pela autora.

No esquema de conceito acima uma propriedade de nota é usada para fornecer informações relacionadas aos conceitos do SKOS, como exemplo, a nota diz que o “Conceito que indica um nome de um estado brasileiro”.

A estruturação das relações entre conceitos são fundamentais na definição de conceitos, de tal modo, os conceitos precisam ser definidos usando documentação legível por humanos, como as notas de escopo ou definições. Na **ISO 25964**, alguns tipos de nota estão associados a conceitos, outras a termos. No SKOS, todas as notas de documentação estão associadas a conceitos.

O modelo SKOS fornece a propriedade `skos:note` para fins de documentação geral, que podem ser mapeadas para as diretrizes existentes para tesouros **ISO 25964**. Ademais, pode ser tratada de forma mais especializada usando as propriedades: `skos:scopeNote`, `skos:definition`, `skos:examplee` e `skos:historyNote`, para os tipos mais específicos de documentação.

A propriedade `skos:scopeNote` (Quadro 45) fornece informações sobre o significado pretendido de um conceito, sobretudo é uma indicação de como o uso de um conceito é limitado na prática de indexação.

Quadro 40 - Propriedade `skos:scopeNote`

```
onde:cidade skos:scopeNote
  "Conceito que representa um tipo de lugar
  geográfico, cosmográfico ou arquitetônico"@pt .
```

Fonte: Elaborado pela autora.

A propriedade `skos:definition` (Quadro 46) fornece uma explicação completa do significado pretendido de um conceito.

Quadro 41 - Propriedade `skos:definition`

```
onde:saoPaulo skos:definition
  "Conceito que representa um nome de lugar
  geográfico, cosmográfico ou arquitetônico,
  usado para indicar relacionamento subordinado
  ou mais estrito"@pt .
```

Fonte: Elaborado pela autora.

```
onde:saoPaulo skos:changeNote [
  rdf:value "Atualizado o termo de `onde:estadoSãoPaulo para
  `onde:saoPaulo"@pt;
  dct:creator ex:DeniseCavalcante;
  dct:date "2022-10-23"
].
ex:DeniseCavalcante rdf:type foaf:Person; foaf:name "Denise
Cavalcante".
```

As propriedades de documentação do SKOS podem ser simplesmente usadas com literais simples RDF, o que possibilita o uso de tags de idioma para rotular propriedades. A documentação pode, portanto, ser fornecida em vários idiomas:

```
ex:brasil rdf:type skos:Concept ;
  skos:prefLabel "Brasil"@pt ;
  skos:prefLabel "Brazil"@en ;
  skos:definition " Conceito que representa um nome de lugar
  geográfico, cosmográfico ou arquitetônico, usado para indicar
  relacionamento mais amplo"@pt ;
  skos:definition "Concept representing a name of geographic,
  cosmographic, or architectural place, used to indicate a broader
  relationship"@en .
```

Quadro 42 - Excel do esquema

A1 fx ConceptScheme URI					
A	B	C	D	E	F
1	ConceptScheme URI	<https://purl.org/purl/ecmi/categorias/>			
2	dct:title	Esquema de conceitos da categoria locais (ONDE) de representação da imagem em movimento			
3	dct:description	Esquema de conceito da tese			
4	dc:creator	Denise Cavalcante			
5					
6	Identificador de conceito (computado automaticamente)	Rótulo do conceito	Rótulo do conceito pai (computado automaticamente)	Rótulo do conceito filho (computado automaticamente)	TopConceptOf inScheme
7	URI	skos:prefLabel	skos:broader	skos:narrower	
8	onde:saopaulo	São Paulo	Brasil		onde:
9	onde:copacabana	Copacabana	Praia		onde:
10	onde:amazonia	Amazonia	Floresta		onde:
11	onde:paris	Paris	Cidade		onde:
12	onde:brasil	Brasil		São Paulo	onde:
13	onde:praia	Praia		Copacabana	onde:
14	onde:floresta	Floresta		Amazonia	onde:
15	onde:cidade	Cidade		Paris	onde:
16					

Fonte: Elaborado pela autora.

6.3 MAPEAMENTO HIERÁRQUICO E ASSOCIATIVO

Uma propriedade fundamental do SKOS é a interoperabilidade entre vocabulários de contextos diferentes, pois com o SKOS é possível afirmar que dois conceitos de esquemas diferentes, possuem significados que podem ser mapeados de forma hierárquica e associativa, mesmo provenientes de contextos diferentes. De tal maneira, foram estabelecidos links de alinhamento (mapeamento) entre conceitos SKOS de dois esquemas de conceito diferentes. Para isso foi criado o **SOC 3** “Esquema de conceitos das categorias de representação da imagem em movimento QUEM/QUANDO/ONDE/OQUE” (fig.66) para estabelecer o alinhamento com o **SOC 2** “Esquema de conceitos da categoria locais (ONDE) de representação da imagem em movimento”. Logo, foram aplicadas as propriedades de mapeamento SKOS (`skos:broadMatch`, `skos:narrowMatch`, `skos:relatedMatch`).

O uso de URI para representar os conceitos, possibilita referenciar inequivocamente cada conceito SKOS em qualquer aplicação SKOS, criando relações semânticas entre conceitos a partir de mapeamentos que utilizam mais de um SOC ao mesmo tempo, e que possuem domínios sobrepostos.

Figura 66 - SOC 3: Esquema de conceitos das categorias de representação da imagem em movimento QUEM /QUANDO/ONDE/OQUE

```

propriedade_mapeamento_skos.ttl
1
2 @prefix onde: <https://purl.org/purl/ecml/categorias/> .
3 @prefix categorias: <https://purl.archive.org/purl/categorias-imagem-movimento#> .
4 @prefix dct: <http://purl.org/dc/terms/> .
5 @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
6 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
7 @prefix wd: <http://www.wikidata.org/entity/> .
8
9 categorias: a skos:ConceptScheme ;
10   dc:title "Esquema de conceitos das categorias de representação da imagem em movimento QUEM, QUANDO, ONDE, OQUE" ;
11   dc:creator "Denise Cavalcante" ;
12   dct:created "2022-06-15"^^xsd:date ;
13   skos:hasTopConcept categorias:personagens, categorias:lugar, categorias:evento .
14
15 categorias:personagens a skos:Concept ;
16   skos:prefLabel "personagens"@pt ;
17   skos:narrower categorias:adultero, categorias:alcoólatra ;
18   skos:note "conceitos retirados do vocabulário da USP"@pt ;
19   skos:topConceptOf categorias: .
20
21 categorias:adultero a skos:Concept ;
22   skos:prefLabel "adultero"@pt ;
23   skos:broader categorias:personagens ;
24   skos:note "conceitos retirados do vocabulário da USP"@pt ;
25   skos:inScheme categorias: .
26
27 categorias:alcoólatra a skos:Concept ;
28   skos:prefLabel "alcoólatra"@pt ;
29   skos:broader categorias:personagens ;
30   skos:note "conceitos retirados do vocabulário da USP"@pt ;
31   skos:inScheme categorias: .
32
33 categorias:pontedasbandeiras a skos:Concept ;
34   skos:prefLabel "Ponte das bandeiras"@pt ;
35   skos:broader categorias:ponte ;
36   skos:note "conceitos retirados do vocabulário da USP"@pt ;
37   skos:inScheme categorias: .
38
39 categorias:ponte a skos:Concept ;
40   skos:prefLabel "Ponte"@pt ;
41   skos:narrower categorias:pontedasbandeiras ;
42   skos:inScheme categorias: .
43
44 categorias:evento a skos:Concept ;
45   skos:prefLabel "evento"@pt ;
46   skos:narrower categorias:mortedarainha, categorias:morte ;
47   skos:note "conceitos retirados do vocabulário da USP"@pt ;
48   skos:topConceptOf categorias: .
49

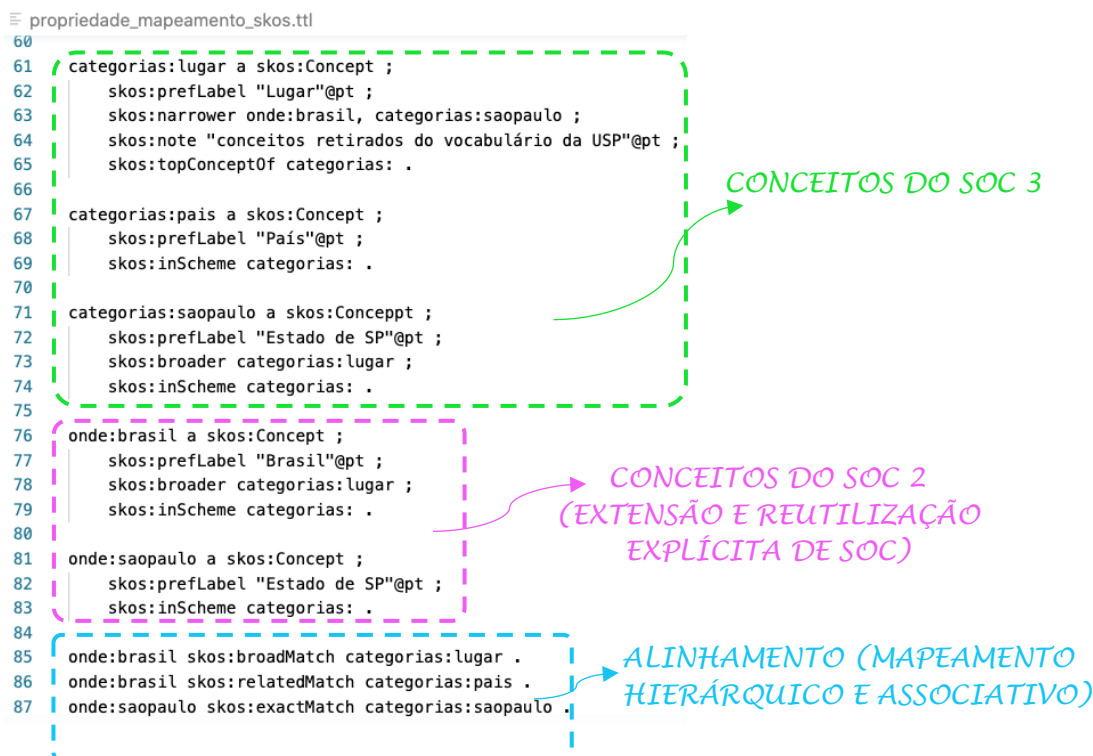
```



Fonte: Elaborado pela autora.

Nesse caso, os alinhamentos (mapeamentos) são inerentes ao significado dos conceitos vinculados, pois as propriedades `skos:broadMatch` e `skos:narrowMatch` são usadas para estabelecer um link de mapeamento hierárquico entre dois conceitos (fig.67). A propriedade `skos:relatedMatch` é usada para indicar um link de mapeamento associativo entre dois conceitos. A figura 65 mostra a extensão do SOC 3 e os alinhamentos reutilizando os conceitos de SOC 2.

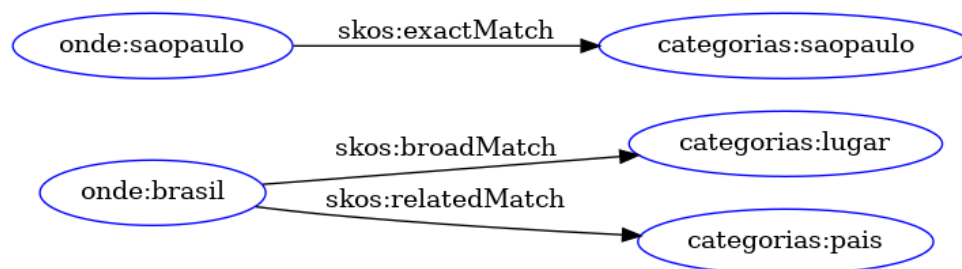
Figura 67 - Alinhamento entre conceitos dos esquemas SOC 2 e SOC 3



Fonte: Elaborado pela autora.

Uma característica fundamental do mapeamento é a possibilidade de afirmar que dois conceitos de esquemas diferentes têm significados comparáveis e especificar como esses significados se comparam, inclusive quando ambos os SOC provêm de contextos diferentes e seguem princípios de modelagem diferentes. Os mapeamentos conceituais são as principais prerrogativas de disponibilizar SOC na Web Semântica usando SKOS.

Figura 68 - Grafo do alinhamento (Mapeamento hierárquico e associativo)



Namespaces:
 onde: <https://purl.org/purl/ecmi/categorias/>
 categorias: <https://purl.archive.org/purl/categorias-imagem-movimento#>
 dct: <http://purl.org/dc/terms/>
 skos: <http://www.w3.org/2004/02/skos/core#>
 xsd: <http://www.w3.org/2001/XMLSchema#>
 wd: <http://www.wikidata.org/entity/>

Renderizado no grapher. Fonte: Elaborado pela autora.

6.4 REDE SEMÂNTICA

Nos resultados foi construído uma rede semântica entre esquemas para as categorias gerais de representação da imagem em movimento e para a categoria (onde:) que representa os lugares. O **SOC 3** “Esquema de conceitos das categorias de representação da imagem em movimento (**QUEM /QUANDO/ONDE/OQUE**)” foi estendido e vinculado, reusando os conceitos de outro esquema **SOC 4** intitulado “Esquema de conceitos dos países da América latina” na figura 69.

Figura 69 - SOC 4: Esquema de conceitos dos países da américa latina

```

categorias_lugares_latinos.ttl
1  @prefix paises: <https://purl.org/purl/ecmi/categorias2/> .
2  @prefix dct: <http://purl.org/dc/terms/> .
3  @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
4  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
5
6  paises: a skos:ConceptScheme ;
7  |   dct:title "Esquema de conceitos dos países da américa latina" ;
8  |   dct:creator "Denise Cavalcante" ;
9  |   dct:created "2022-06-15"^^xsd:date ;
10 |   dct:language "pt" ;
11 |   skos:hasTopConcept paises:brasil, paises:chile, paises:uruguai, paises:argentina.
12
13 paises:brasil a skos:Concept ;
14 |   skos:prefLabel "Brasil"@pt ;
15 |   skos:topConceptOf paises: .
16
17 paises:chile a skos:Concept ;
18 |   skos:prefLabel "Chile"@pt ;
19 |   skos:topConceptOf paises: .
20
21 paises:uruguai a skos:Concept ;
22 |   skos:prefLabel "Uruguai"@pt ;
23 |   skos:topConceptOf paises: .
24
25 paises:argentina a skos:Concept ;
26 |   skos:prefLabel "Argentina"@pt ;
27 |   skos:topConceptOf paises: .
28

```



Fonte: Elaborado pela autora.

Dessa forma, aplicou-se a propriedade `skos:inScheme` para reutilizar os conceitos referentes ao esquema de países da América latina, no caso o conceito `paises:brasil`, no segundo SOC de categorias gerais da imagem em movimento como mostra a figura 70.

Figura 70 - Rede semântica entre SOC 3 e SOC 4 (reutilização do conceito `paises:brasil`)

```

74  |paises:brasil skos:inScheme categorias: .|
75  |-----|
76  categorias:saopaulo a skos:Concept ;
77  |   skos:prefLabel "São Paulo"@pt ;
78  |   |skos:broader paises:brasil ;|
79  |   skos:note "Conceito que indica um nome de um estado brasileiro"@pt;
80  |   skos:inScheme categorias: .
81
82  categorias:bahia a skos:Concept ;
83  |   skos:prefLabel "Bahia"@pt ;
84  |   |skos:broader paises:brasil ;|
85  |   skos:note "Conceito que indica um nome de um estado brasileiro"@pt;
86  |   skos:inScheme categorias: .
--

```

Conceitos do SOC 4 (extensão e reutilização explícita de SOC)

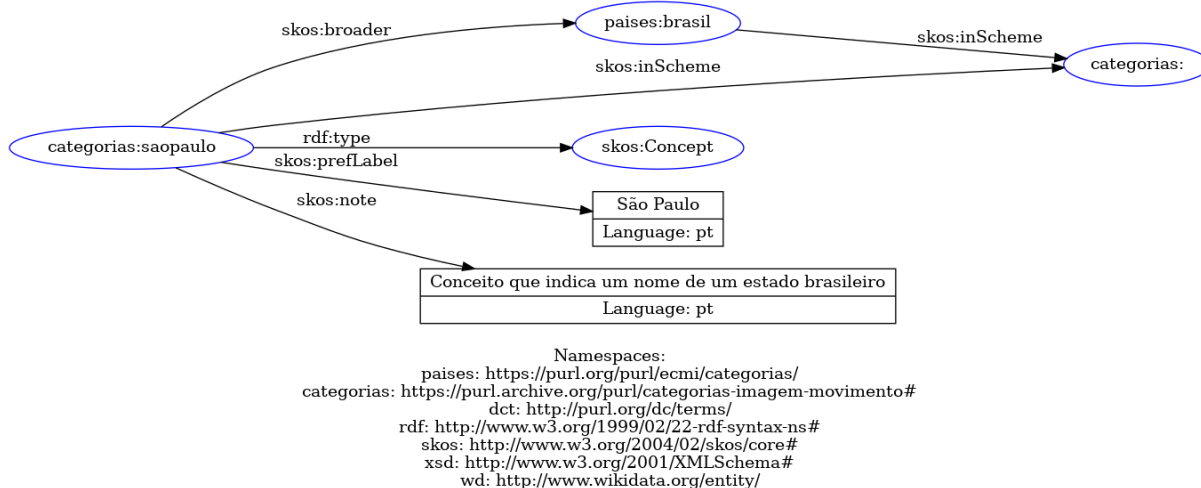
Reutilização de conceito (SOC 4) no esquema SOC 3

Reutilização de conceito (SOC 4) no esquema SOC 3

Fonte: Elaborado pela autora.

O conceito do SOC foi renderizado no serviço de grafos, como mostra a figura 71:

Figura 71 - Conceito categorias:saopaulo renderizado no grafo



Fonte: Elaborado pela autora.

Também foi experimentado renderizar o SOC a partir de planilhas, como na figura 72:

Figura 72 - SOC na forma de planilha

A	B	C	D	E	F
ConceptScheme URI	<https://purl.org/purl/ecmi/categorias/>				
dct:title	Esquema de conceitos das categorias de representação da imagem em movimento QUEM, QUANDO, ONDE, OQUE				
dct:description	Esquema de conceito da tese				
dc:creator	Denise Cavalcante				
Identificador de conceito (computado automaticamente)	Rotulo do conceito	Rótulo do conceito pai (computado automaticamente)	Rótulo do conceito filho (computado automaticamente)	TopConceptOf	inScheme
URI	skos:prefLabel	skos:broader	skos:narrower		
categorias:personagens	Personagens		Adultero, Alcoolátra	Categorias:	
categorias:evento	Evento		Morte da rainha, Morte	Categorias:	
categorias:lugar	Lugar		Estado de SP	Categorias:	
categorias:ponte	Ponte		Ponte das Bandeiras		Categorias:
pais:brasil	Brasil		São Paulo, Bahia	pais:	Categorias:
categorias:adultero	Adultero	Personagens			Categorias:
categorias:alcoolatra	Alcoolátra	Personagens			Categorias:
categorias:pontedasbandeiras	Ponte das Bandeiras	Ponte			Categorias:
categorias:mortedarainha	Morte da rainha	evento			Categorias:
categorias:morte	Morte	evento			Categorias:
categorias:estadodeSP	Estado de São Paulo	Lugar			Categorias:
categorias:saopaulo	São Paulo	paises			Categorias:
categorias:bahia	Bahia	paises			Categorias:

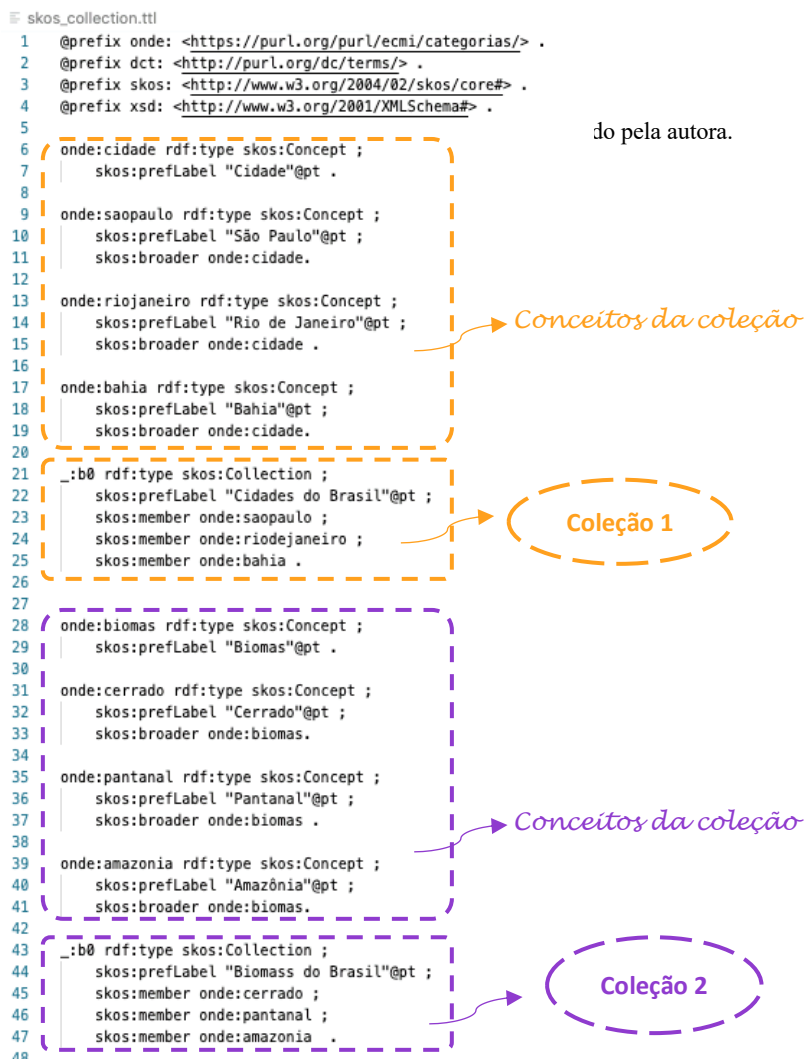
Fonte: Elaborado pela autora.

Depois da elaboração de possível rede semântica, foram encerrados os experimentos com tesouros conceituais, em seguida foi modelado coleções SKOS, os quais são tipos de SOC com visualização independente do modelo de dados SKOS.

6.5 COLEÇÃO DE CONCEITOS

Por último, como a documentação esclarece que a classe `skos:Collection` é disjunta da `skos:Concept` e a representação e visualização sistemática de uma coleção depende de procedimentos externos ao escopo do modelo de dados SKOS, foi desenvolvida uma coleção de conceitos que não estabelece nenhum relacionamento com os alinhamentos anteriormente apresentados nos resultados do trabalho. Por tanto, foi criada uma coleção de conceitos da categoria ONDE usando a classe `skos:Collection`, como apresentado na figura 73. Na **ISO 25964** os objetos `ConceptGroup` e `ThesaurusArray` são semelhantes às classes `skos:Collection`.

Figura 73 - Coleção de conceitos da categoria ONDE

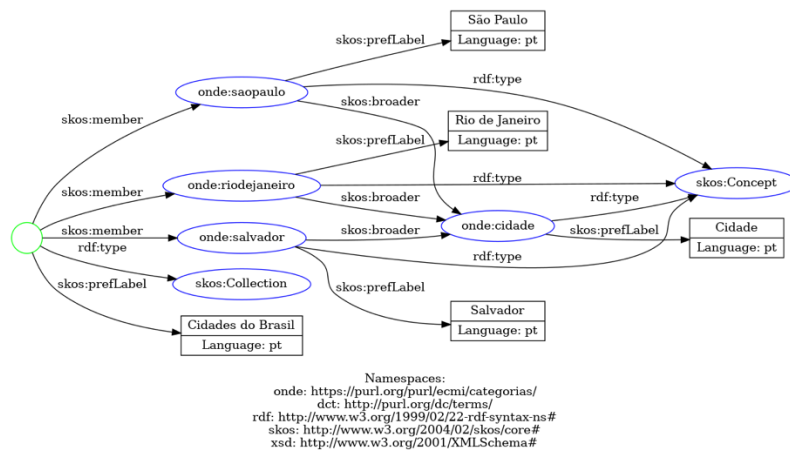


Fonte: Elaborado pela autora.

No primeiro bloco da serialização RDF em Turtle foi criada uma coleção com rótulo lexical preferido “Cidades do Brasil” e definidos os seguintes conceitos SKOS onde:saopaulo, onde:riodejaneiro e onde:bahia como membros da coleção.

E ainda, no segundo bloco da serialização foi criada uma coleção com rótulo lexical preferido “Biomass do Brasil” e definidos os conceitos SKOS onde:cerrado, onde:pantanal e onde:amazonia como membros da coleção.

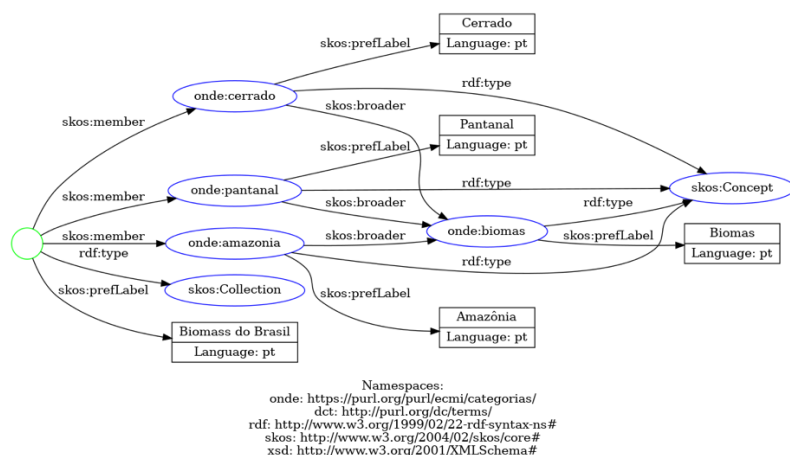
Figura 74 - Renderização da coleção 1 no serviço online de grafo



Fonte: Elaborado pela autora.

A figura 74 e 75 mostra a renderização da coleção 1 e 2 no serviço online de grafos:

Figura 75 - Renderização da coleção 2 no serviço online de grafo



Fonte: Elaborado pela autora.

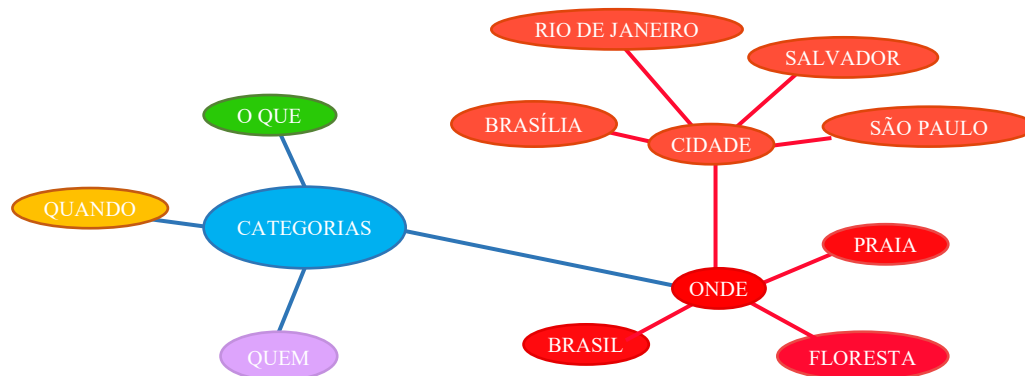
Portanto, a principal diferença entre skos:Collection e skos:conceptScheme, está no escopo e na finalidade: a skos:Collection agrupa

conceitos relacionados num esquema, enquanto o `skos:conceptScheme` representa o esquema completo de vocabulário ou taxonomia, incluindo todos os conceitos e suas propriedades.

6.6 MODELAGEM SEMÂNTICA DA ONTOLOGIA CRIM

A segunda etapa da pesquisa aplicada realizou a modelagem semântica de uma ontologia de domínio, descrevendo a conceituação relacionada ao **domínio genérico da indexação de imagem em movimento**, ou seja, a abordagem é epistemológica e trata os processos de indexação. O trabalho conceitualizou o universo que abrange a prática de designar conceitos por meio de termos para a recuperação de documentos, a partir da representação formal e o conhecimento explícito do domínio. A abordagem no desenvolvimento da hierarquia de classes é o processo *top-down*, com conceitos de nível superior (*top-level*). A classificação das categorias é organizada em **facetar** e os relacionamentos dos conceitos em hierarquias genéricas.

Figura 76 - Representação da estrutura de facetar da ontologia CRIM



Fonte: Elaborado pela autora.

De tal modo, foi desenvolvido a ontologia “**Categorias de Representação da Imagem em Movimento (CRIM)**”, fundamentada numa conceituação analítica e metodológica sobre a representação temática (indexação) e as práticas de estruturação de conceitos presente nos SOC para imagem em movimento. Dessa forma, a ontologia possibilita os processos de anotação semântica de corpora audiovisual, a definição de conceitos e seus relacionamentos através de expressividade semântica, por fim, a estruturação de declarações formais acerca do universo da organização e representação do conhecimento da imagem em movimento. O público-alvo da modelagem semântica são pesquisadores, docentes, discentes que utilizam bases de dados com

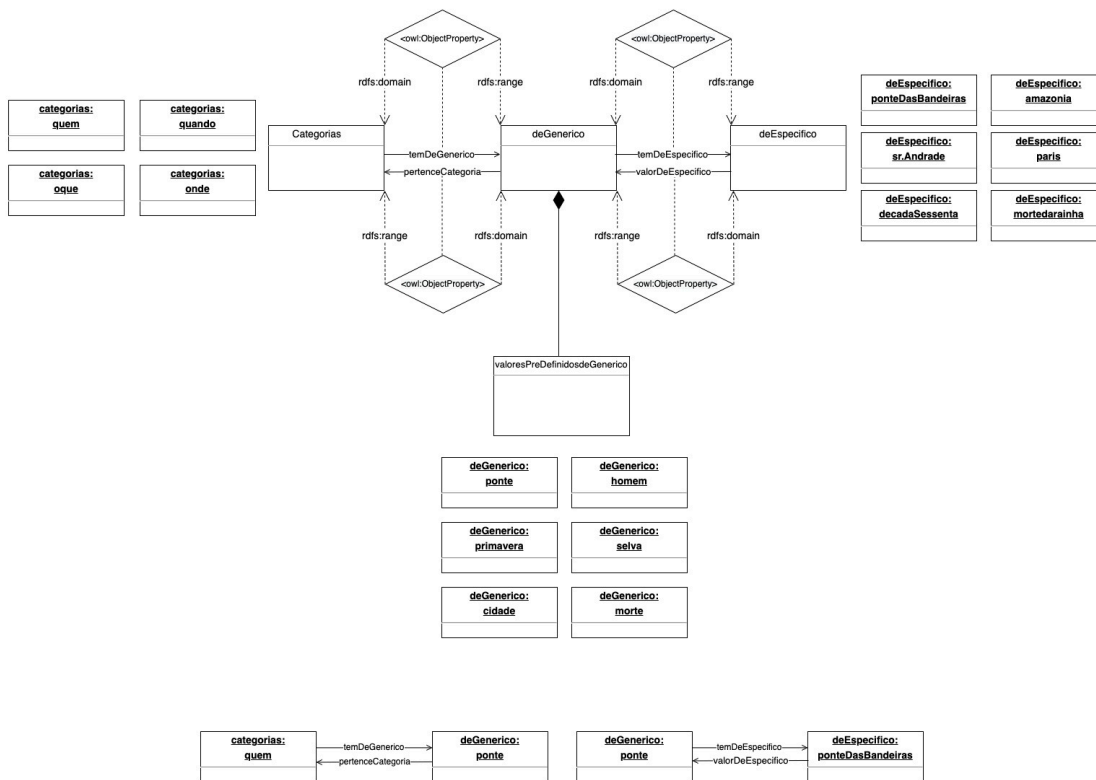
documentos de imagens em movimento. Visando auxiliar processos de indexação de imagem em movimento no ambiente web, integrando os processos de anotação semântica.

As classes e propriedade da ontologia foram criadas com base no referencial teórico-metodológico e analítico dos métodos condensados de Smit (1996), Shatford (1986), Macambyra (2017) e Panofysky (1989) elaborados a partir das categorias de representação (ONDE, QUEM, QUANDO e OQUE) e do níveis de determinação das perguntas: “a imagem é de quê?” e “A imagem é sobre o quê?”, para representar os processos de representação temática da imagem em movimento, ou seja, as etapas de tradução do item informacional em conceitos procedentes da ontologia. De tal modo, a modelagem semântica abordou três perspectivas teórico-metodológicas, a representação temática da imagem, os SOC e o desenvolvimento de ontologias de domínio.

A ontologia Categorias de Representação da Imagem em Movimento (CRIM) está estruturada nos três tipos diferentes de conceitos que formam a representação da imagem em movimento delimitados na pesquisa, 1º) categorias de representação (ONDE, QUEM, QUANDO e OQUE), nível de determinação De Genérico e no nível de determinação De Específico.

Dessa forma, o modelo semântico primeiro foi modelado num diagrama de classe UML adaptados para estrutura ontológica, representando as três classes principais Categorias, De Genérico e De Específico e as propriedades que associam de forma subordinada todas as instâncias de determinada classe e as propriedades inversas que conecta uma única instância a instância da classe superior. Assim como, estão definidas algumas instâncias de classe pertencente a suas respectivas classes na figura 77.

Figura 77 - Diagrama de Classe da ontologia CRIM



Fonte: Elaborado pela autora.

As classes e propriedades definidas em OWL para a ontologia estão no quadro 48:

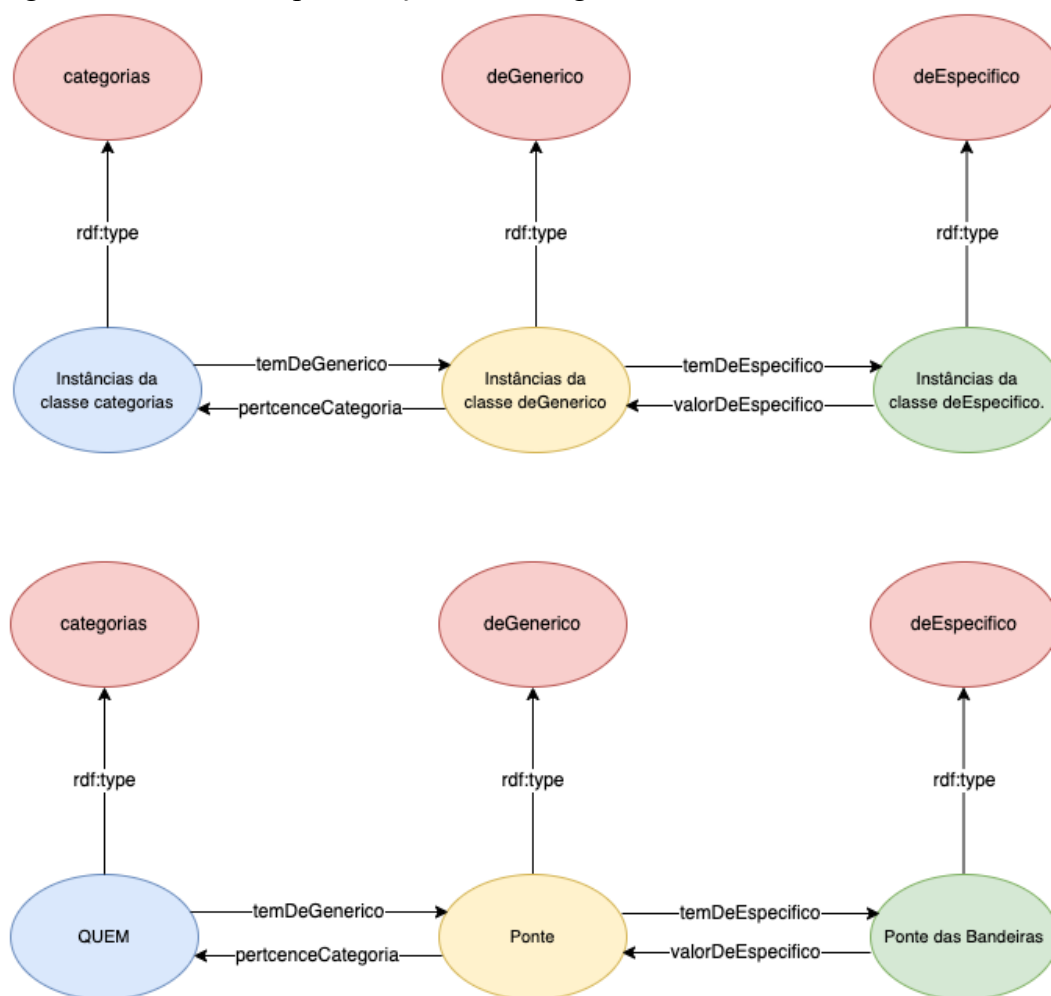
Quadro 43 - Classes e propriedades da ontologia CRIM

URI	Tipo	Descrição	Domínio	Alcance
Categorias	owl:Class	Categorias de representação da imagem em movimento: QUEM, QUANDO, ONDE, O QUE.	-	-
De Genérico	owl:Class	Nível de determinação De Genérico de representação da imagem em movimento, respondendo à pergunta: A IMAGEM É DE QUE?	-	-
De Específico	owl:Class	Nível de determinação De Específico de representação da imagem em movimento, respondendo à pergunta: A IMAGEM É DE QUE?	-	-
temDeGenerico	owl:ObjectProperty	Associar todas as instâncias da classe De Genérico subordinadas a classe Categorias.	Categorias	De Genérico
temDeEspecifico	owl:ObjectProperty	Associar todas as instâncias da classe De Específico subordinadas a classe De Genérico.	De Genérico	De Específico
pertenceCategoria	owl:ObjectProperty	Conectar uma instância da classe De Genérico com uma instância da classe superior Categorias.	De Genérico	Categorias
valorDeEspecifico	owl:ObjectProperty	Conectar uma instância da classe De Específico com uma instância da classe superior De Genérico.	De Específico	De Genérico

Fonte: Elaborado pela autora.

Consequentemente é possível transformar a modelagem conceitual e semântica num grafo RDF, mostrando as classes e os relacionamentos ontológicos, figura 78.

Figura 78 - Grafo de representação da ontologia CRIM



Fonte: Elaborado pela autora.

A partir disso, foi construída uma planilha para organizar os conceitos de acordo o tipo, a classe, os rótulos no idioma português, a descrição de cada conceito e as propriedades de objeto de classe.

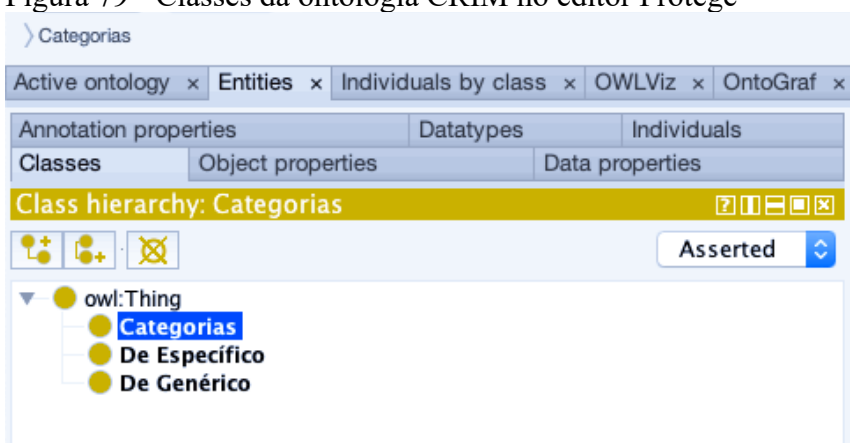
Quadro 44 - Planilha com as entidades da ontologia CRIM

ENTIDADE	TIPO	Pertence a classe	Label @pt	Descrição
Categorias	Classe		Categorias	Categorias de representação da imagem em movimento QUEM, QUANDO, ONDE, O QUE
De Generico	Classe		De Generico	Nível de determinação De Especifico para representação da imagem em movimento respondendo à pergunta A IMAGEM É DE QUE?
De Especifico	Classe		De Especifico	Nível de determinação De Especifico para representação da imagem em movimento respondendo à pergunta A IMAGEM É DE QUE?
QUEM	Conceito (instância)	Categorias	QUEM	Define animado e inanimado, objetos e seres concretos.
QUANDO	Conceito (instância)	Categorias	QUANDO	Define o tempo linear ou cíclico, datas, períodos específicos, tempos recorrentes.
ONDE	Conceito (instância)	Categorias	ONDE	Localização da imagem no "espaço": espaço geográfico ou espaço da imagem.
O QUE	Conceito (instância)	Categorias	O QUE	Define ações, eventos e emoções.
Ponte	Conceito (instância)	De Generico	Ponte	Conceito representando animado ou inanimado, objetos e seres concretos.
Homem	Conceito (instância)	De Generico	Homem	Conceito representando animado ou inanimado, objetos e seres concretos.
Ponte das Bandeiras	Conceito (instância)	De Especifico	Ponte das Bandeiras	Conceito representando especificamente de quem se trata a imagem.
Sr. Andrade	Conceito (instância)	De Especifico	Sr. Andrade	Conceito representando especificamente de quem se trata a imagem.
Primavera	Conceito (instância)	De Generico	Primavera	Conceito representando o tempo cíclico.
1996	Conceito (instância)	De Especifico	1996	Conceito representando o tempo linear.
Selva	Conceito (instância)	De Generico	Selva	Conceito representando um tipo de lugar geográfico, arquitetônico ou cosmográfico.
Cidade	Conceito (instância)	De Generico	Cidade	Conceito representando um tipo de lugar geográfico, arquitetônico ou cosmográfico.
Amazônia	Conceito (instância)	De Especifico	Amazônia	Conceito representando um nome de lugar geográfico, arquitetônico ou cosmográfico.
Bahia	Conceito (instância)	De Especifico	Bahia	Conceito representando um nome de lugar geográfico, arquitetônico ou cosmográfico.
Jogo de Futebol	Conceito (instância)	De Generico	Jogo de Futebol	Conceito que define ações, eventos e emoções.
Morte	Conceito (instância)	De Generico	Morte	Conceito que define ações, eventos e emoções.
Copa do Mundo	Conceito (instância)	De Especifico	Copa do Mundo	Conceito que define eventos individualmente nomeados.
Rainha	Conceito (instância)	De Especifico	Rainha	Conceito que define eventos individualmente nomeados.

Fonte: Elaborado pela autora.

Os resultados foram transformados numa ontologia desenvolvida no editor Protégé, usando a metodologia 101 como base da modelagem conceitual. Assim, foram definidas as classes principais e inseridas as instâncias individuais de cada classe, como definido na figura 79.

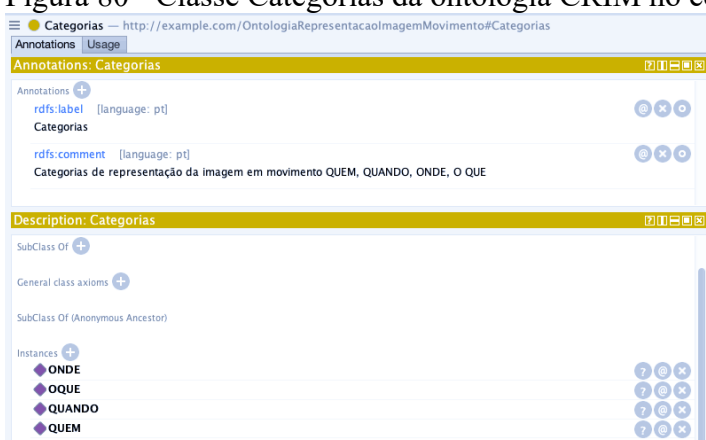
Figura 79 - Classes da ontologia CRIM no editor Protégé



Fonte: Elaborado pela autora.

Em seguida, foram adicionados a classe ex:Categorias as instâncias que definem as categorias de representação da imagem, por tanto, são conceitos que não entram como descritores e não devem ser traduzidos para termos no processo de indexação das imagens, por isso, são considerados como conceitos de nível superior ONDE, QUANDO, O QUE e QUEM, como na figura 80.

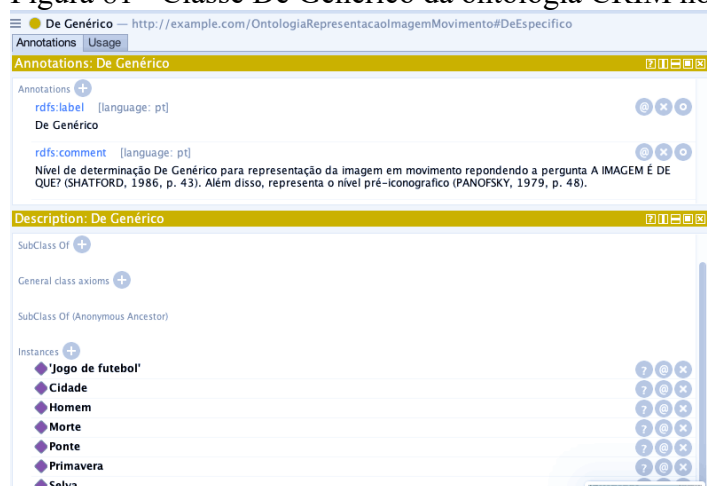
Figura 80 - Classe Categorias da ontologia CRIM no editor Protégé



Fonte: Elaborado pela autora.

A segunda classe definida na ontologia foi ex:DeGenérico e os conceitos genéricos para representar a imagem, que podem ser utilizados no processo de indexação de itens informacionais, como definido na figura 81.

Figura 81 - Classe De Genérico da ontologia CRIM no editor Protégé



Fonte: Elaborado pela autora.

A terceira classe definida foi a ex:DeEspecifico com instâncias individuais que definem entidades nomeadas e são os conceitos específicos, como na figura 82.

Figura 82 - Classe De Específico da ontologia CRIM no editor Protégé

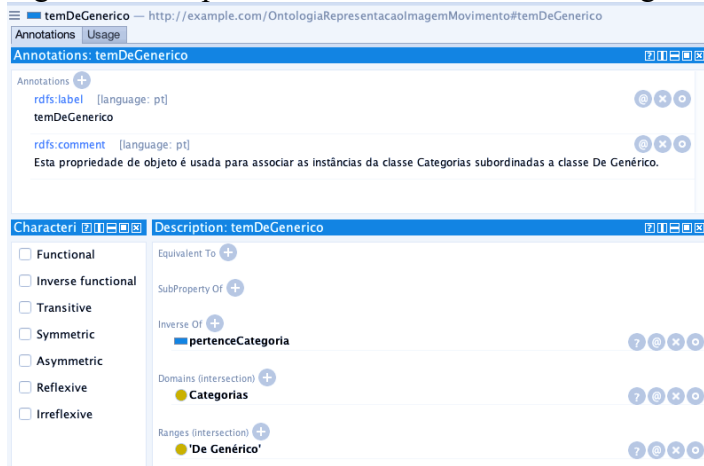


Fonte: Elaborado pela autora.

Em seguida, foram desenvolvidas as propriedades (owl:ObjectProperty) para conectar as instâncias de classe, ou seja, os relacionamentos entre os conceitos. O primeiro

relacionamento criado foi o `ex:temDeGenerico` com domínio a classe `ex:Categorias` e alcance a classe `ex:DeGenérico`, como mostra a figura 83.

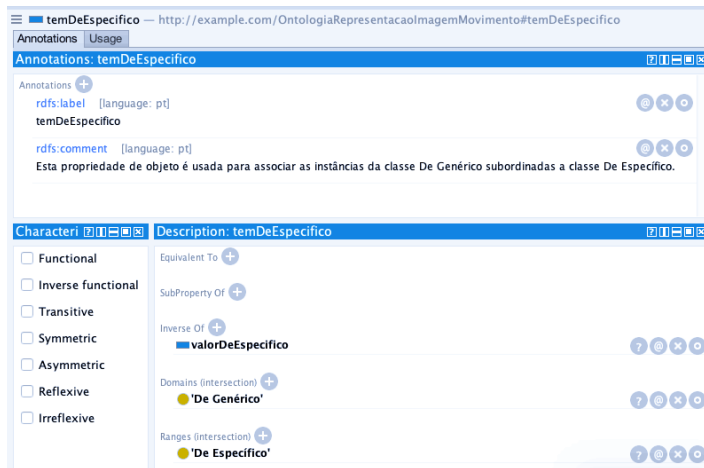
Figura 83 - Propriedade `temDeGenerico` da ontologia CRIM no editor Protégé



Fonte: Elaborado pela autora.

O segundo relacionamento criado foi o `ex:temDeEspecifico` com domínio a classe `ex:DeGenérico` e alcance a classe `ex:DeEspecifico`, como apresentado na figura 84.

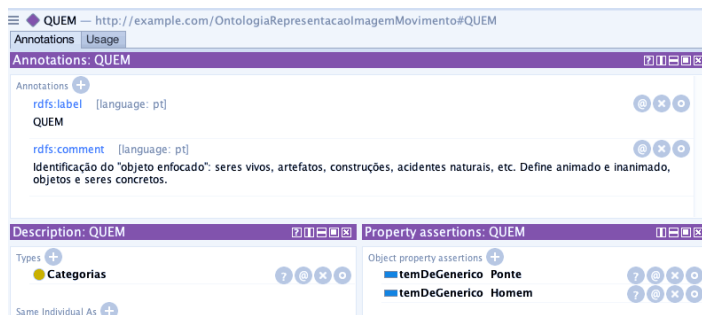
Figura 84 - Propriedade `temDeEspecifico` da ontologia CRIM no editor Protégé



Fonte: Elaborado pela autora.

Por fim, foram adicionados os relacionamentos entre as instâncias de classe através das propriedades OWL. Dessa forma, a instância `ex:QUEM` pertencente a classe `ex:Categorias` recebeu a propriedade `ex:temDeGenerico` para associar a todas as instâncias da classe `ex:DeGenérico`, como definido na figura 85.

Figura 85 - Instância QUEM da classe Categorias da ontologia CRIM no editor Protégé



Fonte: Elaborado pela autora.

Além de que, a instância `ex:Ponte` da classe `ex:DeGenérico` foi associada através da propriedade `ex:temDeEspecifico` com seu conceito mais estrito `ex:PonteDasBandeiras`. Como na figura 86.

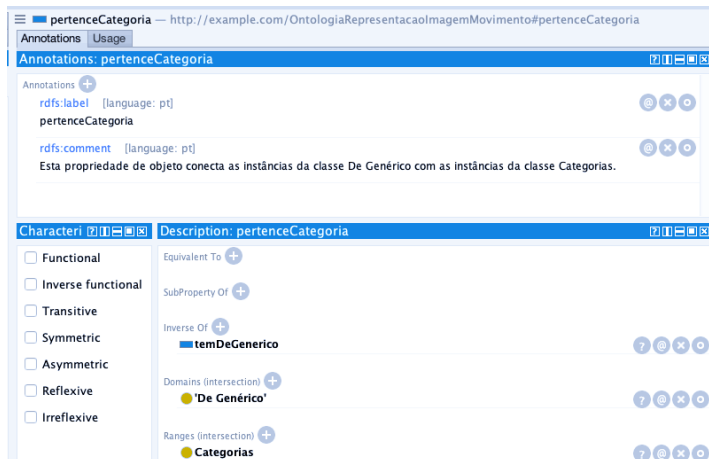
Figura 86 - Instância Ponte da classe De Genérico da ontologia CRIM no editor Protégé



Fonte: Elaborado pela autora.

Além disso, foi criada outra propriedade inversa de `ex:temDeGenerico`, intitulada de `ex:PertenceCategoria`, para indicar que um conceito mais estrito (instância) pertence a determinada categoria de representação superior, como na figura 87.

Figura 87 - Propriedade PertenceCategoria da ontologia CRIM no editor Protégé

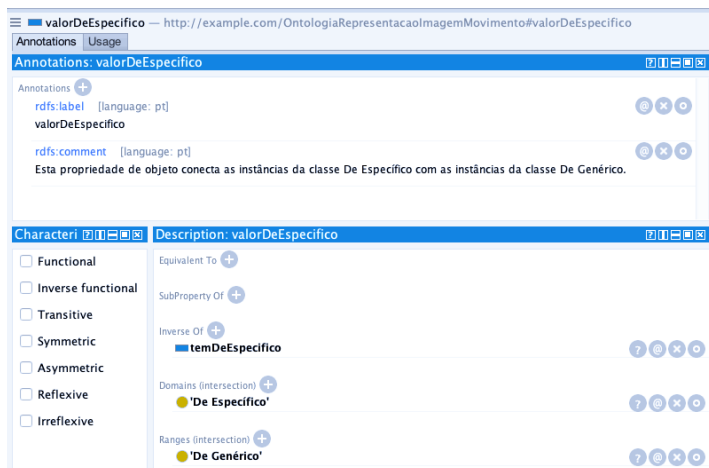


Fonte: Elaborado pela autora.

De modo que, a propriedade ex:PertenceCategoria foi aplicada para criar a relação entre ex:Ponte e a categoria ex:QUEM, como mostra a figura 87.

Ademais, o conceito ex:PonteDasBandeiras é uma entidade nomeada e o valor pré-definido de ex:Ponte, portanto, foi aplicado a propriedade inversa de ex:temDeEspecifico, intitulada valor ex:ValorDeEspecifico (fig.88).

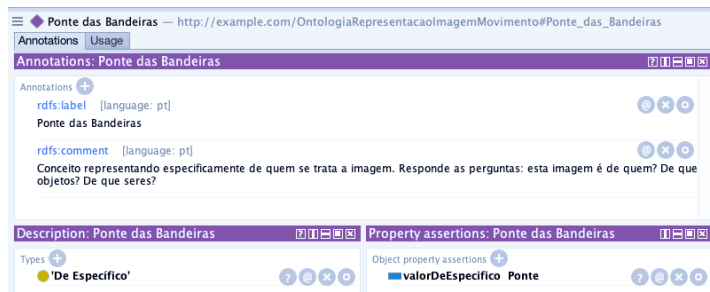
Figura 88 - Propriedade valorDeEspecifico da ontologia CRIM no editor Protégé



Fonte: Elaborado pela autora.

Por conseguinte, a propriedade ex:ValorDeEspecifico foi aplicada para criar a relação entre ex:PonteDasBandeiras e ex:Ponte (fig.89).

Figura 89 - Instância Ponte das Bandeiras da classe De Específico da ontologia CRIM no editor Protégé



Fonte: Elaborado pela autora.

A descrição feita acima da ontologia CRIM demonstrou o processo de criação utilizando uma modelagem direta no editor de ontologia, sem usar nenhuma serialização em RDF/XML ou em Turtle. Entretanto, para entender a modelagem de OWL 2, a partir do processo de serialização RDF, foi modelado parte da ontologia CRIM em Turtle, convertido para RDF/XML e feita a renderização do arquivo no editor Protégé. O que possibilitou experimentar, na próxima subseção, criar modelagens usando conceitos SKOS e classes OWL juntos, que foram combinados para modelar diferentes partes da mesma conceituação e em seguida foi realizado a renderização no editor Protégé

Além do Protégé, também foi realizado a modelagem da ontologia CRIM na serialização Turtle, como a figura 90 abaixo:

Figura 90 - Ontologia CRIM modelada em Turtle

```

@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix crim: <http://example.com/OntologiaRepresentacaoImagemMovimento#> .

crim: a owl:Ontology ;
dcterms:description "Ontologia categorias de representação da imagem em movimento"@pt ;
dcterms:license <http://creativecommons.org/licenses/by/4.0/> ;
dcterms:creator "Denise Cavalcante" ;
dcterms:modified "2023-04-01"^^xsd:date ;
dcterms:title "Ontologia categorias de representação da imagem em movimento"@pt .

crim:categoria a owl:Class ;
rdfs:label "Categorias"@pt .

crim:deGenerico a owl:Class ;
rdfs:label "De Generico"@pt .

crim:deEspecifico a owl:Class ;
rdfs:label "De Especifico"@pt .

crim:temDeGenerico a owl:ObjectProperty ;
rdfs:domain crim:categoria ;
rdfs:range crim:deGenerico ;
rdfs:label "temDeGenerico"@pt .

crim:temDeEspecifico a owl:ObjectProperty ;
rdfs:domain crim:deGenerico ;
rdfs:range crim:deEspecifico ;
rdfs:label "temDeEspecifico"@pt .

crim:pertenceCategoria a owl:ObjectProperty ;
rdfs:domain crim:deGenerico ;
rdfs:range crim:categoria ;
rdfs:label "pertenceCategoria"@pt .

crim:valorDeEspecifico a owl:ObjectProperty ;
rdfs:domain crim:deEspecifico ;
rdfs:range crim:deGenerico ;
rdfs:label "valorDeEspecifico"@pt .

crim:quem a crim:categoria ;
crim:temDeGenerico crim:ponte ;
rdfs:label "quem"@pt .

crim:ponte a crim:deGenerico ;
crim:temDeEspecifico crim:ponteDasBandeiras ;
rdfs:label "ponte"@pt .

crim:ponteDasBandeiras a crim:deEspecifico ;
crim:temDeGenerico crim:ponte ;
rdfs:label "ponte Das Bandeiras"@pt .

```

Classes

Propriedades de objeto

Instanciação de classe

Por conseguinte, foram realizadas consultas SPARQL na ontologia CRIM no editor de ontologia Protégé, para validar os resultados de busca e recuperar as classes, as instâncias e respectivas propriedades da ontologia.

6.6.1 Consulta SPARQL na ontologia CRIM

A partir da ontologia CRIM, foram criadas diferentes consultas para acessar as classes e instâncias da modelo de conhecimento utilizando o software Protégé:

Consulta 1:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>
```

```
SELECT ?onde
      WHERE { :Cidade :temDeEspecifico ?onde }
```

onde
Bahia

Consulta 2:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>
```

```
SELECT ?categoria ?generico
      WHERE { ?generico :pertenceCategoria ?categoria }
```

categoria	generico
ONDE	Selva
QUEM	Homem
OQUE	Morte
QUEM	Ponte
QUANDO	Primavera
ONDE	Cidade

Consulta 3:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>
```

```
SELECT ?generico ?categoria ?especie
      WHERE { ?generico :pertenceCategoria ?categoria.
              ?generico :temDeEspecifico ?especie
            }
```

generico	categoria	especie
Selva	ONDE	Amazonia
Homem	QUEM	Andrade
Homem	QUEM	Marcos
Morte	OQUE	Morte_da_rainha
Ponte	QUEM	Ponte_das_Bandeiras
Primavera	QUANDO	decadaSessenta
Cidade	ONDE	Salvador

Consulta 4:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>

```

```

SELECT ?categorias ?tem_De_Genericos
      WHERE { ?categorias rdf:type :Categorias.
              ?categorias :temDeGenerico ?tem_De_Genericos
            }

```

categorias	tem_De_Genericos
OQUE	Jogo_de_Futebol
OQUE	Morte
QUANDO	Primavera
QUEM	Homem
QUEM	Ponte
ONDE	Selva
ONDE	Cidade

Consulta 5:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>

```

```

SELECT ?categoria ?deGenerico ?tem_De_Especifico ?descricao
      WHERE {

```

```

?deGenerico :temDeEspecifico ?tem_De_Especifico;
            :pertenceCategoria ?categoria;
            rdfs:comment ?descricao
          }

```

categoria	deGenerico	tem_De_Especifico	descricao
ONDE	Selva	Amazonia	"Conceito representando um tipo de lugar geográfico, arquitetônico ou cosmográfico."@pt
QUEM	Homem	Andrade	"Conceito representando animado ou inanimado, objetos e seres concretos."@pt
QUEM	Homem	Marcos	"Conceito representando animado ou inanimado, objetos e seres concretos."@pt
OQUE	Morte	Morte_da_rainha	"Conceito que define ações, eventos e emoções."@pt
QUEM	Ponte	Ponte_das_Bandeiras	"Conceito representando animados ou inanimado, objetos e seres concretos."@pt
QUANDO	Primavera	decadaSessenta	"Conceito representando o tempo cíclico."@pt
ONDE	Cidade	Salvador	"Conceito representando um tipo de lugar geográfico, arquitetônico ou cosmográfico."@pt

Consulta 6:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>

```

```

SELECT ?categoria ?tem_De_Especifico ?descricao
      WHERE {

```

```

:Homem :temDeEspecifico ?tem_De_Especifico;
        :pertenceCategoria ?categoria;
        rdfs:comment ?descricao
      }

```

categoria	tem_De_Especifico	descricao
QUEM	Andrade	"Conceito representando animado ou inanimado, objetos e seres concretos."@pt
QUEM	Marcos	"Conceito representando animado ou inanimado, objetos e seres concretos."@pt

Consulta 7:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>
```

```
SELECT ?deGenerico ?tem_De_Especifico ?descricao
WHERE {
```

```
?deGenerico :temDeEspecifico ?tem_De_Especifico;
:pertenceCategoria :QUEM ;
rdfs:comment ?descricao
}
```

deGenerico	tem_De_Especifico	descricao
Homem	Andrade	"Conceito representando animado ou inanimado, objetos e seres concretos."@pt
Homem	Marcos	"Conceito representando animado ou inanimado, objetos e seres concretos."@pt
Ponte	Ponte_das_Bandeiras	"Conceito representando animados ou inanimado, objetos e seres concretos."@pt

Consulta 8:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>
```

```
SELECT ?categoria ?deGenerico_instancia ?descricao
WHERE {
```

```
?categoria :temDeGenerico ?deGenerico_instancia ;
rdfs:comment ?descricao
}
```

categoria	deGenerico_instancia	descricao
ONDE	Selva	"Localização da imagem no "espaço": espaço geográfico ou espaço da imagem (p. ex. São Paulo ou interior de danceteria). Fonte: (SMIT, 1996)"@pt
QUEM	Homem	"Identificação do "objeto enfocado": seres vivos, artefatos, construções, acidentes naturais, etc. Define animado e inanimado, objetos e seres concretos."@pt
ONDE	Cidade	"Localização da imagem no "espaço": espaço geográfico ou espaço da imagem (p. ex. São Paulo ou interior de danceteria). Fonte: (SMIT, 1996)"@pt
QUANDO	Primavera	"Localização da imagem no "tempo": tempo cronológico ou momento da imagem (p.ex. 1996, noite, verão). Define o tempo linear ou cíclico, datas, períodos específicos, tempos recor
OQUE	Jogo_de_Futebol	"Descrição de "atitudes" ou "detalhes" relacionados ao "objeto enfocado", quando este é um ser vivo (p.ex cavalo correndo, criança trajando roupa do século XVIII). Define ações, event
QUEM	Ponte	"Identificação do "objeto enfocado": seres vivos, artefatos, construções, acidentes naturais, etc. Define animado e inanimado, objetos e seres concretos."@pt
OQUE	Morte	"Descrição de "atitudes" ou "detalhes" relacionados ao "objeto enfocado", quando este é um ser vivo (p.ex cavalo correndo, criança trajando roupa do século XVIII). Define ações, event

Consulta 9:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>
```

```
SELECT ?deGenerico_instancia ?descricao
WHERE {
```

```
:ONDE :temDeGenerico ?deGenerico_instancia ;
rdfs:comment ?descricao
}
```

deGenerico_instancia	descricao
Selva	"Localização da imagem no "espaço": espaço geográfico ou espaço da imagem (p. ex. São Paulo ou interior de danceteria). Fonte: (SMIT, 1996)"@pt
Cidade	"Localização da imagem no "espaço": espaço geográfico ou espaço da imagem (p. ex. São Paulo ou interior de danceteria). Fonte: (SMIT, 1996)"@pt

Consulta 10:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>
```

```
SELECT ?deEspecifico_instancia ?DeEspecifico ?descricao
WHERE {
```

```
?DeEspecifico :valorDeEspecifico ?deEspecifico_instancia ;
rdfs:comment ?descricao
}
```

deEspecifico_instancia	DeEspecifico	descricao
Homem	Marcos	"Conceito representando especificamente de quem se trata a imagem. Responde as perguntas: esta imagem é de quem? De que objetos? De que seres?"@pt
Ponte	Ponte_das_Bandeiras	"Conceito representando especificamente de quem se trata a imagem. Responde as perguntas: esta imagem é de quem? De que objetos? De que seres?"@pt
Primavera	decadaSessenta	"Conceito representando o tempo linear."@pt
Jogo_de_Futebol	CopaDoMundo94	"Conceito que define eventos individualmente nomeados."@pt
Homem	Andrade	"Conceito representando especificamente de quem se trata a imagem. Responde as perguntas: esta imagem é de quem? De que objetos? De que seres?"@pt
Cidade	Bahia	"Conceito representando um nome de lugar geográfico, arquitetônico ou cosmográfico."@pt
Morte	Morte_da_rainha	"Conceito que define eventos individualmente nomeados."@pt
Selva	Amazonia	"Conceito representando um nome de lugar geográfico, arquitetônico ou cosmográfico."@pt

Consulta 11:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>
```

```
SELECT ?deGenerico_instancias ?descricao
      WHERE {

      ?deGenerico_instancias :temDeEspecifico :Marcos ;
      rdfs:comment ?descricao

}
```

deGenerico_instancias	descricao
Homem	"Conceito representando animado ou inanimado, objetos e seres concretos."@pt

Consulta 12:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example.com/OntologiaRepresentacaoImagemMovimento#>
```

```
SELECT ?deGenerico_instancias ?descricao
      WHERE {

      ?deGenerico_instancias :temDeEspecifico :Amazonia ;
      rdfs:comment ?descricao

}
```

deGenerico_instancias	descricao
Selva	"Conceito representando um tipo de lugar geográfico, arquitetônico ou cosmográfico."@pt

6.7 CONCEITOS SKOS E CLASSES OWL

O esquema de conceitos SOC 3 “Categorias de representação da imagem em movimento QUEM/QUANDO/ONDE/OQUE”, parte do resultado da pesquisa descrito na tese, pode ser ilustrado com a seguinte conceitualização SKOS:

```
categorias:evento a skos:Concept ;
  skos:prefLabel "evento"@pt ;
  skos:topConceptOf categorias: .

categorias:mortedarainha a skos:Concept ;
  skos:prefLabel "Morte da rainha"@pt ;
  skos:broader categorias:evento ;
  skos:inScheme categorias: .

categorias:morte a skos:Concept ;
  skos:prefLabel "Morte"@pt ;
  skos:broader categorias:evento ;
  skos:inScheme categorias: .
```

Para sobrepor SKOS com OWL:

```
evento: rdf:type owl:Class .

morte:  rdf:type owl:Class ;
rdfs:subClassOf classeEvento: .
```

```
morteRainha: rdf:type classeMorte: .
```

Para transformar SKOS em OWL:

```
classeEvento: rdf:type owl:Class .
classeMorte:  rdf:type owl:Class ;
rdfs:subClassOf classeEvento: .
individuoMorteRainha: rdf:type classeMorte: .
```

A abordagem, indo do mais para menos formal (OWL para SKOS), começa numa conceituação formal usando OWL e é gerado uma conceituação semiformal usando SKOS. Um exemplo, pode ser expresso a partir da seguinte conceituação de ontologia formal:

```
ex2:ClasseBioma rdf:type owl:class .
ex2:ClasseAmazonia rdf:type owl:class ;
  rdfs:subClassOf ex2:ClasseBioma.
ex2:IndividuoFlorestaAmazonica rdf:type ex2:ClasseAmazonia .
```

No padrão que sobrepõe OWL com SKOS, o SKOS é usado para sobrepôr alguns triplos adicionais, por exemplo:

```
ex2:ClasseBioma rdf:type skos:Concept.
ex2:ClasseAmazonia rdf:type skos:Concept;
  skos:broader ex2:ClasseBioma .
ex2:IndividuoFlorestaAmazonica rdf:type skos:Concept;
  skos:broader ex2:ClasseAmazonia .
```

Se os dois conjuntos de triplos forem mesclados, causa o cruzamento dos fluxos SKOS e OWL.

Para transformar OWL em SKOS, o processo é uma "transformação" e é criado um conjunto de URIs para a saída. Por exemplo, é possível gerar os seguintes triplos:

```
ex:bioma rdf:type skos:Concept .
ex:amazonia rdf:type skos:Concept ;
  skos:broader ex:bioma.
ex:florestaAmazonia rdf:type skos:Concept ;
  skos:broadet ex:amazonia .
```

Nesse padrão, a URI @prefix ex2: <https://:exemplo.com/ex#2>, foi substituída pela nova URI @prefix ex: <https://:exemplo.com/ex#>.

A ontologia CRIM resultado da pesquisa, que descrita na tese, pode ser ilustrada com a seguinte conceitualização OWL:

```
classeCategorias: a owl:Class .
IndividuoOnde: rdf:type classeCategorias: .
```

Para sobrepor OWL com SKOS:

```
classeCategorias: rdf:type skos:Concept.
classeOnde: rdf:type skos:Concept;
    skos:broader classeCategorias: .
```

Para transformar OWL em SKOS

```
categorias: rdf:type skos:Concept .
onde: rdf:type skos:Concept ;
    skos:broader categorias: .
```

Neste contexto, OWL e SKOS são usados juntos, combinados para modelar diferentes partes da mesma conceitualização, que serão híbridos formais/semiformais -- parte OWL, parte SKOS. Por exemplo, podemos usar OWL para definir uma classe de Organizações e suas propriedades relevantes:

```
ex:categoriaLugar rdf:type owl:class .
ex:TopicoDaCategoria rdf:type owl:ObjectProperty;
    rdfs:domain ex:categoriaLugar ;
    rdfs:range ex:conceitoTopicoCategoria .
ex:conceitoTopicoCategoria rdf:type owl:Class ;
    rdfs:subClassOf skos:Concept .
```

Podemos então usar SKOS para definir alguns tópicos de interesse, por exemplo:

```
ex:lugarNatureza rdf:type ex:conceitoTopicoCategoria;
    skos:prefLabel "Lugar de natureza"@pt.
ex:praia rdf:type ex:conceitoTopicoCategoria;
    skos:prefLabel "Praia"@pt;
    skos:broader ex:lugarNatureza .
ex:campo rdf:type ex:conceitoTopicoCategoria;
    skos:prefLabel "Campo"@pt;
    skos:broader ex:lugarNatureza .
```

Esse padrão, embora os vocabulários SKOS e OWL sejam usados no mesmo grafo, os "fluxos SKOS e OWL" são efetivamente mantidos separados.

Nos resultados da tese, foi definido uma modelagem que segue o padrão no qual OWL e SKOS foram usados juntos, combinados para modelar diferentes partes da mesma conceituação. Dessa forma, gerando um Híbrido formal/semiformal, com parte OWL, parte SKOS, ou seja, parte ontologia CRIM e parte SOC 3.

```

categorias: rdf:type owl:Class ;
            rdfs:subClassOf skos:Concept .
deGenerico: rdf:type owl:Class ;
            rdfs:subClassOf skos:Concept .

```

Definição do conceito como classe OWL e subclasse SKOS

```

temDeGenerico: a owl:objectProperty;
               rdfs:domain categorias: ;
               rdfs:range deGenerico: .
pertenceCategoria: a owl:objectProperty ;
                   rdfs:domain deGenerico: ;
                   rdfs:range categorias: .

```

Definição de propriedades

```

quem: rdf:type categorias: ;
       skos:prefLabel "Quem"@pt ;
       temDeGenerico: ponte: .
ponte: rdf:type deGenerico: ;
        pertenceCategoria: quem: ;
        skos:prefLabel "Ponte"@pt.

```

Definição de instâncias

O último padrão, que combina SKOS e OWL, envolve a adição de rótulos e documentação (notas) a uma ontologia formal representada usando OWL, acrescentando informações legíveis por humanos, como rótulos, notas, apenas com as propriedades de rotulagem e documentação (notas) do SKOS. Essa situação não envolve o uso de `skos:Concept`, não tendo influência na semântica de `skos:Concept`.

O SKOS é um vocabulário RDF/OWL que pode ser estendido para atender a requisitos específicos de modelagem. Da mesma forma, os recursos do SKOS também podem ser utilizados como complemento a outros vocabulários. Podendo ocorrer a reutilização de propriedades de rotulagem SKOS para descrever recursos que não são necessariamente conceitos SKOS:

```
<http://exemplo/organizacao/universidadedesapaulo#>
```

```

rdf:type owl:Class ;
rdfs:label "USP";
skos:prefLabel "Universidade de São Paulo"@pt .

```

Um sistema, ao exibir um rótulo deste recurso, pode identificar o termo "Universidade de São Paulo" como o rótulo preferencial, em vez do rótulo compatível "USP".

De tal modo, a pesquisa realizou a modelagem semântica de híbridos formais/semiformais para modelar diferentes partes da mesma conceituação da ontologia CRIM e dos SKOS resultados da pesquisa utilizando a serialização Turtle, como mostra a figura 91:

Figura 91 - Prefixos e URIS das classes OWL e indivíduos SKOS

```

@prefix crim: <https://purl.archive.org/purl/categorias-imagem-movimento#> .
@prefix quem: <https://purl.archive.org/purl/categorias-imagem-movimento#quem> .
@prefix evento: <https://purl.archive.org/purl/categorias-imagem-movimento/#evento> .
@prefix onde: <https://purl.archive.org/purl/categorias-imagem-movimento/#onde> .
@prefix deGenerico: <https://purl.archive.org/purl/categorias-imagem-movimento#deGenei> .
@prefix categorias: <https://purl.archive.org/purl/categorias-imagem-movimento#catego> .
@prefix deEspecifico: <https://purl.archive.org/purl/categorias-imagem-movimento#deEs> .
@prefix temDeGenerico: <https://purl.archive.org/purl/categorias-imagem-movimento#tem> .
@prefix temDeEspecifico: <https://purl.archive.org/purl/categorias-imagem-movimento#t> .
@prefix pertenceCategoria: <https://purl.archive.org/purl/categorias-imagem-movimento> .
@prefix valorDeEspecifico: <https://purl.archive.org/purl/categorias-imagem-movimento> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dct: <http://purl.org/dc/terms/> .

```

Prefixo e URI da ontologia CRIM

Prefixos e URI dos conceitos SKOS e indivíduos OWL

Prefixos e URI das classes OWL, que são subclasses SKOS

Prefixos e URI das propriedades OWL, que são subclasses SKOS

Prefixos e URI padrões da web semântica

Fonte: Elaborado pela autora.

Assim, foi criada a ontologia CRIM em OWL como mostra a figura 92:

Figura 92 - Criação da ontologia CRIM

```

crim: a owl:Ontology ;
skos:prefLabel "Ontologia Crim"@pt ;
dct:description "Ontologia categorias de representação da imagem em movimento"@pt ;
dct:license <http://creativecommons.org/licenses/by/4.0/> ;
dct:creator "Denise Cavalcante" ;
dct:modified "2023-04-01"^^xsd:date .

```

Criação da ontologia CRIM

Fonte: Elaborado pela autora.

Em seguidas foram criadas as classes OWL que também são subclasse SKOS (Fig.93):

Figura 93 - Classes OWL e subclasses SKOS

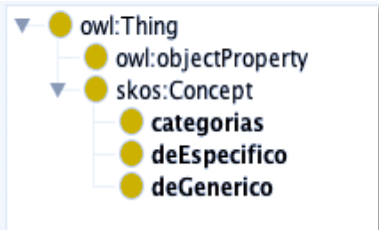
```

categorias: rdf:type owl:Class ;
skos:prefLabel "Categorias"@pt ;
skos:definition "Categorias de representação da imagem em movimento QUEM, QUANDO, ONDE, O QUE"@pt ;
rdfs:subClassOf skos:Concept .

deGenerico: rdf:type owl:Class ;
skos:prefLabel "DeGenerico"@pt ;
skos:definition "Nível de determinação De Genérico para representação da imagem em movimento QUEM, QUANDO, ONDE, O QUE"@pt ;
rdfs:subClassOf skos:Concept .

deEspecifico: rdf:type owl:Class ;
skos:prefLabel "DeEspecifico"@pt ;
skos:definition "Nível de determinação De Específico para representação da imagem em movimento QUEM, QUANDO, ONDE, O QUE"@pt ;
rdfs:subClassOf skos:Concept .

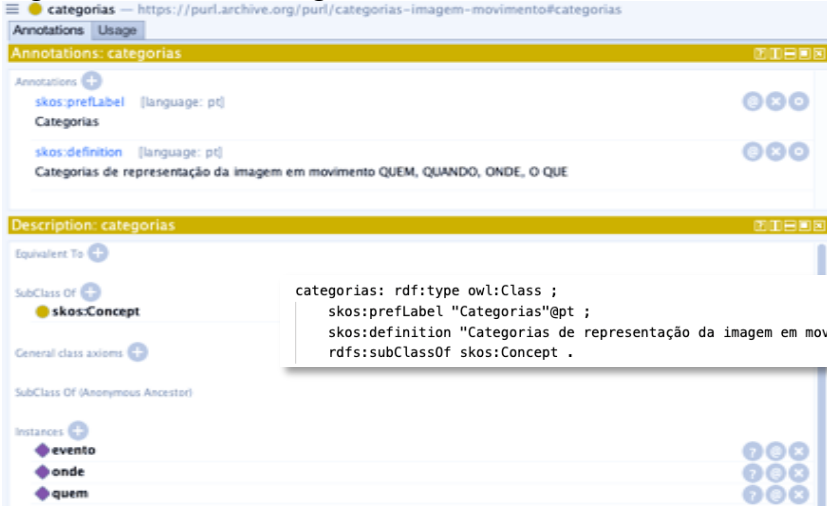
```



Fonte: Elaborado pela autora.

A figura 94 mostra a classe :Categoria modelada de duas formas, serializado em Turtle e no editor de ontologia Protégé:

Figura 94 - Classe :Categoria serializado em Turtle e no editor de ontologia Protégé



```

categorias: rdf:type owl:Class ;
skos:prefLabel "Categorias"@pt ;
skos:definition "Categorias de representação da imagem em movimento QUEM, QUANDO, ONDE, O QUE"@pt ;
rdfs:subClassOf skos:Concept .

```

Fonte: Elaborado pela autora.

A figura 95 mostra as propriedades da ontologia, serializadas em Turtle e no editor de ontologia Protégé:

Figura 95 - Propriedades serializadas em Turtle e no editor de ontologia Protégé

```

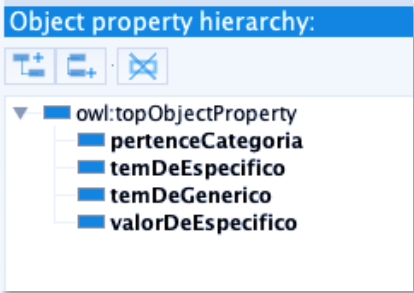
temDeGenerico: a owl:ObjectProperty;
  rdfs:domain categorias: ;
  rdfs:range deGenerico: .

temDeEspecifico: a owl:ObjectProperty ;
  rdfs:domain deGenerico: ;
  rdfs:range deEspecifico: .

pertenceCategoria: a owl:ObjectProperty ;
  rdfs:domain deGenerico: ;
  rdfs:range categorias: .

valorDeEspecifico: a owl:ObjectProperty ;
  rdfs:domain deEspecifico: ;
  rdfs:range deGenerico: .

```



Fonte: Elaborado pela autora.

A figura 96 mostra as instâncias da ontologia, serializadas em Turtle e no editor de ontologia Protégé:

Figura 96 - Instâncias serializadas em Turtle e no editor de ontologia Protégé


```

quem: rdf:type categorias: ;
  skos:prefLabel "Quem"@pt ;
  temDeGenerico: quem:ponte .

quem:ponte rdf:type deGenerico: ;
  temDeEspecifico: quem:ponteDasBandeiras ;
  pertenceCategoria: quem: ;
  skos:prefLabel "Ponte"@pt.

quem:ponteDasBandeiras rdf:type deEspecifico:
  valorDeEspecifico: quem:ponte ;
  skos:prefLabel "Ponte das Bandeiras"@pt .

```

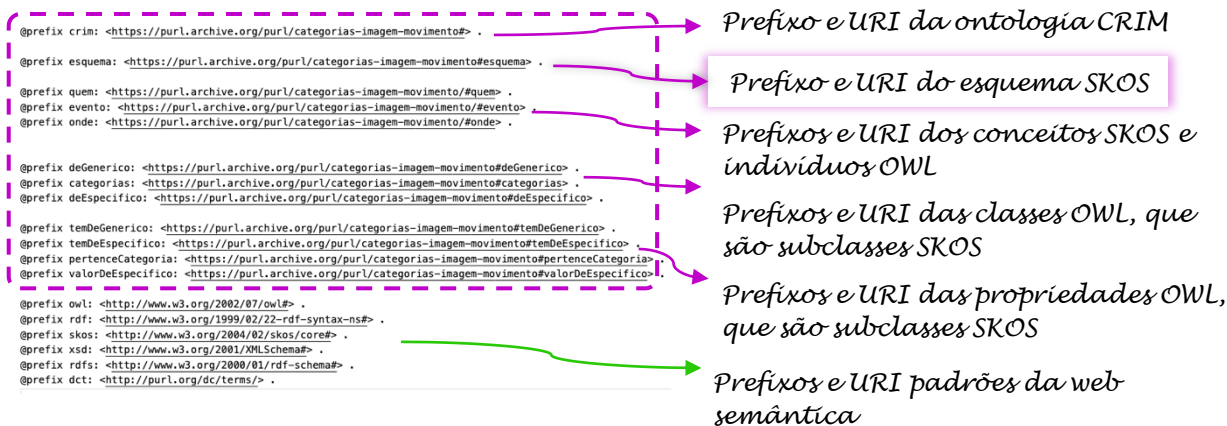


Fonte: Elaborado pela autora.

Fonte:

Ainda, foi modelado a ontologia CRIM usando conceitos de um esquema de conceitos SKOS:

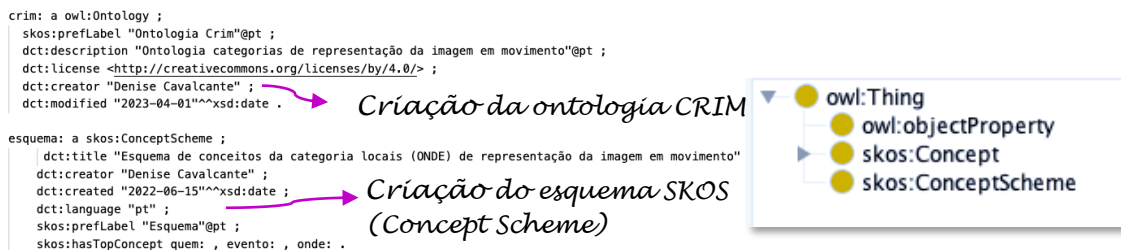
Figura 97 - Prefixo e URIs



Fonte: Elaborado pela autora.

De tal modo, foram criados a Ontologia CRIM e o esquema de conceitos SKOS como mostra a figura 98:

Figura 98 - Ontologia CRIM e o esquema de conceitos SKOS



Fonte: Elaborado pela autora.

Igualmente, foi criada a esquema de conceitos SKOS no Protégé, como mostra a figura 99:

Figura 99 - Esquema de conceitos SKOS

The screenshot shows a web browser interface for a SKOS schema. The browser address bar shows the URL: <https://purl.archive.org/purl/categorias-imagem-movimento#esquema>. The page title is 'esquema'. Below the browser interface, there are several annotations listed:

- skos:prefLabel** [language: pt]: Esquema
- dcterms:created** [type: xsd:date]: 2022-06-15
- dcterms:creator**: Denise Cavalcante
- dcterms:language**: pt
- dcterms:title**: Esquema de conceitos da categoria locais (ONDE) de representação da imagem em movimento
- skos:hasTopConcept**: evento
- skos:hasTopConcept**: onde
- skos:hasTopConcept**: quem

A code window on the right side of the page displays the following RDF/JSON-LD snippet:

```

esquema: a skos:ConceptScheme ;
  dct:title "Esquema de conceitos da categoria locais (ONDE) de representação da imagem em movimento" ;
  dct:creator "Denise Cavalcante" ;
  dct:created "2022-06-15"^^xsd:date ;
  dct:language "pt" ;
  skos:prefLabel "Esquema"@pt ;
  skos:hasTopConcept quem: , evento: , onde: .

```

Fonte: Elaborado pela autora

Ao utilizar híbridos formais/semiformais, é possível representar conceitos complexos de maneira formal, aproveitando a capacidade de inferência e raciocínio da OWL, enquanto ainda se mantém a capacidade de descrever conceitos com o SKOS.

6.8 METADADOS DO CORPUS

Para ser aplicada com a ontologia, foi selecionado um corpus de obras filmicas formado por filmes em domínio público. Para a justaposição da ontologia de domínio CRIM e com as anotações audiovisuais, usamos a modelagem de metadados usado no corpus audiovisual do projeto AdA Filmontology⁷⁴ para o nosso corpus audiovisual.

Desse modo, o corpus de metadados, para catalogação, foi dividido em metadados gerais do filme e os metadados do arquivo de vídeo. Na ontologia CRIM, os metadados estão representados em três classes: coleção, filme e vídeo. Assim, a modelagem de metadados codificados em RDF, utilizou diferentes vocabulários da Web Semântica, como DBpedia Ontology⁷⁵, Schema.org⁷⁶ e o DCMI Metadata Terms⁷⁷.

⁷⁴ Metadado do corpus AdA Project. Disponível em: <https://github.com/ProjectAdA/public/tree/master/ontology>

⁷⁵ <https://www.dbpedia.org/>

⁷⁶ <https://schema.org/>

⁷⁷ <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

Os filmes foram definidos como do tipo classe `dbo:film` e com as seguintes propriedades:

```
fcrim:1 a dbo:Film ;
    rdfs:label          "Ganga Bruta" ;
    crim:mediaURI      mcrim:1231 ;
    dcterms:isPartOf <http://example.com/CRIM/recurso/Corpus> ;
    dbo:filmVersion    " ver " ;
    dbo:releaseDate    "1933-00-00" ;
    dbo:year            "1933" .
```

A classe `Corpus` da ontologia `CRIM`, foi conectada a cada filme através da propriedade `dcterms:isPartOf`, o que indica que o filme descrito está incluído física ou logicamente no corpus. Portanto, pode ser aplicada a propriedade de objeto “`crim:mediaURI`” que tem como domínio a classe de filmes (`dbo:film`) e como range a classe de vídeo (`schema:VideoObject`) para conectar a descrição do filme com o URI do vídeo.

Os itens de mídia foram definidos como do tipo classe `schema:VideoObject` e com as seguintes propriedades:

```
mcrim:1231 a schema:VideoObject ;
    rdfs:label "Ganga Bruta_Video" ;
    schema:duration "000" ;
    schema:url      "https://youtu.be/8T8qYtERWvA" .
```

Os metadados do corpus de imagem em movimento também foram serializados em Turtle como mostra a figura 100:

Figura 99 - Metadados do corpus de imagem em movimento em Turtle

```

1 corpus_metadado.ttl
2
3 <!-- prefixo da media -->
4 @prefix mcrim: <http://example.com/OntologiaRepresentacaoImagemMovimento/recurso/media> .
5 @prefix crim: <http://example.com/OntologiaRepresentacaoImagemMovimento/recurso/> .
6 @prefix fcrim: <http://example.com/OntologiaRepresentacaoImagemMovimento/recurso/filme> .
7 @prefix schema: <http://schema.org/> .
8 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
9 @prefix dbo: <http://dbpedia.org/ontology/> .###
10 @prefix owl: <http://www.w3.org/2002/07/owl#> .
11 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
12 @prefix xml: <http://www.w3.org/XML/1998/namespace> .
13 @prefix dcterms: <http://purl.org/dc/terms/> .
14 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
15 @prefix dctypes: <http://purl.org/dc/dcmitype/> .
16
17
18 mcrim:1231
19   a schema:VideoObject ;
20   rdfs:label "Ganga Bruta_Video" ;
21   schema:duration "000" ;
22   schema:url "https://youtu.be/8T8qYTERWVA" .
23
24
25 mcrim:1232
26   a schema:VideoObject ;
27   rdfs:label "Brasa Dormida_Video" ;
28   schema:duration "000" ;
29   schema:url "https://youtu.be/EIWIWUVdMg" .
30
31
32
33 fcrim:1 a dbo:Film ;
34   rdfs:label "Ganga Bruta" ;
35   crim:mediaURI mcrim:1231 ;
36   dcterms:isPartOf <http://example.com/OntologiaRepresentacaoImagemMovimento/recurso/Corpus> ;
37   dbo:filmVersion " ver " ;
38   dbo:releaseDate "1933-00-00" ;
39   dbo:year "1933" .
40
41
42 fcrim:2 a dbo:Film ;
43   rdfs:label "Brasa Dormida" ;
44   crim:mediaURI mcrim:1232 ;
45   dcterms:isPartOf <http://example.com/OntologiaRepresentacaoImagemMovimento/recurso/Corpus> ;
46   dbo:filmVersion " ver " ;
47   dbo:releaseDate "1929-02-23" ;
48   dbo:year "1929" .
49
50
51 <http://ada.filmontology.org/resource/Corpus>
52   a dctypes:Collection ;
53   dcterms:hasPart fcrim:1, fcrim:2 .
54

```

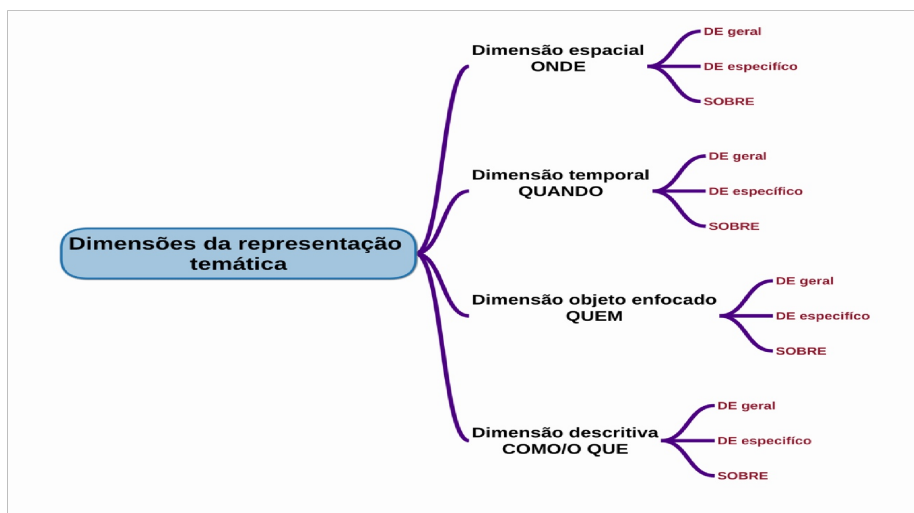
Fonte: Elaborado pela autora.

6.9 ANOTAÇÃO SEMÂNTICA DA IMAGEM

Em prosseguimento, os resultados preliminares também apontaram os possíveis cenários de aplicação do *Web Annotation Data Model*, evidenciando que aplicado a anotação profunda da imagem, possibilita a anotação da imagem no nível da sub imagem (fragmento da imagem) ou nível temporal (fragmento temporal do vídeo).

A partir disso, será definido os termos no domínio da representação da imagem em movimento, os conceitos do domínio, as classes do domínio, as hierarquias, atributos das classes e as restrições de valores e os indivíduos de cada classe. Para criar uma conceituação da representação temática no nível da imagem na totalidade e no nível da sub imagens, aplicando a anotação de regiões específicas da imagem. Para anotar as imagens será aplicado os conceitos da imagem indexados e anotados semanticamente.

Figura 100 - Dimensões da representação temática



Fonte: Baseado em Smit (1996).

Dessa forma, as anotações serão codificadas usando o Web Annotation Data Model do W3C, Media Fragments URI. Por conseguinte, a anotação é uma relação entre recursos, sendo o *body* (a descrição), que no caso representa um conceito referente alguma dimensão da representação temática, e o alvo (*target*) de um recurso, por exemplo, um filme que terá anotado uma região da imagem.

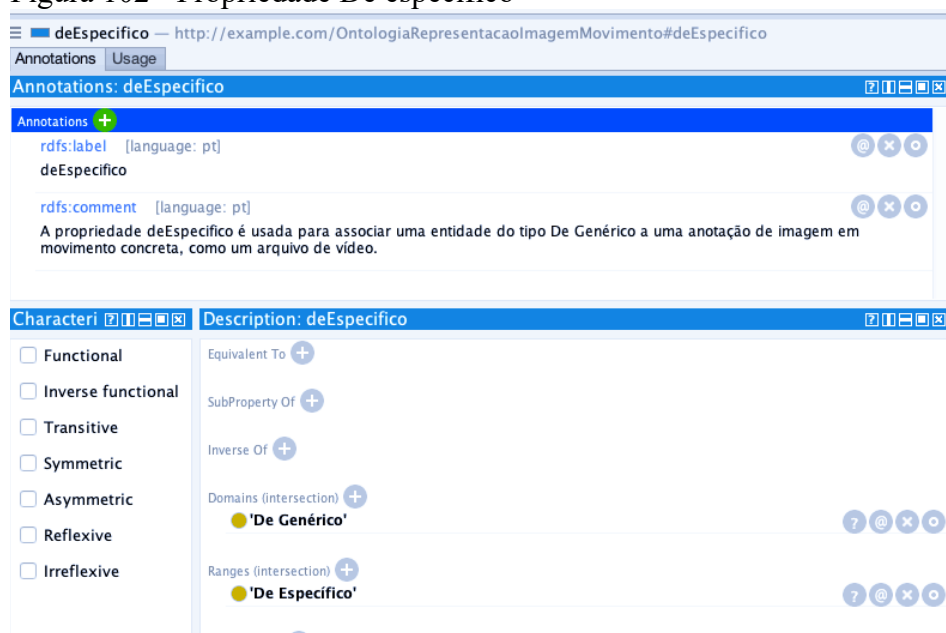
Figura 101 - Imagem anotada no nível da dimensão espacial



Fonte: Elaborado pela autora.

Logo, os metadados do corpus de obras filmicas agem como alvo (*target*) na estrutura do *Web Annotation Data Model* e foi adotado a codificação RDF usada no corpus do projeto AdA Filmontology. Para conectar os metadados das obras ao modelo de anotação foi necessário criar na ontologia CRIM, uma propriedade nomeada de *deEspecifico* e usada para associar uma entidade do tipo De Genérico a um objeto concreto, como uma mídia de vídeo, assim como, também ocorre na AdA *Filmontology*.

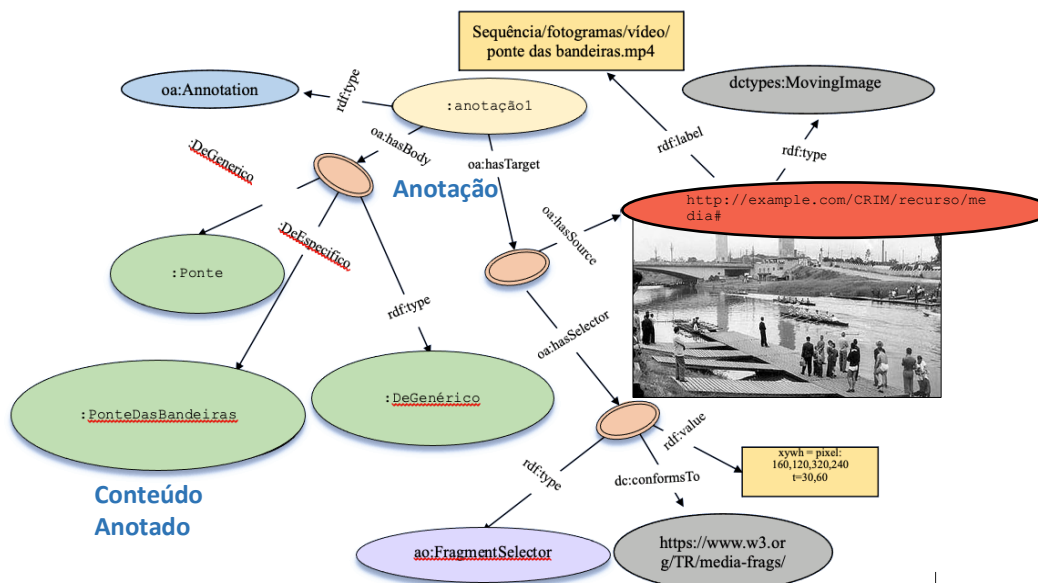
Figura 102 - Propriedade De específico



Fonte: Elaborado pela autora.

A figura 105 mostra a anotação de um recurso de vídeo com a ontologia CRIM e o modelo de anotação Web Annotation Data Model.

Figura 103 - Anotação de recurso de vídeo com a ontologia CRIM e o Web Annotation Data Model



Fonte: Elaborado pela autora.

Com a anotação profunda da imagem no nível da sub imagem, é possível anotar uma região da imagem com alguma informação relevante, como algum artefato da cultura popular ou uma escultura artística da paisagem. Na figura 105 o fotograma do filme O Canto da Saudade⁷⁸ do cineasta Humberto Mauro⁷⁹, está destacado um artefato cultural, o instrumento musical acordeão.

⁷⁸ O Canto da Saudade é um filme brasileiro de 1952, escrito, produzido e dirigido por Humberto Mauro, sendo o seu último longa-metragem.

⁷⁹ Humberto Mauro (Volta Grande, 30 de abril de 1897 — 5 de novembro de 1983) foi um dos pioneiros do cinema brasileiro. Fez filmes entre 1925 e 1974, sempre com temas brasileiros.

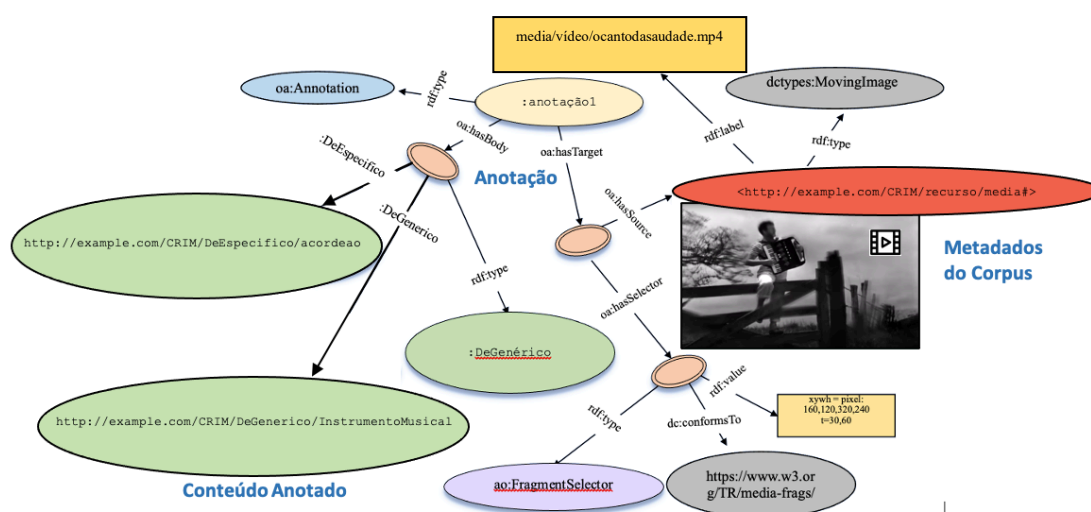
Figura 104 - Fotograma do filme "O canto da saudade"



Fonte: Elaborado pela autora.

Ainda na figura 105, o acordeão foi anotado com um recurso e uma URI HTTP, apenas como exemplo, no caso seria possível através da propriedade foaf:topic conectar a um vocabulário de valor, com o valo dessa tripla, sendo por exemplo, oriundo do tesouro de objetos patrimônio cultural brasileiro. A foaf:topic é uma propriedade que marca um tópico de alguma página ou documento, relaciona um documento a algo sobre o qual o documento se trata. O domínio dessa propriedade implica ser um documento, as imagens são subclasses da classe foaf:Document, portanto, são também documentos. Imagens digitais (como JPEG, PNG, bitmaps GIF, diagramas SVG, etc.) são exemplos de foaf:Image.

Figura 105 - Anotação de recurso com a ontologia CRIM e o Web Annotation Data Model



Fonte: Elaborado pela autora.

A pesquisa no interior de imagens, ou seja, a busca por recursos visuais discretos nas imagens, pode ser suportada pela maioria das ferramentas de anotação de imagens. O uso de vocabulários SOC para indexar anotações, melhora muito a recuperação do conteúdo da imagem, permitindo que os usuários pesquisem as ideias e conceitos por trás dos detalhes pictóricos (Clarke, 2015, p. 17, tradução nossa).

Dessa forma, as implicações da pesquisa, mostram que a anotação de imagem profunda (em inglês *deep image annotation*) envolve a anotação semântica de imagem usando KOS LOD, como tesouros conceituais modelados em SKOS, para o ambiente web que tem como objetivo a recuperação e o acesso de conceitos das imagens.

7 CONCLUSÕES

Desde seus primórdios a imagem em movimento foi enviesada pelas características do real, devido à reprodução de referentes “isto foi” e sua capacidade de reprodutibilidade técnica. A polissemia da imagem e a dificuldade de definir os significados do conteúdo dos documentos, são tangenciados pelos aspectos de denotação e conotação da imagem. São inúmeras as informações da imagem em movimento e ela engloba diferentes aspectos de linguagens, conteúdo, técnico, narrativo, cultural, entre outros. A transcodificação de imagem em texto envolve criar termos que têm relação com os assuntos de um domínio específico, todavia, nas imagens em movimento, assim como, nas obras literárias de ficção, existem desafios para criar condições de cunhar conceitos e terminologias para a área, na forma de linguagens artificiais.

A pesquisa identificou na literatura as categorias de representação da imagem fotográfica, ou seja, de indexação e classificação, as quais são ONDE, QUANDO, O QUE e QUEM. De modo que, os aspectos pré-iconográficos e iconográficos da imagem em movimento, que provocam a condição de que **toda imagem fotográfica é genérica e específica ao mesmo tempo**. Com sua qualidade de ícone, índice e símbolo, faz com que seja possível definir categorias que auxiliam na representação desse tipo de documento, as categorias ONDE, QUANDO, O QUE, QUEM, genérica e específicas.

No ambiente web estão presentes diferentes tipos de imagens em movimento que podem ser organizadas e representadas a partir de SOC e de anotação semântica. A anotação profunda se refere a identificar conceitos em regiões específicas da imagem, ou ainda, a criação índices de visualização de subimagens de áreas anotadas, essas etapas criam metadados no nível da anotação. De maneira que, apresentam a possibilidade de descoberta e recuperação de assuntos presentes no conteúdo, que de outra forma apresentariam dificuldades de serem sistematizados e estruturados, a partir de instrumentos de recuperação da informação. Bem como, oferecem desafios para a criação de termos e relações num vocabulário controlado ou de justapor a informações descritivas e a representação temática.

O trabalho identificou projetos internacionais, como o OASIS, que criaram ferramentas de anotação semântica no nível de metadados; o *I-Media-Cities* de anotação semântica manual e automática assistida por visão computacional. A ontologia AdA, que desenvolveu uma plataforma de exploração de anotação semântica dos metadados de um corpus de obras audiovisuais, alimentada por uma ontologia de domínio, que engloba a conceituação de dimensões inerentes a disciplina de análise fílmica e por último. Ademais, os experimentos da FIAF, com a ontologia InterMurnau, com enfoque na catalogação das obras incluídas, fazendo

uma reconciliação dos metadados espalhados pelas bases de dados de arquivos filmicos da Europa que contêm no seu acervo de filmes do cineasta do expressionismo alemão, F. W. Murnau.

As imagens em movimento acompanham a organização e a representação do conhecimento no século XXI, não estando mais apenas em ambientes físicos, como bibliotecas, arquivos, museus e demais unidades de informação, mas também no ambiente web. Portanto, os SOC acompanham a criação dos metadados dos documentos na web, possibilitando reconciliar dados e interoperabilizar informações entre bases de dados, assim como, entre sistemas de informação. De modo, a estabelecer relações de mapeamento, redes semânticas, reutilização de SOC, inferências, serendipidades, ente outros. Devido a isso, no ambiente web a identificação de conceitos e suas relações, são importantes pelo fato de ser um ambiente dinâmico, em constante mudança, o que faz de SOC com terminologias fixas, como tesouros lexicais, não sejam indicados para organizar o conhecimento na web.

A demanda em estabelecer a categorização de conceitos no ambiente web, faz com que a teoria do conceito ampare a normalização e as terminologias dos SOC na web, como tesouros conceituais e ontologias. Devidos as características normativas do uso da lógica, da semântica e da linguística na organização e representação da informação, através das inferências da lógica, na constituição dos campos semânticos e da composicionalidade, e nos processos de formação do signo linguístico (Signo, significado e significante).

A ambiguidade da linguagem natural em sistemas de busca, pode ser controlada por SOC semânticos para estabelecer buscas semânticas e desambiguações de resultados, através do mapeamento de entidade em classes, subclasses, instancias e relacionamentos entre instâncias.

O projeto da Web Semântica surge como modo de conectar dados através das normativas do *linked data* do W3C, utilizando linguagens baseadas em RDF, como o OWL e o modelo de dado para SOC via web, o SKOS. O diferencial do SKOS é que ele foi criado para ser usado na web, ele utiliza conceitos para representar o conhecimento e não termos lexicais. O que significa que o conceito, unidade de conhecimento, constitui propriedades que indicam relacionamentos e atributos. Ao mesmo tempo, permite que rótulos lexicais sejam adicionados, alterados, retirados, marcados como preferidos, alternativos, escondidos e permite inserção de tags de idioma, criando rótulos preferidos em diferentes idiomas.

O SKOS reflete as relações semânticas usadas em tesouros (ISO25964), com três propriedades padrões: *skos:broader*, *skos:narrower* e *skos:related*. Além disso, tem propriedades de mapeamentos, como *skos:closeMatch* e o *skos:exactMatch*, quem estabelecem

links de mapeamento (alinhamento) entre conceitos SKOS em diferentes esquemas de conceito. Ademais, conceitos de diferentes esquemas de conceito também podem ser mapeados usando as propriedades: *skos:broadMatch*, *skos:narrowMatch* e *skos:relatedMatch*.

Nesse contexto, o KOS LOD, se refere a todos os vocabulários de valor, como tesouros, listas de termos, esquemas de classificação, arquivos de autoridade, entre outros, além desses, são considerados KOS LOD, as ontologias leves na estrutura da Web Semântica, como o SKOS.

Ao contrário das ontologias leves, uma ontologia formal é uma representação formal e explícita do conhecimento em um domínio específico, que descreve conceitos, relações entre os conceitos e restrições sobre essas relações. De modo que, uma ontologia define os termos e as relações utilizadas para representar o conhecimento em um determinado contexto. As ontologias no ambiente web, são geralmente representadas usando linguagens formais, como a *Web Ontology Language* (OWL) ou o RDF. A SPARQL (*Protocol and RDF Query Language*), é uma linguagem padrão para a consulta de grafos RDF.

O *Web Annotation Model* é um conjunto de especificações e padrões desenvolvidos pelo World Wide Web Consortium (W3C) para permitir a anotação de recursos na web. Uma anotação web refere-se à capacidade de adicionar informações ou metadados a qualquer recurso da web, como uma página da web, uma imagem, um documento ou um vídeo, permitindo definir um alvo (*target*) que será anotado e o corpo da anotação (*body*).

A pesquisa aplicada partiu da modelagem semântica de microtesouros conceituais, na forma de **esquema de conceitos**, para estruturar SOC **que podem ser aplicados em documentos audiovisuais** do tipo imagem em movimento e publicados no ambiente web. Em seguida realizou a modelagem semântica de uma ontologia, usando semântica formal para **conceitualizar o domínio da indexação da imagem em movimento** no ambiente web.

Os resultados incluíram experimentos aplicados de modelagem conceitual, semântica e lógica, de microtesouros conceituais no modelo de dados SKOS, estabelecendo relações semânticas entre os conceitos, ao diferenciar o nível lexical do conceitual. Cada categoria ONDE, QUEM, QUANDO e O QUÊ, deu nome a um esquema de conceito SKOS, ou seja, cada categoria de representação temática é um SOC diferente. Os conceitos gerais e específicos se relacionam via relações semânticas do tesouro. Assim, um conceito da categoria ONDE e de nível geral, como, por exemplo, Cidade, representa um **tipo** de lugar geográfico, cosmográfico ou arquitetônico. Já o conceito da mesma categoria, mas de nível específico, representa um **nome** de lugar geográfico, cosmográfico ou arquitetônico. Esta representação de tipo e nome, resguarda o significado, enquanto os rótulos estabelecem o termo, ou seja, a forma verbal que serve como descritor.

Além de que, realizamos o alinhamento entre conceitos SKOS de esquemas diferentes, criando links entre conceitos com significados que podem ser mapeados de forma hierárquica e associativa. De forma que, por exemplo, um conceito da categoria ONDE do tipo Cidade, pode ter um relacionamento **hierárquico** com outro microtesauro que trate de estados ou países. Ou ainda estabelecer relação **associativa**, por exemplo, com um microtesauro sobre mapas ou áreas urbanas.

Ademais, testamos a criação de redes semânticas SKOS e a reutilização de conceitos de esquemas de modelagens diferentes entre si. Ao reutilizar um conceito advindo de outro esquema, foram estabelecidos vínculos e extensões no esquema que recebeu e fez o reuso dos conceitos. Isso foi possível aplicando propriedades que definiram que um determinado conceito, também pertence a outro esquema. Isso permitiu alcançar a interoperabilidade de dados e um fluxo de redes semânticas na web entre diferentes SOC.

Ainda na pesquisa aplicada, realizamos a modelagem da ontologia CRIM, que definiu classes, instâncias e propriedades de relacionamento entre classes. Assim, visamos representar epistemologicamente a conceituação das categorias de indexação QUEM, QUANDO, ONDE e O QUE e seus respectivos níveis: geral e específico. De forma, que ao usar propriedades de objeto com domínio e alcance, é possível definir as relações entre instâncias das classes Categorias, De Genérico e De Específico. Os experimentos permitiram a reconciliação de dados com vocabulários externos, a aplicação de vocabulários de valor e alinhamentos semânticos.

Portanto, a ontologia leve e a formal, foram utilizadas para anotar imagens em movimento com a *Web Annotation Data Model*, sendo empregues propriedades, da ontologia CRIM e dos esquemas SKOS, para realizar anotações. Possibilitando analisar qual tipo de classe e instância da ontologia pertence a anotação. Assim como, quais tesouros conceituais podem ser utilizados como valor da anotação. A anotação da imagem profunda possibilitou a criação de índices de conceitos, das partes que formam a imagem total. Sendo possível partir da imagem para o texto, ou, ao contrário, partir do texto para as regiões da imagem.

Como pesquisas e aplicações futuras, planejamos continuar esta pesquisa e os resultados do trabalho, estendendo para a construção de tesouros conceituais para a web. A partir da base de dados de uma biblioteca universitária e desenvolver a ontologia CRIM completa com todas as instâncias de classe, conectando e enriquecendo semanticamente a partir de outros vocabulários da Web Semântica e a publicação da ontologia como um *triplestore*⁸⁰.

⁸⁰ Um armazenamento triplestore ou RDF é um banco de dados criado especificamente para o armazenamento e recuperação de triplos [1] por meio de consultas semânticas. Um triplo é uma entidade de dados composta de sujeito- predicado -objeto, como "Bob tem 35 anos" ou "Bob conhece Fred". (Wikipédia)

Do mesmo modo, continuaremos a construção dos metadados do corpus das obras do cineasta Humberto Mauro, usando vocabulários da Web Semântica, para a criação de uma ontologia do cineasta a partir dos catálogos e bases de dados de instituições de arquivos filmicos brasileiros.

Por fim, a pesquisa tem relevância científica por explorar soluções para problemas, no contexto científico, presentes no ambiente web, no patrimônio cultural e na organização do conhecimento. Bem como, a relevância tecnológica parte da aplicação prática dos resultados em áreas interdisciplinares da tecnologia da informação, como: o setor audiovisual, a ciência da informação e a ciência da computação. Além de que, a relevância social da pesquisa é a promoção da educação e da cultura, por meio da difusão e acesso aos bens culturais do patrimônio cultural e de projetos educacionais no ambiente web.

Os aspectos de inovação, de modo geral, se concentram na resolução de problemas no ambiente web e desafios tecnológicos. Como a crescente demanda da organização de dados via categorizações e classificações, para texto e imagem. E o aumento da aplicação de taxonomias, vocabulários e ontologias em recursos na web. De modo específico, empregamos tecnologias semânticas para modelar significado e contexto aos dados e informações. Os resultados auxiliaram na interoperabilidade de dados e na recuperação de informação na web. E ampararam a modelagem e representação de domínios do conhecimento. A pesquisa desenvolveu SOC que permitem a modelagem semântica de metadados legíveis por humanos e máquinas, para dar significados aos recursos e informação que estão na web com recursos LOD.

REFERÊNCIAS

- AGUSTÍN LACRUZ, M. C. El contenido de las imágenes y su análisis en entornos documentales. *In*: GÓMEZ DÍAZ, R.; AGUSTÍN LACRUZ, M. C. (ed.). **Polisemias visuales: aproximaciones a la alfabetización visual en la sociedad intercultural**. Salamanca: Ediciones Universidad de Salamanca, 2010.
- AUMONT, J.; MARIE, M. **A análise do filme**. Lisboa: Texto e Grafia, 2004.
- AUMONT, J. **A imagem**. Campinas: Papyrus, 1993.
- ALEXIEV, V.; ISAAC, A.; LINDENTHAL, J. On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI). **International Journal on Digital Libraries**, New York, v. 17, n. 1, p. 39–48, 2016.
- ALMEIDA, M. B. **Revisiting ontologies: a necessary clarification**. [S. l.: s. n.], 2013.
- ALMEIDA, M. B.; SOUZA, R. R. Avaliação do espectro semântico de instrumentos para organização da informação 10.5007/1518-2924.2011v16n31p25. **Encontros Bibli**, Florianópolis, v. 16, n. 31, p. 25–50, 2011.
- BETHÔNICO, J. Signos audiovisuais e Ciência da Informação: uma avaliação. **Encontros Bibli**, Florianópolis, v. 11, n. 2, p. 58–78, 2006.
- BINDING, C.; TUDHOPE, D. Improving interoperability using vocabulary linked data. **International Journal on Digital Libraries**, New York, v. 17, n. 1, p. 5–21, 2016.
- BRAVO, B. R. El tratamiento documental del mensaje audiovisual. **Documentary treatment of the audio-visual message**, [s. l.], v. 19, n. 38, p. 140–160, jan. 2005.
- BAADER, F. *et al.* **Introduction to description logic**. Cambridge: Cambridge University Press, 2017.
- BARTHES, R. **Análisis estructural del relato**. Buenos Aires: Tiempo Contemporáneo, 1997.
- BARTHES, R. **A câmara clara: nota sobre a fotografia**. Rio de Janeiro: Nova Fronteira, 1984.
- BARRETO, A. A. Uma história da Ciência da Informação. *In*: TOUTAIN, L. M. B. B. (org.). **Para entender a Ciência da Informação**. Salvador: EDUFBA, 2007. p. 13-34.
- BAZIN, A. **O cinema: ensaios**. São Paulo: Brasiliense, 1991.
- BENJAMIN, W. **Sobre arte, técnica, linguagem e política**. Lisboa: Relógio D'Água Editores, 1992.
- BERNERS-LEE, T. **Linked data: design issues**. [S. l.: s. n.], 2006. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 25 nov. 2019.

BERNERS-LEE, T. **Plenary at WWW Geneva 94**. [S. l.: s. n.], 1994. Disponível em: <https://www.w3.org/Talks/WWW94Tim/>. Acesso em: 22 jul. 2019.

BERNERS-LEE, T. **Metadata architecture**. [S. l.: s. n.], 1997. Disponível em: <https://www.w3.org/DesignIssues/Metadata.html>. Acesso em: 11 abri. 2021

BERNERS-LEE, T. **Meaning**. [S. l.: s. n.], 1999. Disponível em: <https://www.w3.org/DesignIssues/Meaning.html>. Acesso em: 5 mar. 2019

BERNERS-LEE, T. **Generic resources**. [S. l.: s. n.], 1992. Disponível em: <https://www.w3.org/DesignIssues/Generic.html>. Acesso em: 17 maio 2021

BERNERS-LEE, T. **What do HTTP URIs Identify?**, [S. l.: s. n.], 2002. Disponível em: <https://www.w3.org/DesignIssues/HTTP-URI.html#L920>. Acesso em: 2 jun. 2019.

BRASCHER, M. Semantic relations in knowledge organization systems. **Knowledge Organization**, Frankfurt, v. 41, n. 2, p. 175-180, 2014.

BUCKLAND, M. What kind of Science can Information Science be? **Journal of the American Society for Information Science and Technology (ASIS&T)**, New York, v. 63, n. 1, p. 1-7, 2012. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/asi.21656>. Acesso em: 31 abr. 2019.

BOYCE, S.; PAHL, C. **Developing domain ontologies for course content**. Educational Technology & Society, 2007.

CANÇADO, M. **Manual de semântica**. Belo Horizonte: UFMG, 2008.

CAPURRO, R. Epistemologia e Ciência da Informação. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 5., 2019. **Anais [...]** Belo Horizonte: ECI/UFMG, 2003. Disponível em: http://www.capurro.de/enancib_p.htm. Acesso em: 08 jun. 2019.

CALDERA-SERRANO, J.; CARO-CASTRO, C. **Limitaciones y problemas de la lectura de imagen en movimiento**. [S. l.: s. n.], 2013. Acesso em: 21 out. 2020.

CALDERA-SERRANO, J.; SÁNCHEZ-JIMÉNEZ, R. Recuperación de secuencias de información audiovisual con rdf y smil. **El Profesional de la Información**, Barcelona, v. 18, n. 3, p. 291–300, 2009.

CARACENI, S.; GUIDAZZOLI, A.; IMBODEN, S.; LIGUORI, M. C.; MONTANARI, M.; SCIPIONE, G.; TROTTA, G. I. **Media-cities: a digital ecosystem enriching a searchable treasure trove of audio visual assets**, DataverseNL. [S. l.], 2019.

CASTRO, C. C. Vocabulários estruturados, Web Semântica y Linked Data: oportunidades y retos para los profesionales de la documentación 1. *In*: SEMINÁRIO DE ESTUDOS DA INFORMAÇÃO: ARQUIVOLOGIA, BIBLIOTECONOMIA E CIÊNCIA DE INFORMAÇÃO: IDENTIDADES, CONTRASTES E PERSPECTIVAS DE INTERLOCUÇÃO, 2., Niterói. **Anais [...]** Niterói: Universidade Federal Fluminense, 2011. p. 26-28.

CLARKE, D. Deep image annotation: making a difference in knowledge organization. *In: INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION, FOURTH BIENNIAL CONFERENCE*, London, 2015. **Proceedings** [...] [S. l.: s. n.], 2015. p. 1–21.

CLARKE, D. **Using linked data as a framework for enterprise KOS Dublin Core**. DC2018. Porto: [s. n.], 2018.

CORDEIRO, R. I. D. N. Análise de Imagens e Filmes: alguns princípios para sua indexação e recuperação. **Ponto de Acesso**, Salvador, v. 7, n. 1, p. 67, 2013.

CORDEIRO, R. I. D. N.; LA BARRE, K. Análise de facetas e obra fílmica. **Informação & Informação**, Londrina, v. 16, n. 2, p. 180, 15 2011.

CORDEIRO, R. I. DE N.; AMÂNCIO, T. Análise e representação de filmes em unidades de informação. **Ciência da Informação**, Brasília, DF, v. 34, n. 1, p. 89-94, 2005.

COSTA, F. **O primeiro cinema: espetáculo, narração e domesticação**. Rio de Janeiro: Azougue, 1995.

COPI, I. **Introdução à lógica**. São Paulo: Mestre Jou, 1978.

DAHLBERG, I. A Referent-oriented, Analytical Concept Theory of Interconcept. **International Classification**, Munchen, v. 5, p. 122-151, 1978a.

SOARES, M. F.; OLIVEIRA, B. M. J. F. Librarian's image: an analysis in movie films TT - Imagem bibliotecaria(o): uma análise em películas cinematográficas. **Biblionline**, [s. l.], v. 1, n. 1, 2005.

DE SMEDT, J.; ISAAC, A.; CLARKE, S. D.; LINDENTHAL, J.; ZENG, M. L.; TUDHOPE, D. S.; WILL, L.; ALEXIEV, V. ISO 25964 part 1: thesauri for information retrieval: RDF/OWL vocabulary, extension of SKOS and SKOS-XL. **Education for Library & Information Science**, State College, v. 60, n. 2, p. 139–151, 2019. Disponível em: <http://purl.org/iso25964/skos-thes>. Acesso em: 25 nov. 2019.

DOBRESKI, B.; ZENG, M. L.; KOVARI, J.; QIN, J. **Linked data education and training: past, present, and future**. [S. l.]: Proc Assoc Inf Sci Technol, 2020.

DOMÍNGUEZ-DELGADO, R.; LÓPEZ-HERNÁNDEZ, M.-Á. Análisis documental del contenido fílmico en seis filmotecas españolas TT - Film content analysis at six major Spanish film libraries. **El Profesional de la Información**, Barcelona, v. 25, n. 5, p. 787–794, 2016.

DOMÍNGUEZ-DELGADO, R.; LÓPEZ HERNÁNDEZ, M. Á. Film content analysis on FIAF cataloguing rules and CEN metadata standards. **Proceedings of the Association for Information Science and Technology**, [s. l.], v. 54, n. 1, p. 655–657, 2017a.

DOMÍNGUEZ-DELGADO, R.; LÓPEZ-HERNÁNDEZ, M.-Á. Una perspectiva histórica del análisis documental de contenido fílmico TT - A historical perspective of the documentary analysis of filmic content. **Documentación de las Ciencias de la Información**, Madrid, v. 40, p. 73–90, 2017b.

DUBOIS, P. **O ato fotográfico e outros ensaios**. Campinas: Papirus, 1993.

EDMONDSON, R. **Arquivística audiovisual: filosofia e princípios**. Brasília, DF: UNESCO, 2017.

FIACCARINI, A; BIESBROUCK, B; MACCONNACHIE, S. **Extending the FIAF moving image cataloguing manual and fiaf glossaries using linked open data**. [S. l.]: FIAF Cataloguing and Documentation Commission, 2017.

FIORIN, J. L. Teoria dos signos. *In*: FIORIN, J. L. (org.) **Introdução à lingüística: objetos teóricos**. São Paulo: Contexto, 2002. v. 1, p. 55-74.

GALVÃO, M. C. B.; RICARTE, I. L. M. Revisão sistemática da literatura: conceituação, produção e publicação. **Logeion: Filosofia da Informação**, Rio de Janeiro, v. 6, n. 1, p. 57–73, 2019. DOI: 10.21728/logcion.2019v6n1.p57-73. Disponível em: <https://revista.ibict.br/fiinf/article/view/4835>. Acesso em: 25 set. 2021.

GAUDREAU, A. L’histoire du cinéma revisitée: le cinema des premiers temps, 1988. **CinémAction. Les théories du cinéma aujourd’hui**, [s. l.], v. 47, p. 102-108, 1988.

GONÇALVES, A. C. B. Os novos paradigmas da imagem em movimento: em busca de metalinguagens de representação para bases de dados virtuais visando a recuperação de conteúdo semântico. **DataGramZero**, [s. l.], v. 24, n. 2, p. 19, 2018.

GONZÁLEZ, J. A. M. Hacia la primacía de los conceptos sobre los términos en los vocabularios para la web semántica. **Anuario ThinkEPI**, [s. l.], v. 7, p. 173, 2013.

GONZÁLEZ, J. A. M.; ARILLO, J. R. **O conteúdo da imagem**. Curitiba: Ed. da UFPR, 2003.134 p.

GRUBER, T. R. A Translation approach to portable ontology specifications. **Knowledge Acquisition**, London, v. 2, p. 199-220, 1993.

GUARINO, N. Formal ontology in information systems. *In*: **FORMAL ONTOLOGY IN INFORMATION SYSTEMS**, Trento, 1998. **Proceedings [...]** Amsterdam: IOS Press, 1998. p. 3-15.

GUIZZARDI, G.; HERRE, H.; WAGNER, G. On the general ontological foundations of conceptual modeling. *In*: SPACCAPIETRA, S., MARCH, S.T., KAMBAYASHI, Y. (ed.) **Conceptual Modeling — ER**. Heidelberg: Springer, 2002. v. 2503, p. 65.

HEFTBERGER, A.; DUCHESNE, P. **Cataloguing practices in the age of linked open data: wikidata and wikibase for film archives**. Brussels: International Federation of Film Archives, 2020. Disponível em: <https://www.fiafnet.org/pages/E-Resources/Cataloguing-Practices-Linked-Open-Data.html>. Acesso em: 17 ago. 2020.

HEFTBERGER, A.; DUCHESNE, P. Merging Murnau: Wikibase and Linked Open Data for Archives. **Journal of Film Preservation**, Brussels, n. 105, p. 47-54, 2021.

HJØRLAND, B. What Is Knowledge Organization (KO)?. **Knowledge Organization**, Frankfurt, v. 35, n. 2/3, p. 86–102, 2008.

HJØRLAND, B. Theories are knowledge organizing systems (KOS). **Knowledge Organization**, Frankfurt, v. 42, n. 2, p. 113–128, 2015.

HJØRLAND, B. **Knowledge organization**. [S. l.: s. n.], 2016. (Nota técnica).

HODGE, G. **Systems of knowledge organization for digital libraries: beyond traditional authority files**. Washington, DC: Council on Library and Information Resources, 2000.

HYVÖNEN, E. Using the semantic web in digital humanities: shift from data publishing to data-analysis and serendipitous knowledge discovery. **Semantic Web**, Amsterdam, v. 11, n. 1, p. 187-193, 2020.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados: em busca da web do conhecimento**. [S. l.]: Novatec Editora. 2014.

ISAAC, A.; SUMMERS, E. (ed.). **SKOS simple knowledge organization system primer**. [S. l.]: W3C Working Group, 2009. Disponível em: <https://www.w3.org/TR/skos-primer/>. Acesso em: 22 set. 2020.

KOBASHI, N; FRANCELIN, M. Conceitos, categorias e organização do conhecimento. **Informação & Informação**, Londrina, v. 16, n. supl, p. 1–24, 2011. DOI: 10.5433/1981-8920.2011v16nesp.p1. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/>. Acesso em: 22 set. 2020.

KOBASHI, N. **Elementos lógicos e linguísticos na organização e representação da informação**. Brasília, DF: CAPES: UAB; Rio de Janeiro, RJ: Departamento de Biblioteconomia, FACC/UFRJ, 2019.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Brasília, DF: Briquet de Lemos, 2004. Disponível em: <https://bibliotextos.files.wordpress.com/2014/07/livro-indexac3a7c3a3o-e-resumosteoria-e-prc3a1tica-lancaster.pdf>. Acesso em: 20 de jul. 2020.

LIBRARY OF CONGRESS. *Site*. Washington, 21. Disponível em: <https://www.loc.gov/>. Acesso em: 20 de jul. 2020.

LIMA, V. M. A. **A documentação audiovisual: tópicos para o ensino de biblioteconomia**. São Paulo: ECA-USP, 2016. V. 1, p. 190. Disponível em: <https://www.eca.usp.br/acervo/producao-academica/002749728.pdf>. Acesso em: 05 jul. 2022.

LICHTENSTEIN, A.; PLANK, M.; NEUMANN, J. TIB's Portal for Audiovisual Media: Combining Manual and Automatic Indexing. **Cataloging and Classification Quarterly**, Philadelphia, v. 52, n. 5, p. 562–577, 4 jul. 2014.

LOPES, E. **Fundamentos da lingüística contemporânea**. 4. ed. São Paulo: Cultrix, 1980.

LÓPEZ-YEPES, A. **Patrimonio audiovisual iberoamericano Online (2013-2017)**. España, Argentina, Bolivia, Brasil, Colombia, Costa Rica, Cuba, Chile, Ecuador, El Salvador, Guatemala, México, Paraguay, Perú, Uruguay, Venezuela. Madrid: Servicio Documentación Multimedia. Dpto. Biblioteconomía y Documentación. Universidad Complutense Madrid, 2017. v. 1.

LÓPEZ-YEPES, A.; SÁNCHEZ-JIMÉNEZ, R.; PÉREZ-AGÜERA, J.-R. **Tratamiento de la documentación audiovisual en el entorno digital**: iniciativas de metadatos y lenguajes de descripción multimedia. El profesional de la información. Barcelona: EPI SCP, 2003. Disponível em: <http://eprints.rclis.org/3975/>. Acesso em: 3 fev. 2019.

MACAMBYRA, M. **Seres imaginários na academia**: como indexar filmes com o Vocabulário da USP. São Paulo: Imagem falada, 2013. Disponível em: <https://imagemfalada.wordpress.com/2013/06/02/seres-imaginarios-na-academia/>. Acesso em: 3 fev. 2019.

MACAMBYRA, M. **Sobre o que é esse filme?** São Paulo: Imagem falada, 2017. Disponível em: <https://imagemfalada.wordpress.com/2017/03/02/indexacao-assunto-filmes/>. Acesso em: 4 fev. 2019.

MACAMBYRA, M. **Quadro para indexação de filmes**. São Paulo: Imagem falada, 2020. Disponível em: <https://imagemfalada.wordpress.com/2020/04/19/quadro-para-indexacao-de-filmes/>. Acesso em: 5 de fev. 2022.

MAZZOCCHI, F. Knowledge organization system (KOS). **Knowledge organization**, Frankfurt, v. 45, n.1, p. 54-78, 2018.

MORRIS, C. **Fundamentos da teoria dos signos**. São Paulo: Universidade de São Paulo, 1976.

PETTER, M. Linguagem, língua, linguística. *In*: FIORIN, J. L. (org.). **Introdução à linguística**: objetos teóricos. São Paulo: Contexto, 2002.

MACEDO, L. Reler Metz: por uma semiologia dos códigos e das escritas fílmicas. *In*: 42º CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO – INTERCOM, 42., 2019, Belém. **Anais [...]** São Paulo: Intercom, 2019.

MANOLA, F.; MILLER, E. (ed.). **RDF Primer**. [S. l.]: W3C, 2004. Disponível em: <https://www.w3.org/TR/rdf-primer/>. Acesso em: 3 fev. 2019.

MENDONÇA, F. M.; FELIPE, E. R. Construção de ontologias: problemas, metodologias, editores e direcionamentos. *In*: ALMEIDA, M. (org.). **Representação do conhecimento, ontologias e linguagem**: pesquisa aplicada em Ciência da Informação. Curitiba: CRV, 2020. v. 1, p. 17–46.

METZ, C. **A significação no cinema**. São Paulo: Perspectiva, 1972.

METZ, C. **Linguagem e cinema**. São Paulo: Perspectiva, 1980.

MCGUINNESS, D. L.; VAN HARMELEN, F. Web ontology language. *In*: **W3C recommendation**. [S. l.]: W3C, 2004. (W3C Recommendation).

MILES, A.; BECHHOFFER, S. **SKOS simple knowledge organization system reference**. [S. l.]: World Wide Web Consortium, 2009. (W3C Recommendation). Disponível em: <http://www.w3.org/TR/skos-reference/>. Acesso em: 3 fev. 2019.

MILES, A.; BECHHOFFER, S. **Using OWL and SKOS**. [S. l.: s. n.], 2008. Disponível em: <https://www.w3.org/2006/07/SWD/SKOS/skos-and-owl/master.html>. Acesso em: 29 nov. 2022.

MOURA, M. A. *et al.* Linguagens de indexação em contextos cinematográficos: a experiência de elaboração do tesauto eletrônico do cinema brasileiro. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 24, n. 2, 19 abr. 2018.

MOTA, D.; KOBASHI, N. Semantic Web and Pragmatic Web: Critical Discussion of Versioning on the Web and Conceptual Limits. **Tendências da Pesquisa Brasileira em Ciência da Informação**, Joao Pessoa, v. 9, n. 2, p. 1–18, 2017.

NICHOLS, B. **Introdução ao documentário**. Campinas: Papirus, 2005.

NÖTH, W.; GURICK, A. A teoria da informação de Charles S. Peirce. **TECCOGS: Revista Digital de Tecnologias Cognitivas**, São Paulo, v. 4, p. 1-29, 2011.

NOY, N. F.; MCGUINNESS, D. L. **Ontology development 101: a guide to creating your first ontology**. [S. l.: s. n.], 2020. Disponível em: <http://www.wksl.stanford.edu/people/dlm/papers/ontology-tutorialnoy-mcguinness.pdf>. Acesso em: 02 abr. 2020.

PANOFSKY, E. **O significado nas artes virtuais**. Barbacena: Editorial Presença, 1989.

PARRA, H. Z. M. Tendências democráticas e autoritárias nas tecnologias de comunicação digital em redes cibernéticas. *In*: ENCONTRO ANUAL DA ANPOCS, 38., 2014, Caxambu. **Anais [...]** Caxambu: ANPOCS, [versão preliminar– 29/08/2014]. GT04 - Ciberpolítica, ciberativismo e cibercultura.

PASTOR-SÁNCHEZ, J. A. *et al.* **Audio-visual semantics: propuesta de una ontología para la descripción de secuencias audiovisuales**. [S. l.: s. n.], 2019. Disponível em: <http://eprints.rclis.org/38545/>. Acesso em: 02 abr. 2020.

PASTOR-SANCHEZ, J.-A.; MARTÍNEZ-MENDEZ, F. J.; RODRIGUEZ-MUÑOZ, J. V. Advantages of thesaurus representation using the Simple Knowledge Organization System compared with proposed alternatives. **Information Research**, Boraas, v. 14, n. 4, p. 1-15, 2009. Disponível em: <http://www.informationr.net/ir/14-4/paper422.html>. Acesso em: 22 jan. 2022.

PASTOR-SÁNCHEZ, J. A. Los tesauros en la web semántica: SKOS y la norma ISO 25964. El lenguaje sobre el patrimonio. **Estándares documentales para la descripción y gestión de colecciones**, [S. l.], p. 47–72, 2016. Disponível em: <https://sede.educacion.gob.es/publiventa/el-lenguaje-sobre-el-patrimonio-estandares-documentales-para-la-descripcion-y-gestion-de-colecciones/cultura-sociedad-lenguaje-patrimonio-historico-artistico/20819C>. Acesso em: 22 jan. 2022.

PASTOR-SÁNCHEZ, J. *et al.* **Audio-Visual Semantics: propuesta de una ontología para la descripción de secuencias audiovisuales**. [S. l.: s. n.], 2019. Disponível em: <http://eprints.rclis.org/38545/>. Acesso em: 22 jan. 2022.

PEIRCE, C. S. **Semiótica**. São Paulo: Editora Perspectiva, 2005.

PENA, C. N. Tesouro documental para la representacion de contenidos educativos en videos. The documentary thesaurus for representation of educational contents in videos.

Investigacion Bibliotecologica: Archivonomia, Bibliotecologia, e Informacion, Ciudad Universitaria, v. 17, n. 34, p. 144–174, jan. 2003.

PIZZANI, L. *et al.* A arte da pesquisa bibliográfica na busca do conhecimento. **RDBCI**: Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, v. 10, n. 2, p. 53–66, 2012. DOI: 10.20396/rdbci.v10i1.1896. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/1896>. Acesso em: 25 set. 2021.

GRACY, K. F. Enriching and enhancing moving images with Linked Data. **Journal of Documentation**, Bradford, v. 74, n. 2, p. 354-371, 2018.

ROCKO, E. B. **El documentalista en el entorno actual de los medios de comunicación audiovisuales como gestor de imágenes**. [S. l.]: SEDIC, 2009. Disponível em: <http://eprints.rclis.org/13875/>. Acesso em: 25 set. 2021.

ROSA, A. B. A. R.; SANTOS, R. A. A.; SIMIONATO, A. C. A documentação audiovisual para o ensino da biblioteconomia e ciência da informação no Brasil. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, v. 8, n. 2, 2018.

RIO, Zona Norte. Direção: Nelson Pereira dos Santos. Intérprete: Grande Otelo; Jece Valadão; Paulo Goulart. Roteiro: Nelson Pereira dos Santos. Brasil: [s. n.], 1957. DVD, pb, 90 min.

SANTOS, F. E. P. *et al.* Definição de metadados e critérios de indexação para documentário em repositório audiovisual. **Revista Brasileira de Biblioteconomia e Documentação**, Sao Paulo, v. 14, n. 3, p. 237–261, 2018.

SAUSSURE, F. *et al.* **Aula 2 e 3 curso de linguística geral**. [S. l.: s. n.], 2006.

SASTRE, N. **El centro de documentación fotográfica y audiovisual de Inspai, centro de la imagen de la diputació de Girona**. [S. l.: s. n.], 2009. Acesso em: 21 out. 2020.

SANDERSON, R.; CICCARESE, P.; VAN DE SOMPEL, H. **Open annotation draft data model body**. [S. l.: s. n.], 2013. Disponível em: <http://www.commonsemantics.com/oa/OpenAnnotationDataModelPrimer.html>. Acesso em: 25 dez. 2020.

SANDERSON, R.; CICCARESE, P.; YOUNG, B. **Web annotation data model**. [S. l.]: W3C, 2015.

SANDERSON, R.; CICCARESE, P.; YOUNG, B. **Web Annotation vocabulary**. [S. l.]: W3C, 2017. Disponível em: <https://www.w3.org/TR/annotation-vocab/>. Acesso em: 4 out. 2019.

SAUERMAN, L.; CYGANIAK, R. **Cool URIs for the Semantic Web**. [S. l.: s. n.], 2008. Disponível em: <http://www.w3.org/TR/cooluris>. Acesso em: 4 out. 2019.

- SCHREIBER, G.; RAIMOND, Y. **RDF 1.1 primer**. [S. l.: s. n.], W3C Working Group Note, 2014. Disponível em: <http://www.w3.org/TR/rdf11-primer/>. Acesso em: 22 ago. 2020.
- SILVA, A. O. A. DA; CORDEIRO, R. I. N. Indexação de filmes: um estudo sobre a representação do audiovisual. **Encontro Nacional de Pesquisa em Ciência da Informação**, [s. l.], v. 24, n. 2, 19 abr. 2018.
- SMIT, J. W. **A representação da imagem**. São Paulo: USP, 1996. p. 28–36.
- SMIT, J. A informação na Ciência da Informação. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 3, n. 2, p. 84-101, jul./dez. 2012.
- SOUZA, J. J. Publicações científicas sobre documentação audiovisual na Ciência da Informação e Ciências da Comunicação de 1984 a 2019. **Biblionline**, [s. l.], v. 16, n. 2, p. 43–30, 19 abr. 2020.
- SOUZA, R. R.; SILVA, D. L.; Modelagem conceitual baseada em ontologias na organização de documentos multimídia. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador. **Anais [...]** Salvador: ENANCIB, 2016. p. 1–20.
- SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems. **Knowledge Organization**, Frankfurt, v. 39, n. 3, 179-192, 2012.
- SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jul.1996. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>. Acesso em: 17 mar. 2019.
- SILVEIRA, D. T.; CÓRDOVA, F. P. A pesquisa científica. *In*: GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.
- SHATFORD, S. Analyzing the subject of a picture: a theoretical approach. **Cataloging and Classification Quarterly**, New York, v. 6, n. 3, p. 39-62, 1986.
- SOWA, J. F. Ontology, metadata, and semiotics. **In Conceptual Structures: Logical, Linguistic, and Computational Issues: 8th International Conference on Conceptual Structures, ICCS 2000**, Darmstadt, Germany, August 14-18, 2000.
- SOWA, J.F. Future Directions for Semantic Systems. **Tolk, A., Jain, L.C. (eds) Intelligence-Based Systems Engineering. Intelligent Systems Reference Library**, vol 10. Springer, Berlin, Heidelberg, 2011.
- SUNDSTRÖM, A. D. S. S.; MORAES, J. B. E. DE; ALBUQUERQUE, A. C. Filme de ficção para a Ciência da Informação: um estudo sobre as abordagens de organização e representação temática. **Encontros Bibli**, Florianópolis, v. 24, n. 54, p. 124–134, 4 jan. 2019.
- SWARTOUT, W.; TATE, A. Ontologies. **IEEE Intelligent Systems**, Piscataway, v. 1, n. 14, p. 18-19, 1999.

TENNIS, J.; CALZADO-PRADO, J. Ontologies and the Semantic Web: problems and perspectives for LIS professionals. **Ibersid: revista de sistemas de información y documentación**, Zaragoza, v. 1, p. 303–311, 2007.

TUGENDHAT, E.; WOLF, U. **Propedêutica lógico-lingüística**. Petrópolis: Vozes, 1997.

TODOROV, T. Las Categorías del Relato Literario. *In*: BARTHES, R. *et al.* **Análisis Estructural del Relato**. Buenos Aires: Tiempo Contemporáneo, 1974. p. 155-192.

TURNER, J. Representing and accessing information in the stockshot database at the National Film Board of Canada. **Canadian Journal of Information Science**, London, v. 15, n.4, dez. 1990.

TURNER, J. Indexing film and video images for storage and retrieval. **Information Services and Use**, Amsterdam, v. 14, n. 3, p. 225–236, 1994.

TURNER, J. M.; MATHIEU, S. Audio description text for indexing films. **IFLA Conference Proceedings**, The Hague, p. 1–11, nov. 2007.

VATANT, B. Porting library vocabularies to the semantic web, and back: a win- win round trip. *In*: INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS CONGRESS, 76 th., 2010, Gothenburg. **Proceedings** [...] [*S. l.: s. n.*], 2010. p. 1-11.

VAN MALSSSEN, K. The FIAF moving image cataloguing manual. **The Journal of Film Preservation**, [*s. l.*], v. 96, 2017.

VIEIRA, R. *et al.* **Web semântica: ontologias, lógica de descrição e inferência**. Tutorial do XI Simpósio Brasileiro de Sistemas Multimídia e Web (WEBMEDIA05). [*S. l.: s. n.*], 2005.

ZABALA-VÁZQUEZ, J.; SÁNCHEZ-GALÁN, M.-B. **Advertising documents in film archives**: importance, present situation and difficulties of analysis. . [*S. l.: s. n.*], 2014.

ZENG, M. L. Semantic enrichment for enhancing lam data and supporting digital humanities. Review article. **Profesional de la Informacion**, Barcelona, v. 28, n. 1, p. 1–36, 2019.

ZENG, M. L.; MAYR, P. Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review. **International Journal on Digital Libraries**, Heidelberg, v. 20, n. 3, p. 1–22, 2018.

