



INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES

Autarquia associada à Universidade de São Paulo

**Desenvolvimento de uma Camada Semântica para um Protótipo
de Repositório de Dados de Pesquisas de Análise por Ativação
Neutrônica**

GLAUBER MAUCH DE CARVALHO

Dissertação apresentada como parte dos requisitos para obtenção do Grau de Mestre em Ciências na Área de Tecnologia Nuclear – Aplicações;

Orientador:

Mário Olímpio de Menezes

SÃO PAULO

2020

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Autarquia associada à Universidade de São Paulo

**Desenvolvimento de uma Camada Semântica para um Protótipo
de Repositório de Dados de Pesquisas de Análise por Ativação
Neutrônica**

Versão Corrigida
Versão Original disponível no IPEN

GLAUBER MAUCH DE CARVALHO

Dissertação apresentada como parte
dos requisitos para obtenção do Grau de
Mestre em Ciências na Área de
Tecnologia Nuclear – Aplicações;

Orientador:
Prof. Dr. Mário Olímpio de Menezes

SÃO PAULO
2020

Autorizo a reprodução e divulgação total ou parcial deste trabalho,
para fins de estudo e pesquisa, desde que citada a fonte

Como citar:

MAUCH DE CARVALHO, G. ***Desenvolvimento de uma Camada Semântica para um Protótipo de Repositório de Dados de Pesquisas de Análise por Ativação Neutrônica.*** 2020. 64 p. Dissertação (Mestrado em Tecnologia Nuclear), Instituto de Pesquisas Energéticas e Nucleares, IPEN-CNEN/SP, São Paulo. Disponível em: (data de consulta no formato: dd/mm/aaaa)

Ficha catalográfica elaborada pelo Sistema de geração automática da Biblioteca IPEN/USP,
com os dados fornecidos pelo(a) autor(a)

Mauch de Carvalho, Glauber
Desenvolvimento de uma Camada Semântica para um Protótipo
de Repositório de Dados de Pesquisas de Análise por Ativação
Neutrônica / Glauber Mauch de Carvalho; orientador Mário
Olimpio de Menezes. -- São Paulo, 2020.
64 p.

Dissertação (Mestrado) - Programa de Pós-Graduação em
Tecnologia Nuclear (Aplicações) -- Instituto de Pesquisas
Energéticas e Nucleares, São Paulo, 2020.

1. Ontologia. 2. Web Semântica. 3. e-Science. I. Olímpio
de Menezes, Mário, orient. II. Título.

AGRADECIMENTOS

Aos meus pais, Waldir Mauch de Carvalho e Marlene Aparecida Pereira de Carvalho, e a minha irmã Glaucia Mauch de Carvalho, com amor, admiração e gratidão por sua compreensão, por prestigiar e apoiar-me em cada conquista de minha vida e na concretização deste projeto.

Ao meu amigo Carlos Farias Felgueiras que contribuiu tirando dúvidas e dando sugestões valiosas.

Ao meu amigo Jurandir Alves de Sousa que me deu apoio e motivação para seguir em frente e não desanimar, bem como o meu amigo Rafael Silveira, que me auxiliou no decorrer do trabalho.

Aos meus amigos de sala que fizeram com que os dias fossem menos tensos, alegres e produtivos.

Ao Prof. Dr. Mário Olímpio de Menezes pela orientação cuidadosa, por meio de trocas de experiências e conhecimento que contribuíram para o desenvolvimento do projeto e, acima de tudo, por ter me acolhido e aceitado a idéia de realizar um projeto inovador no âmbito da Web Semântica.

Ao Instituto de Pesquisas Energéticas e Nucleares, IPEN, pela infraestrutura acadêmica, equipamento concedido, pela condução do programa de pós-graduação em nível de mestrado, possibilitando o acontecimento desta pesquisa.

*“Não vos amoldeis às estruturas deste mundo,
mas transformai-vos pela renovação da mente,
a fim de distinguir qual é a vontade de Deus:
o que é bom, o que Lhe é agradável, o que é perfeito.
(Bíblia Sagrada, Romanos 12, 2)*

RESUMO

Carvalho, Glauber M, **Desenvolvimento de uma Camada Semântica para um Protótipo de Repositório de Dados de Pesquisas de Análise por Ativação Neutrônica** 2020. 60 p. **Dissertação** (Mestrado em Tecnologia Nuclear) - Instituto de Pesquisas Energéticas e Nucleares - IPEN - CNEN/SP. São Paulo.

Um dos pilares da ciência é a possibilidade de reprodução dos resultados de pesquisas científicas por pesquisadores independentes, tornando possível a validação dos métodos, dos resultados e suas conclusões. Para que isto seja possível no cenário atual, onde a produção de dados científicos tomou uma proporção gigantesca (*Big Data*), métodos sistemáticos de armazenamento, curadoria e disponibilização dos dados precisam ser implementados. Diante do grande volume de dados disponíveis, todo o processo científico é impactado, possibilitando o surgimento de um novo paradigma científico: a e-Science ou e-Ciência. Neste trabalho foram desenvolvidos uma ontologia para o domínio dos dados da área de Análise por Ativação Neutrônica e de um protótipo de repositório de dados de pesquisa. A metodologia adotada para a construção da ontologia foram o Léxico Aplicado da Linguagem (LAL) e a Ontology Development 101 (Método 101). A integração da camada semântica com o SGBD PostgreSQL foi emulada através do Protégé e sua ferramenta Ontop, realizando uma conexão direta com o banco de dados relacional. Nesta emulação, consultas semânticas foram realizadas em SPARQL. Os resultados apresentados demonstram os ganhos oriundos desta integração e apontam para vantagens de se ter uma camada semântica em um repositório de dados, incluindo, mas não se limitando, à maior possibilidade de reuso dos dados visto que podem ser melhor entendidos a partir da ontologia que os descreve. O protótipo de repositório foi desenvolvido utilizando-se um *framework* de desenvolvimento Web de código aberto – Django, com a linguagem de programação Python, o que possibilitou bastante flexibilidade e agilidade neste processo.

Palavras-chaves: Ontologia, Web Semântica, e-Science

ABSTRACT

Carvalho, Glauber M, **Development of a Semantic Layer for a Prototype of Research Data Repository in Neutron Activation Analysis** 2020. 60 p. **Dissertação** (Mestrado em Tecnologia Nuclear) - Instituto de Pesquisa Energética e Nucleares - IPEN - CNEN/SP. São Paulo.

One of the pillars of Science is the possibility of replicating the results of scientific research by independent researchers, making possible the validation of methods, results and their conclusions. For this to be possible in the current scenario, where the production of scientific data grows fast (Big Data), systematic methods of data storing, curating and delivering need to be implemented. Faced with the avalanche of available data, the entire scientific process is impacted, enabling the emergence of a new scientific paradigm called e-Science. That new paradigm, called "The Fourth Paradigm of Scientific Exploration", distinguishes data-intensive science from traditional computational science. In this work, an ontology was developed for the domain of data in the area of Analysis by Neutron Activation and a prototype of research data repository. The integration of the semantic layer with the PostgreSQL SGBD was emulated through Protegè and its Ontop tool, making a direct connection to the relational database. In this emulation, semantic queries were performed in SPARQL. The results presented demonstrate the gains from that integration and point to several advantages of having a semantic layer in a data repository, including, but not limited to, a greater possibility of data reuse since these data can be better understood because of all description that the ontology provides. Also, as a result of these work, the methodology developed for the creation of the ontology is presented in details, allowing its application in other correlated domains in IPEN-CNEN/SP as well as in other institutions. The repository prototype was developed using the open source web development framework Django along with Python programming language, in a way that provided a lot of flexibility and agility to the process.

Keywords: Ontology, Web Semantic, e-Science

LISTA DE ILUSTRAÇÕES

Figura 1 – Pirâmide de Dados Científicos	15
Figura 2 – (a) Gestão de Dados científicos na ciência tradicional; (b) Gestão de Dados científicos na e-Science	16
Figura 3 – Hierarquia das Ontologias	25
Figura 4 – Espectro Semântico de Ontologias	25
Figura 5 – Perdas conceituais desde a percepção do mundo até a representação do conhecimento	26
Figura 6 – Tela de Interface do Protégé 5.5.0	32
Figura 7 – Dados gerados em cada etapa da AAN	37
Figura 8 – Diagrama de Caso de Uso da AAN	40
Figura 9 – Conceitualização da Ontologia	44
Figura 10 – Hierarquia de Classes	45
Figura 13 – Hierarquia da classe ObjetoAnalise	45
Figura 11 – Propriedades da OntoAAN	46
Figura 14 – Hierarquia da classe Amostra	46
Figura 12 – Atributos da OntoAAN	47
Figura 15 – Hierarquia da classe Aliquota	47
Figura 16 – Hierarquia da classe AliquotalIrradiada	48
Figura 17 – Hierarquia da classe Metodo	48
Figura 18 – Hierarquia da classe Contagem	48
Figura 19 – Hierarquia da classe Irradiacao	49
Figura 20 – Hierarquia da classe Equipamento	49
Figura 21 – Hierarquia da classe ReatorNuclear	49
Figura 22 – Hierarquia da classe Espectrometro	49
Figura 23 – Hierarquia da classe do PadraoNaolIrradiado	50
Figura 24 – Classe MaterialReferenciaCertificado	50
Figura 25 – Classe PadraoPipetado	50
Figura 26 – Classe Elemento	50
Figura 27 – Hierarquia da classe Elemento	51
Figura 28 – Hierarquia da classe MaterialReferenciaCertificadoIrradiado	51
Figura 29 – Hierarquia da classe PadraoPipetadoIrradiado	51
Figura 30 – Hierarquia da classe AnaliseEspectral	51
Figura 31 – Arquiteura Ontop com a camada semântica	52
Figura 32 – Mapeamento da Amostra no formato R2RML	53
Figura 33 – Consulta Sparql classe Amostra	54

Figura 34 – Classes do Protótipo	56
Figura 35 – Classe Amostra	57
Figura 36 – Interface da listagem das Amostras	57

LISTA DE TABELAS

Tabela 1 – Elementos do Dublin Core	33
Tabela 2 – Lista de Termos Candidatos	36
Tabela 3 – Regras de formação do LAL	38
Tabela 4 – Léxico da AAN	42
Tabela 5 – Propriedades do objeto	43

LISTA DE ABREVIATURAS E SIGLAS

IPEN	Instituto de Pesquisas Energéticas e Nucleares
CRPq	Centro do Reator Nuclear de Pesquisas
LAN	Laboratório de Análise por Ativação Neutrônica
AAN	Análise por Ativação com Nêutrons
IA	Inteligência Artificial
IEEE	Institute of Electrical and Electronics Engineers
SUMO	Standard Upper Merged Ontology
WEB	World Wide Web
RDF	Resource Description Framework
XML	eXtensible Markup Language
W3C	World Wide Web Consortium
HTML	HyperText Markup Language
DARPA	Defense Advanced Research Projects Agency
DAML	DARPA Agent Markup Language
OIL	Ontology Inference Layer
OWL	Ontology Web Language
SPARQL	Simple Protocol and RDF Query Language
SGDB	Sistemas Gerenciadores de Bancos de Dados
LAL	Léxico Ampliado da Linguagem
Udi	Universo de Informação

SUMÁRIO

1	Introdução	12
1.1	Paradigmas das Ciencia	12
1.2	Web Semântica	13
1.3	A necessidade da disponibilização dos dados científico	13
1.4	e-Science	14
1.5	Classificação dos dados científicos	15
1.6	Gestão de Ciclo de Vida de Dados de Pesquisas Científicas	16
1.7	A importância dos repositórios de dados semânticos	17
1.8	Princípio da Análise por Ativação com Nêutrons	17
1.9	Objetivos Gerais	20
1.10	Objetivo Específico	20
1.11	Justificativa	20
2	Fundamentos Teóricos	22
2.1	Repositório Digitais	22
2.2	Ontologia	23
2.2.1	Tipo de Ontologias	24
2.2.2	Engenharia de Ontologia	25
2.2.3	Metodologia de Desenvolvimento de Ontologia	26
3	Metodologia	29
3.1	Ferramentas e Tecnologias	29
3.1.1	Resource Description Framework (RDF)	29
3.1.2	Web Ontology Language (OWL)	31
3.2	Ferramentas para Desenvolvimento e Avaliação de Ontologias	31
3.2.1	Protégé	31
3.2.2	Dublin Core	32
3.2.3	Ontop	33
3.2.4	SPARQL	33
3.3	Ferramentas para Desenvolvimento do Protótipo	34
3.3.1	Python	34
3.3.2	Django	34
3.3.3	PostgreSQL	34
3.4	Fase I - Construção da Ontologia	35
3.4.1	Construção do Léxico	36
3.4.1.1	Levantamento de Informação	36
3.4.1.2	Modelagem do LAL	37

3.4.1.3	Análise do LAL	37
3.4.1.4	Mapeamento do LAL	37
3.5	Fase II - Construção do Protótipo do Repositório	38
3.5.1	Descrição dos atores	38
3.5.2	Requisitos Funcionais	39
3.5.3	Diagrama de Caso de Uso	39
4	Resultados e Discussão	41
4.1	Modelagem da Ontologia	41
4.1.1	Construção e Análise do LAL	41
4.1.2	Mapeamento do LAL	41
4.2	Construção da Ontologia usando o Framework Protégé	41
4.2.1	Conceitualização da OntoAAN	41
4.2.2	Identificador de Recurso Internacional (IRI)	42
4.2.3	Estrutura de classes e subclasses da OntoAAN	43
4.2.4	Relacionamentos da OntoAAN	43
4.2.5	Atributos da OntoAAN	44
4.2.6	OntoAAN	45
4.3	Mapeamento e consulta do Protégé com o PostgreSQL	52
4.4	Protótipo do Repositório de AAN	54
5	Conclusão	58
Referências	60

1 INTRODUÇÃO

A ciência pode ser compreendida como toda competência adquirida por meio de estudo, pesquisa ou do desenvolvimento prático corroborado em princípios baseados na reflexão, na observação e na experimentação, culminando em teorias que possibilitem serem elaboradas, aprimoradas, ou refutadas, a fim de que a quantidade e a qualidade da informação seja preservada (CHIBENI, 2013).

1.1 Paradigmas das Ciências

Segundo Kuhn (1962), os paradigmas são pressupostos científicos que permitiram a evolução da ciência, fomentando a criação de modelos de representação para a realização científica, possibilitando deparar-se com problemas e soluções modelares, para a sociedade científica. Atualmente os paradigmas da ciência estão balizados em 4 modelos, sendo estes:

- **Ciência Empírica**

Consiste em sistema artificial, ou natural, estabelecido pelas leis naturais que produzem fenômenos empíricos que podem ser analisados pelos métodos de observação bem como por experiência sensorial comuns a toda a ciência.

- **Ciência Teórica**

Faz-se menção às primeiras experiências em explicar fenômenos por meio de modelos teóricos, aos quais destacaram-se, as Leis de Kepler, as Leis de Newton e a Lei de Boyle-Mariotte que viabilizaram um melhor esclarecimento em relação aos fenômenos observados empiricamente como também conjecturaram o comportamento de novos fenômenos (CORDEIRO et al., 2013).

- **Ciência Computacional**

Baseia-se na complexidade das simulações que permitem observar a evolução de um fenômeno em relação ao tempo, a partir de um modelo previamente desenvolvido. O progresso deve-se ao fato de que a criação de modelos grandes e complexos, tornaram-se inviáveis para serem resolvidos por métodos analíticos. Esta tendência computacional teve como preâmbulo as atividades do Eniac para a realização de simulações de modelos balísticos e posteriormente sua atuação em diversas áreas do conhecimento. Com o advento de computadores mais rápidos e eficientes e concomitantemente ao progresso

da computação paralela e distribuída, o que viabilizou a outras ciências desenvolverem e testarem seus modelos numéricos. Consequentemente, estes avanços computacionais permitiram aferir simulações de modelos complexos, culminando na geração de grandes quantidades de dados (CORDEIRO et al., 2013).

- **Ciência Exploratória de Dados**

Se refere a métodos de obtenção de resultados científicos por meio de utilização de computação intensiva, usualmente paralela, e/ou grande volume de dados. A e-Science, está relacionada à descoberta e ao compartilhamento do conhecimento na forma de dados experimentais, vocabulários teóricos ricos, publicações e serviços reutilizáveis que sejam úteis à comunidade científica (MARQUES, 2014).

1.2 Web Semântica

O termo Web Semântica foi designado por Tim Berners-Lee, cujo principal objetivo foi permitir aos computadores pesquisarem, combinarem e processarem de forma inteligente os conteúdos da Web com base no significado que eles possuem para os seres humanos. Além disso, surgiu a ideia de estender o entendimento da informação para as máquinas, desta maneira, agentes de software conseguiriam realizar tarefas complexas no conteúdo da Internet, para isso, seriam suportados utilizando o conceito de metadados definido pelo W3C. No entanto, é necessário apresentar novos formalismos e linguagens para alcançar tais objetivos. A Lógica de Descrição torna-se uma ferramenta importante, raciocinador e representação de conhecimento.

1.3 A necessidade da disponibilização dos dados científico

A disponibilização dos dados subjacentes aos artigos científicos publicados, tem sido estabelecida como uma exigência por algumas revistas inovadoras no meio científico. Marques (2014), cita frase de pesquisadora da Unicamp: “O compartilhamento de dados para reutilização ou reprodução de experimentos exige conhecer sua origem e entender como foram produzidos, associando à informação, métodos, algoritmos ou técnicas adotadas, e ainda ter acesso ao software necessário para processá-los, o que torna o processo bastante complexo. Sem isso, pode não ser possível reproduzir o experimento original ou reutilizar o dado em uma outra pesquisa”.

A complexidade e a abundância de recursos de dados em um ambiente de e-Science requer suporte para a gestão do conhecimento e dos metadados, uma vez que dados são sabidamente difíceis de compartilhar, encontrar, acessar, interpretar e reutilizar (MARQUES, 2014).

No centro do desenvolvimento científico está a descoberta de novos conhecimentos; a geração, o suporte e a manutenção do conhecimento formam a base do desafio científico. Um dos pilares da ciência é a possibilidade de reprodução dos resultados de pesquisas científicas por pesquisadores independentes, tornando possível a validação dos métodos, dos resultados e suas conclusões. Para que isto seja possível no cenário atual, onde a produção de dados científicos tomou uma proporção gigantesca (*Big Data*), métodos sistemáticos de armazenamento, curadoria e disponibilização dos dados precisam ser implementados. Neste contexto, surge um novo paradigma científico: a e-Science, (ou e-Ciência), denominado “O quarto paradigma da exploração científica”, distingue a ciência de dados intensiva da ciência computacional tradicional (GRAY, 2009).

1.4 e-Science

Conforme definido e utilizado correntemente no Brasil (FAPESP Call for Research Proposals – e-Science), o termo e-Science (ou e-Ciência, no Brasil) se refere a métodos de obtenção de resultados científicos através da utilização de computação intensiva, usualmente paralela, e/ou grande volume de dados. "e-Science", em última instância, está relacionada à descoberta e ao compartilhamento do conhecimento na forma de dados experimentais, vocabulários teóricos ricos, publicações e serviços reutilizáveis que sejam úteis à comunidade científica.

De acordo com Marques (2014), a disponibilização dos dados subjacentes aos artigos científicos publicados, tem sido colocada como exigência por algumas revistas inovadoras no meio científico. Neste mesmo artigo, cita frase de pesquisadora da Unicamp: “O compartilhamento de dados para reutilização ou reprodução de experimentos exige conhecer sua origem e entender como foram produzidos, associando à informação, métodos, algoritmos ou técnicas adotados, e ainda ter acesso ao software necessário para processá-los, o que torna o processo bastante complexo. Sem isso, pode não ser possível reproduzir o experimento original ou reutilizar o dado em uma outra pesquisa”.

Figura 1 – Pirâmide de Dados Científicos



Fonte: Demchenko et al. (2015)

1.5 Classificação dos dados científicos

Os dados de pesquisa (ou dados científicos) podem ser classificados de acordo com suas propriedades de identificação e ligações e podem ser representados em uma pirâmide, conforme mostrado na figura 1.

- Dados Brutos (*Raw data*), são aqueles oriundos diretamente das observações, dos experimentos.
- Dados estruturados (*Structured data*) são os que já sofreram algum tipo de processamento, limpeza, para atender algum objetivo de pesquisa (modelo formal). Podem ser armazenados em repositórios de modo que todos os colaboradores tenham acesso.
- Dados publicados (*Published data*) são os que acompanham uma publicação (artigo, tese, etc) e que dão suporte às hipóteses ali exaradas. Algumas revistas aceitam (ou requerem) o envio de tais dados juntamente com o artigo já no momento do envio para publicação.
- Dados ligados e embutidos em publicações, não só acompanham a publicação, mas são parte integrante desta publicação, não sendo possível obter a publicação sem os dados embutidos.

Para garantir a reusabilidade dos dados, não é suficiente apenas sua disponibilização; informação acerca dos processos envolvidos na transformação dos dados crus até a geração dos dados publicados se torna um importante aspecto da gestão de dados científicos. A curadoria (que garante a proveniência)

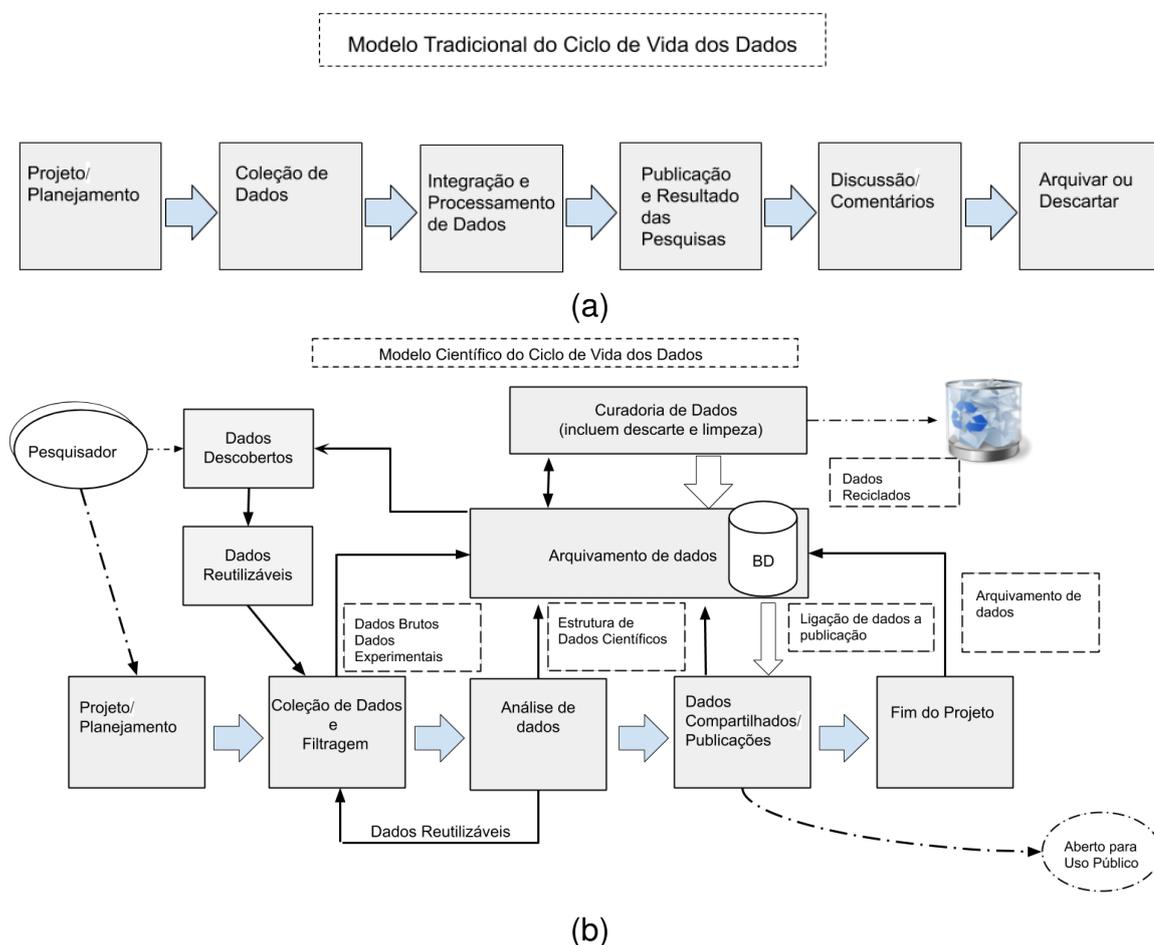
de dados científicos também precisa ser levado em conta neste processo de gestão.

A reusabilidade dos dados depende grandemente do quanto se entende destes dados. Ou seja, o entendimento da semântica dos dados publicados é um aspecto importante para garantir esta reusabilidade; este é um processo que nem sempre pode ser automatizado e, mesmo com a avalanche de dados na era do *Big Data*, a provisão de uma semântica clara se torna uma condição necessária para a reutilização eficiente de dados publicados.

1.6 Gestão de Ciclo de Vida de Dados de Pesquisas Científicas

Dentro deste novo contexto da e-Science, identifica-se um novo modelo de Gestão de Ciclo de Vida de Dados de Pesquisas Científicas (*Data Life Cycle Management in Scientific Research – SDLM*). Este novo modelo é apresentado na figura 2, juntamente com o modelo tradicional de gestão de dados científicos pré-e-Science.

Figura 2 – (a) Gestão de Dados científicos na ciência tradicional; (b) Gestão de Dados científicos na e-Science



Fonte: Demchenko et al. (2015)

1.7 A importância dos repositórios de dados semânticos

Para fazer frente a este novo cenário, um conjunto complexo e amplo de conhecimentos, sistemas, métodos e tecnologias estão envolvidos. Modelos matemáticos, repositórios digitais e gerenciamento de dados, novos hardwares, softwares, protocolos, ferramentas e serviços, são alguns dos itens necessários para se atender às demandas deste novo paradigma da ciência.

Um meio de compartilhar dados na incumbência de gerar conhecimento é através dos repositórios de dados. No entanto, estes tendem a ter alcance limitado devido à falta de padrões de formato de dados e à falta de protocolos padronizados de acesso a dados, o que pode resultar em problemas para a descoberta e reutilização de dados para humanos e sistemas de computadores.

Para fazer frente a este novo cenário, um conjunto complexo e amplo de conhecimentos, sistemas, métodos e tecnologias estão envolvidos. Modelos matemáticos, repositórios digitais e gerenciamento de dados, novos hardwares, softwares, protocolos, ferramentas e serviços, são alguns dos itens necessários para se atender às demandas deste novo paradigma da ciência.

Devido a dificuldade dos repositórios de dados tradicionais em trabalhar com dados não estruturados, estes deram espaço as tecnologias semânticas que estão cada vez mais ganhando espaço na e-Science. Metodologias, ferramentas e outros componentes semânticos (ontologias) estão sendo cada vez mais utilizados. Modelagem de conhecimento, verificação de hipóteses baseada em lógica, integração de dados semânticos, composição de aplicativos e descoberta de conhecimento e análise de dados integrada para diferentes domínios de conhecimento estão cada vez mais presentes nesses contextos.

1.8 Princípio da Análise por Ativação com Nêutrons

No presente trabalho, foi feito um extenso estudo acerca do Método de Análise por Ativação com Nêutrons nos laboratórios do CRPq-Ipen para a obtenção dos dados gerados nas pesquisas que fazem uso do método.

A Análise por Ativação com Nêutrons (AAN) originou-se com os trabalhos de Hevesy e Levi em 1936, mas não se tornou um método prático de análise até o desenvolvimento do reator nuclear como uma fonte de nêutrons de alta intensidade. A introdução dos detectores de raios gama por cintilação NaI(Tl) e o de alta resolução Ge(Li) estabeleceram a AAN como uma técnica analítica sensível e de boa precisão e exatidão, útil para análises multielementares quali e quantitativas dos elementos principais, elementos menores e elementos

traço, em amostras dos mais variados campos de aplicação (MOREIRA, 2002; CATHARINO, 2002).

Na maioria dos métodos de AAN, o nuclídeo produto é radioativo e a atividade induzida, A_t (em Bq, ou desintegrações por segundo) no final da irradiação de tempo t_i em segundos está relacionada ao número de átomos no nuclídeo alvo, N (portanto com a massa do elemento), a probabilidade de reação ou seção de choque, σ (em cm^2), o fluxo de nêutrons, ϕ (em $cm^{-2}s^{-1}$) e a constante de decaimento, λ (em s^{-1}). Esta atividade é expressa pela equação 1.1:

$$A_t = N \sigma \phi (1 - e^{-\lambda t_i}) \quad (1.1)$$

A atividade máxima induzida, A_{max} , obtida para um tempo de irradiação muito maior do que a meia-vida do nuclídeo produto (na prática, cerca de sete meias-vidas), é dada por:

$$A_{max} = N \sigma \phi \quad (1.2)$$

Na AAN, utilizamos as energias características dos raios gamma para analisar, algumas vezes em níveis de “traço”, a composição de amostras desconhecidas (KNOLL, 1989). Cada fóton de raio gama tem uma energia discreta e esta energia é característica do isótopo fonte. Isto forma a base da espectrometria de raios gama. Através da medição das energias dos fótons de raios gama, pode-se determinar a fonte da radiação (IAEA, IAEA-TECDOC-1363, July 2003).

O espectro de energia dos raios gama representa a distribuição de energia dos fótons emitidos na fonte. Cada linha do espectro (ou espectro de “emissão”) mostra a energia e a intensidade relativa das emissões de raios gama nas séries de decaimento nuclear.

A maioria dos laboratórios de AAN utiliza detectores de Ge hiperpuro. Este detector opera à temperatura do nitrogênio líquido (77 K); o cristal de germânio é montado no interior de um recipiente a vácuo, montado sobre um criostato, conectado termicamente por um tubo de cobre, o “dedo frio”. O tipo mais comum de detector de Ge hiperpuro é o detector coaxial, que é usado para medir raios gama na faixa de energia de 60 keV a 3 MeV. Levando-se em conta a eficiência do detector ε , a atividade induzida A_t em um elemento após determinado tempo de irradiação t_i , que pode ser medida por detector adequado, é dada pela seguinte equação 1.3:

$$A_t = \frac{\varepsilon \sigma \phi N_A m f (1 - e^{-\lambda t_i})}{M} \quad (1.3)$$

onde:

ε eficiência do detector;

N_A Número de Avogadro;

m massa do elemento, em gramas;

f abundância isotópica do nuclídeo alvo;

M massa atômica do elemento;

No Laboratório de Ativação Neutrônica do CERPq, o método AAN empregado corriqueiramente é o método comparativo, no qual a amostra desconhecida tem sua massa determinada e é irradiada simultaneamente com um padrão (material de concentração conhecida do elemento a ser determinado), e após determinado período de tempo, a intensidade e energia dos picos de raios gama são medidos. A comparação entre as atividades específicas induzidas nos padrões e nas amostras desconhecidas é a base para o cálculo da concentração do elemento na amostra. Estas atividades resultantes, A_a e A_p , respectivamente, são expressas pelas equações 1.4 e 1.5:

$$A_a = \frac{\varepsilon \sigma \phi N_A m_a f (1 - e^{-\lambda t_i})}{M} \quad (1.4)$$

$$A_p = \frac{\varepsilon \sigma \phi N_A m_p f (1 - e^{-\lambda t_i})}{M} \quad (1.5)$$

onde os índices a e p se referem a amostra e o padrão, respectivamente.

A razão entre as equações 1.4 e 1.5 resulta em:

$$\frac{A_a}{A_p} = \frac{m_a}{m_p} \quad (1.6)$$

A concentração C_a do elemento na amostra é, então, expressa por:

$$C_a = \frac{A_a C_p m_p e^{\lambda(t_a - t_p)}}{A_p m_a} \quad (1.7)$$

onde:

C_p é a concentração do elemento no padrão;

t_a é o tempo de decaimento da amostra;

t_p é o tempo de decaimento do padrão.

O termo exponencial na equação 1.7 leva em consideração o decaimento da amostra e do padrão, uma vez que estes não são medidos simultaneamente.

1.9 Objetivos Gerais

O intuito deste estudo foi desenvolver uma camada semântica, através de uma Ontologia de Dados e criar um protótipo funcional de Repositório Semântico de dados para as pesquisas em análise por ativação neutrônica desenvolvidas no Centro do Reator de Pesquisas - CERPq, do IPEN-CNEN/SP, que abrangem diversas áreas de aplicação, tais como dados relacionados a amostras ambientais, coletadas com indicação de geolocalização a fim de disponibilizar os dados, após o processo de curadoria, no repositório semântico de e-Science do IPEN.

1.10 Objetivo Específico

- Construção da ontologia de AAN
 - Conceitualização da Ontologia
 - Construção das Classes
 - Construção das Propriedades
- Desenvolvimento do protótipo do repositório de dados.
- Desenvolvimento da camada semântica - integração ontologia com o PostgreSQL usando o plugin Ontop no editor Protégé.

1.11 Justificativa

Diante o exposto, foi construído um protótipo de repositório semântico a qual faz uso de uma ontologia de domínio denominada OntoAANI, o protótipo do repositório de dados terá a função de auxiliar no compartilhamento e na gestão de dados gerados pelas pesquisas Análise por Ativação Neutrônica dentro dos laboratórios do CRPq.

Desta descrição da técnica de Análise por Ativação Neutrônica podemos perceber que para a obtenção dos resultados de uma pesquisa, isto é, a análise de uma amostra, vários dados devem ser obtidos ou utilizados durante o processo. Infelizmente, muitos destes dados das pesquisas *desaparecem* após a publicação de um artigo e muitos não entrarão na versão final da tese ou dissertação do aluno. E com isso, perde-se a oportunidade de reutilização de dados tão importantes, ou que foram obtidos a um alto custo, e que poderiam agregar em outras pesquisas na mesma área ou em áreas diferentes.

Quando consideramos o IPEN e todos os seus Centros de Pesquisa, com seus diversos grupos, pesquisadores, estudantes, etc., este problema de perda de

dados de pesquisa é grandemente amplificado. Tomando-se apenas a análise por ativação neutrônica, cujas aplicações estendem-se por diversas áreas, como meio ambiente, arqueologia, forênsica, nutrição e medicina, entre outras, verificamos a grande necessidade de que seus dados de pesquisa sejam corretamente armazenados, curados, processados e disponibilizados à comunidade científica e à sociedade em geral.

Diante deste cenário de oportunidades, o IPEN lançou o projeto e-Science, que visa a criação de uma infraestrutura que possa embarcar essa nova metodologia e tecnologia e que permita a implantação de fluxos de trabalho, processos e bancos de dados buscando solucionar o que hoje se transformou em um importante gargalo da pesquisa científica: o desafio de organizar, classificar, descrever corretamente, selecionar, compartilhar e garantir acesso ao gigantesco volume de dados gerados nos últimos anos em todas as áreas supracitadas, e cruzar dados e variáveis diversas e possibilitar análises mais abrangentes que as atuais inerentes à especificidade de cada área.

2 FUNDAMENTOS TEÓRICOS

Neste tópico encontra-se os conceitos que fundamentam o projeto do protótipo do repositório semântico para o laboratório de análises por ativação com nêutrons.

2.1 Repositório Digitais

A tecnologia da informação e comunicação têm impactado positivamente não somente na gestão empresarial, tanto nas empresas públicas quanto nas privadas, mas também na educação e em outros setores, haja visto o crescente número de publicações científicas nos últimos anos. Para Almeida e Bax (2003) é imprescindível fazer uso de técnicas com o objetivo de melhorias na manutenção e organização deste enorme volume de dados, para proporcionar a recuperação de informação, ou para identificar padrões ou correlações entre diferentes dados que de alguma forma gere algum tipo de conhecimento ou benefício. Afim de solver este problema mostram - se como recurso os repositórios digitais (RDs), este são produtos que ofertam acesso a documentos e metadados, com recursos para descrição dos documentos através de metadados e busca para recuperá-los ??).

A disponibilização de repositórios digitais destinados à gestão da informação científica e acadêmica, cujo objetivo consiste em preservação da produção científica, por meio do pleno funcionamento de um conjunto de sistemas interligados: base de dados, interface, comunicação com servidor, indexação e estatísticas ??), aos quais sejam capazes de armazenar arquivos de diversos tipos de formatos, objetivando a difusão dos trabalhos científicos. Porém o gerenciamento destas dos repositórios tem-se mostrado um enorme desafio, uma vez que percebeu-se a necessidade de elaborar meios e ferramentas de pesquisas de aplicações genéricas que produzam resultados satisfatórios na busca de informação o que inclui verificação e análise de documentos.

Surge neste contexto a ontologia como um instrumento corroborativo, um conjunto de primitivas de representação que cria um modelo (GRUBER, 1993); podendo assumir vários formatos, no entanto, deve apresentar um vocabulário específico de termos e especificações de seu significado, bem como definições e indicações de como os conceitos estão interrelacionados; resultando em estrutura de domínio e restrições de possíveis interpretações de seus termos. A ontologia atualmente está sendo utilizada em setores específicos, tais como: indústria, inteligência

artificial e principalmente na Web, pois são mais eficientes que os tradicionais banco de dados relacionais no que diz respeito à categorização, modelagem, inserção de regras e padrões na codificação do conhecimento.

2.2 Ontologia

O conceito de ontologia surgiu na Filosofia (Metafísica) com Aristóteles, e compreendia o estudo da natureza do “ser” e a “existência” visando explicar todas as coisas do mundo de forma conceitual. Para a computação, as ontologias permitem modelar formalmente uma estrutura de um sistema, evitando interpretações ambíguas, devido ao vocabulário controlado fornecido para representar o conhecimento, que é, de alguma forma, conceitualizado. Para Gruber (1993), uma área de conhecimento formalmente representada baseia-se numa conceitualização, isto é, os objetos, conceitos e outras entidades que se supõe existir em alguma área; esta definição resume o grande interesse por esta área de conhecimento, pois ela fornece uma estrutura conceitual comum sobre a qual é possível gerar bases de conhecimentos compartilháveis e reutilizáveis, bem como prover interoperabilidade e ajuntamento das informações. Conforme indicado por Bontcheva et al. (2006) pode-se afirmar formalmente que uma ontologia é um conjunto de relacionamentos entre quatro elementos expressamente representados por $O = C, R, I, A$, sendo que:

- C:** é o conjunto de classes que representam os conceitos em um domínio de interesse.
- R:** é o conjunto de relações entre os conceitos do domínio.
- I:** é o conjunto de instâncias provenientes das classes.
- A:** é o conjunto de axiomas do domínio, são responsáveis por restringir e inferir regras.

Os elementos apresentados formam a estrutura necessária para o desenvolvimento da representação do conhecimento. O uso das ontologias é de suma importância quando se pretende trabalhar com a representatividade de conhecimento na Web, uma vez que, elas fazem parte da estrutura da Web Semântica, um outro aspecto também importante é a capacidade das ontologias conseguirem trabalhar com aplicações que fazem uso de grande quantidade de dados de maneira extremamente eficiente (MCGUINNESS, 2004; HORROCKS; SATTLER; TOBIES, 2000). As ontologias oferecem apoio para solucionar diversos problemas que os

cercam ao desenvolver aplicações com base em representação formal, ou seja, bases de conhecimento. Dentre as quais elas visam resolver, podem ser citadas:

- Premissas básicas são ignoradas nas bases de dados, culminando na impossibilidade de compartilhar o conhecimento representado.
- Ausência de modelos genéricos aos quais possam ser construídas bases de dados e aplicações de forma simplificada.

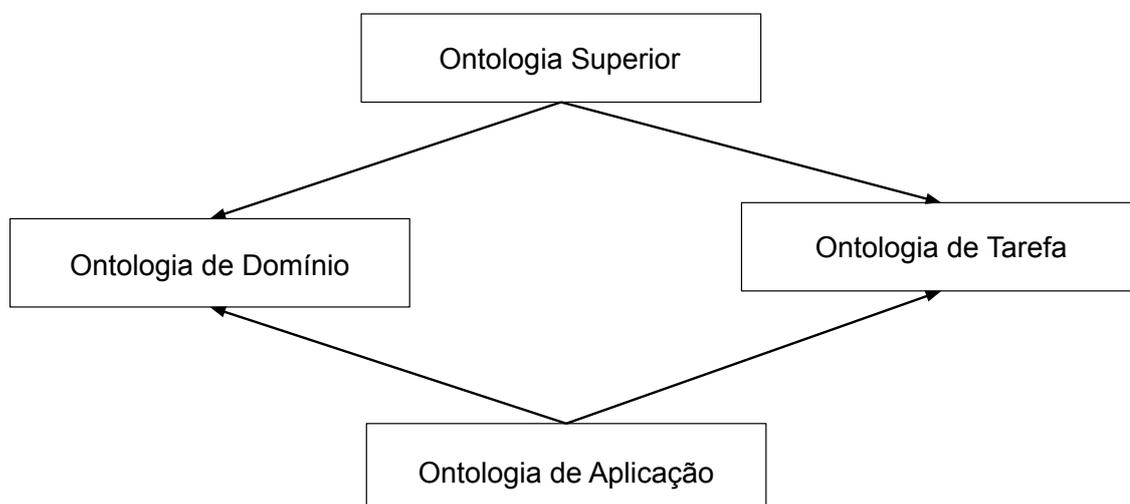
Portanto, os benefícios das ontologias são enormes, principalmente quando os dados publicados na Web são checados automaticamente ao usar conjuntos de regras de inferências que validem a consistência dos mesmos e por oferecer uma anuência quanto a criar modelo de dados e a importância que é proporcionada para a compreensão e a utilização dos dados.

2.2.1 Tipo de Ontologias

No sistema de classificação proposto por Guarino (1998) o critério é a generalização da ontologia. Neste sistema, uma ontologia pode ser classificada em termos de o quão próximo ao modelo ela representa o conhecimento e do quão precisa é essa representação, conforme apresentado na figura 3.

- Ontologias de nível superior
Descrevem os conceitos mais genéricos, normalmente independentes de domínio e podem ser reutilizados na confecção de novas ontologias.
- Ontologias de domínio
Tratam do vocabulário de um domínio específico, onde os conceitos são especializados a partir da ontologia de nível superior.
- Ontologias de tarefa
Tratam da descrição do vocabulário de uma tarefa, ainda de modo genérico; também é construída a partir da especialização de conceitos presentes na ontologia de nível superior
- Ontologias de aplicação
São as ontologias mais específicas; os conceitos representados nestas ontologias estão relacionados a entidades do domínio, são os papéis desempenhados por estas entidades em algum processo ou tarefa

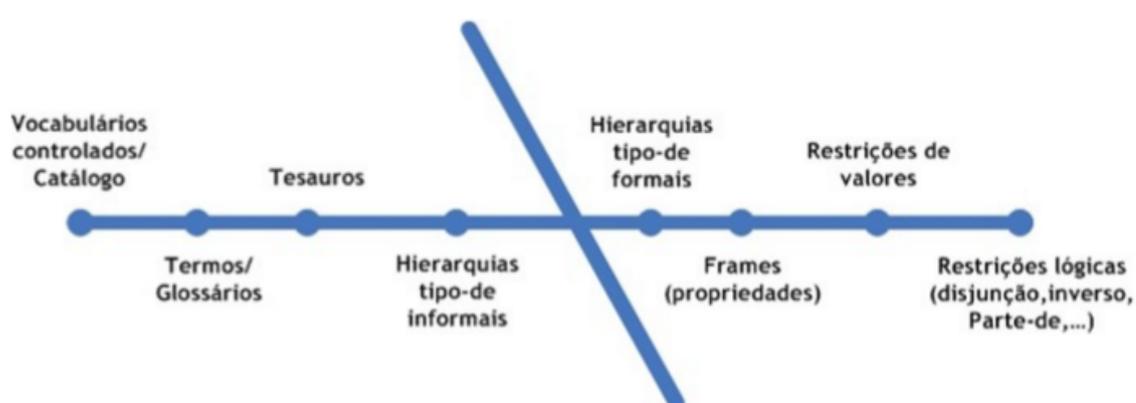
Figura 3 – Hierarquia das Ontologias



Fonte: Guarino (1998)

No entanto, para Lassila e McGuinness (2001) a proposta de classificação segue uma linha que parte da mais simples para a mais complexa. Esta proposta de classificação está ilustrada na figura 4.

Figura 4 – Espectro Semântico de Ontologias



Fonte: Lassila e McGuinness (2001)

2.2.2 Engenharia de Ontologia

Com a evolução da Engenharia de Conhecimento, surge a Engenharia de Ontologias responsável por elaborar estudos e tecnologias para a geração de

bases de conhecimento para sistemas especialistas (ISOTANI; BITTENCOURT, 2015). Fatores como a complexidade, requisitos e restrições relativas ao processo de desenvolvimento da ontologia contribuíram para a necessidade de empregar metodologias que definem um processo a ser adotado ao longo da criação eficiente da ontologia.

2.2.3 Metodologia de Desenvolvimento de Ontologia

Uma metodologia pode ser compreendida como um conjunto de métodos e técnicas integradas que visam construir um arcabouço teórico para sustentar o trabalho intelectual intensivo, a fim de auxiliar no processo de elaboração de uma ontologia, tornando-se fundamental para que haja uma conformidade desta ontologia com o mundo real. No entanto, da observação do mundo real para a conceitualização, ocorrem perdas conceituais, da mesma forma que da conceitualização para a construção da ontologia, ocorrem perdas devido à limitação da própria linguagem utilizada (MCGUINNESS, 2004; HORROCKS; SATTLER; TOBIES, 2000). A figura 5 ilustra estas perdas conceituais.

Figura 5 – Perdas conceituais desde a percepção do mundo até a representação do conhecimento



Fonte: Isotani e Bittencourt (2015)

Segundo Calero Coral; Ruiz (2006), o ciclo de vida de uma ontologia pode ser entendido como constituído das seguintes etapas:

- Especificação: Estabelece o início das atividades, indica o motivo pela qual a ontologia será construída e quais usuários farão uso.
- Conceitualização: Organiza e estrutura o domínio de conhecimento a qual deseja que a ontologia deva representar.

- Formalização: Construção do modelo conceitual, a qual se estabelece os conceitos, as relações e os axiomas.
- Implementação: Construção da ontologia usando alguma ferramenta de edição, ou alguma linguagem específica para o desenvolvimento de ontologia.
- Manutenção: Após a construção da ontologia é necessário que ocorra a manutenção da mesma sempre que necessário.

Segundo (CASELLAS, 2011), as metodologias de ontologias podem ser divididas em três grupos:

- Metodologias Inspiradas na Inteligência Artificial: Nestas metodologias, os processos estão mais próximos ao processo de aquisição de conhecimento. Algumas metodologias são TOVE, Enterprise Model Approaches e IDEF5.
- Metodologias Inspiradas na Engenharia de Software: Nestas metodologias, os processos são inspirados no processo de desenvolvimento de software. Este tipo de metodologia advém de modelos tradicionais de desenvolvimento de software como o RUP (Rational Unified Process). Alguns exemplos de metodologias são: METHODOLOGY, Ontology Development 101 (Método 101) e RapidOWL.
- Metodologias Inspiradas na Interação Humano-Computador: estas metodologias têm o foco voltado para o usuário final. Alguns exemplos de metodologias são: DILIGENT e HCOME.

Além destas metodologias citadas por Casellas, Breitman (2005) apresenta o Léxico Aplicado da Linguagem (LAL), o LAL é um método para auxiliar na elicitação de identificação de símbolos próprios do domínio de conhecimento ao qual deseja representar. A identificação destes símbolos e de seus significados ajuda na compreensão do Universo de Informação (Udi) da aplicação. As descrições dos símbolos no LAL são definidas através da noção e impacto de seus conceitos no domínio da aplicação. De acordo com a classificação metodologias de Casellas a metodologia do LAL pode ser inserida dentro do grupo de Metodologias Inspiradas na Engenharia de Software, uma vez que o LAL esta fortemente acoplado com Engenharia de Requisitos.

O Método 101, consiste na idéia de que o desenvolvimento de ontologias deve ser um processo interativo, a fim de delimitar o “como fazer” de algumas

atividades, como exemplo, pode-se fazer inferência à conceituação, cujos passos são criteriosamente pormenorizados.

Para (NOY; MCGUINNESS, 2001) os passos referentes a esta metodologia estão balizados em: “Determinar o domínio e o escopo da ontologia; Considerar o reuso de ontologias existentes; Enumerar termos importantes para a ontologia; Definir as classes e suas respectivas hierarquias; Definir as propriedades de classes; Definir os valores das propriedades; Criar circunstâncias.” . Segundo (NOY; MCGUINNESS, 2001), o primeiro passo no processo de construção de uma ontologia está fundamentado em minimizar o foco de interesse para um domínio específico, na definição do uso e dos usuários da ontologia. Em situações que envolvam ontologias de nível superior, faz-se necessário maior atenção uma vez que possibilita realizar a ligação de várias ontologias de um domínio específico por meio de algum conceito e relação.

Um expõente desta metodologia é Wine Ontology (ontologia do vinho) desenvolvida e demonstrada passo a passo pelos autores da metodologia. Outros exemplos de ontologia análisadas que foram analisadas: foaf, travel.owl, koala.owl.

3 METODOLOGIA

Neste capítulo são descritos as ferramentas e os processos empregados, que foram divididos em duas etapas;

- Fase 1- Refere-se à construção da Ontologia, resultando na confecção de uma ontologia para o domínio da AAN (Análise por Ativação Neutrônica) que é feita no LAN-CPRq.
- Fase 2 - Construção do Protótipo do Repositório de Dados, que é uma atividade de desenvolvimento de software.

3.1 Ferramentas e Tecnologias

Segundo (STAAB; STUDER, 2007), o processo de representação relativo às metodologias, possibilita a construção e avaliação da ontologia. Logo se faz necessário o uso de ferramentas que auxiliem na construção de ontologias, já que, o próprio ambiente de engenharia de ontologia fornece meios para extrair módulos das ontologias existentes, mantendo intactos os vínculos lógicos. Entre as ferramentas utilizadas no desenvolvimento de ontologias, estão o OntoEdit, a WebOnto, a WebOde e o Protégé.

No aspecto que envolve a avaliação, verificação e manipulação (por meio de regras e axiomas para a aquisição de conhecimento) de ontologias, tem-se as chamadas máquinas de inferência (reasoners), como por exemplo, Fact, Ontop e Pellet, Hemmit.

Outro componente tão importante quanto as ferramentas de criação é a linguagem de consulta, pois esta permite gerar o conhecimento por meio de consultas sobre o domínio de conhecimento concebido, no caso, para a OntoAAN foi usada a linguagem SPARQL.

Neste tópico, estão as tecnologias para o desenvolvimento do projeto de OntoAAN.

3.1.1 Resource Description Framework (RDF)

O Resource Description Framework (RDF), é uma linguagem baseada em XML que possui um conjunto de atribuições que possibilita codificar, adicionar e representar metadados (dados sobre dados) para recursos disponíveis na Web

(MILLER, 1998). A Extensible Markup Language (XML), também permite que agentes computacionais possam processá-los de forma automatizada a RDF, além disso possibilita representar relações entre objetos.

Segundo (BRAATZ; BRANDT, 2008), a RDF permite representar relações entre os objetos, são dados por um conjunto de instruções no qual contém uma tripla, constituída por um sujeito (recurso que é identificado URI), predicado (relacionamento entre um recurso e um objeto, ou uma propriedade do sujeito) e objeto (dependendo do predicado, pode ser um outro recurso ou um valor para a propriedade do sujeito), no qual definem um modelo simples para descrever as inter-relações entre recursos em termos de propriedades e valores nomeados, que também podem ser URIs, e consiste em um formato de dado baseado em grafos direcionados.

As URI podem ser classificados como:

- URL (Uniform Resource Locator), define um localizador/endereço para um determinado recurso, a partir de um protocolo existente (ISOTANI; BITTENCOURT, 2015).
- URN (Unified Resource Name), representa um nome para um determinado recurso, afim de, garantir a unicidade de forma global de um recurso, mesmo quando este não está disponível. Já IRI I (International Resource Identifier) consiste na generalização do URI, a diferença entre eles é que o URI, que é baseado nos caracteres ASCII, enquanto o IRI amplia o número de caracteres (ISOTANI; BITTENCOURT, 2015).

O RDF, se torna limitado quando não é capaz de prover mecanismo para declarar propriedades, esta função é designada ao RDF Schema (esquema RDF), uma vez que possibilita definir atributos sobre os objetos. Um esquema RDF define não apenas as propriedades dos objetos, mas também pode definir os tipos de recursos que estão sendo descritos.

A partir da capacidade do XML de definir esquemas de marcação personalizados e na abordagem flexível do RDF para representar dados, foram criadas diversas linguagens para representação de ontologias tais como a DARPA Agent Markup Language (DAML), Ontology Inference Layer (OIL), e a Web Ontology Language (OWL), esta foi a base para a construção da ontologia de AAN.

3.1.2 Web Ontology Language (OWL)

A Web Ontology Language (OWL) é uma linguagem de marcação semântica usada para compartilhar e publicar ontologias na Web, foi desenvolvida com o propósito de ser uma extensão do Resource Description Framework (RDF).

De acordo com (FRANK VAN et al., 2003), a sua construção está fundamentada à linguagem DAML + OIL cuja finalidade consiste na representação de conhecimento com base em ontologias, dessa forma, fornecendo meios para que aplicações possam processar as informações, otimizando a capacidade de processamento do conteúdo da Web por agentes de software, uma vez que o XML, RDF e RDF Schema (RDF-S) são suportados pelo OWL, isso gera maior interpretabilidade das aplicações que irão fazer uso do conteúdo da Web.

Conforme a W3C, mencionado por (SCHREIBER; DEAN, 2004), a linguagem OWL possuem três sub-linguagens (ou dialetos): OWL-Lite, OWL-DL e OWL-Full, que são descritas a seguir:

- OWL-Lite - suporta apenas hierarquia de classificação e restrições simples. Ela possui uma complexidade formal menor que o OWL DL;
- OWL-DL - corresponde a lógica de descrição, sendo assim, são suscetíveis a raciocínio automático, contribuindo para a verificação de inconsistências na ontologia.
- OWL-Full - sub linguagem mais expressiva do que as outras, no entanto, a preocupação maior se dá por conta de uma maior expressividade do que capacidade de decisão.

O emprego desta linguagem OWL, na construção da ontologia de AAN deu-se por conter a estrutura necessária para a descrição da mesma.

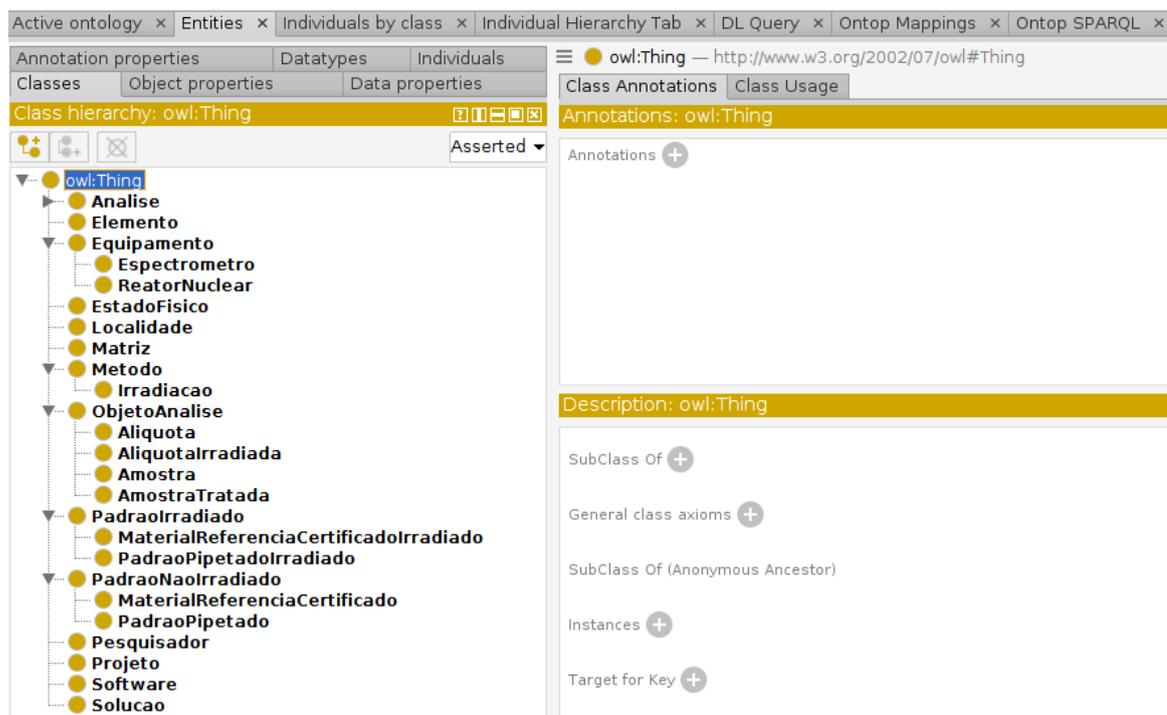
3.2 Ferramentas para Desenvolvimento e Avaliação de Ontologias

Neste tópico, serão apresentados os elementos essenciais para a construção da OntoAAN, sendo eles: editor Protégé, o raciocinador Ontop e a linguagem de consulta SPARQL.

3.2.1 Protégé

O Protégé é um editor de ontologias de código aberto que fornece um conjunto de ferramentas que auxiliam na construção de sistemas baseados em co-

Figura 6 – Tela de Interface do Protégé 5.5.0



Fonte: autor

nhecimento, pois possui uma farta coleção de estruturas que permitem a modelagem, visualização e manipulação de ontologias em diversos formatos de representação.

Conforme indicado por Yu (2011), o Protégé pode ser descrito como uma coleção de plug-in, pois sua arquitetura viabiliza a sua extensão, podendo ser utilizada para alterar e estender o comportamento do Protégé, bem como, substituí-los individualmente ou como um todo a fim de modificar completamente a interface e o comportamento do Protégé. A versão do Protégé usada na construção da OntoAAN foi a 5.5.0, a qual tem a interface mostrada na figura 10.

Na parte superior, do lado esquerdo, são exibidas as abas para visualização das principais janelas do Protégé dentre as quais podemos destacar (Class, Object properties e Data properties). Os plug-in Ontop Mapping e Ontop SPARQL foram adicionados, o primeiro permite definir o mapeamento entre a ontologia e um banco de dados relacional e o segundo permite fazer as inferências dentro da ontologia.

3.2.2 Dublin Core

O Dublin Core é um padrão criado com o objetivo de padronização de metadados e ao desenvolvimento de vocabulário especializado com o intuito de descrever os dados, afim de, viabilizar que sistemas inteligentes possam recuperar informações. O Dublin Core é atualmente um padrão ANSI(ANSI/NISO Z39.85) e

norma ISO(ISO Standard 15836-2003 - fevereiro de 2003).

Os elementos que compõem o padrão Dublin Core são:

Tabela 1 – Elementos do Dublin Core

Elemento	Definição
Assunto (subject)	O tópico abordado pelo trabalho
Título (title)	Nome do objeto
Criador (creator)	Responsável(eis) pelo conteúdo intelectual do objeto
Descrição (description)	Descrição do conteúdo do objeto
Editor (publisher)	Agente ou agência responsável por disponibilizar o objeto
Outro agentes (contributor)	Pessoas que fez (fizeram) contribuições significativas para o objeto.
Data (date)	Data da publicação
Tipo de Objeto (type)	Gênero do objeto, se ficção, novela, poema ou dicionário
Formato (format)	Manifestação física do objeto
Identificador (identifier)	Cadeia ou número utilizado para identificar unicamente aquele objeto
Relacionamento (relation)	Relacionamento com outros objetos
Fonte (source)	outros objetos, eletrônicos ou físicos, dos quais estes foi derivado (caso seja aplicável)
Linguagem (language)	Linguagem do conteúdo intelectual
Cobertura (coverage)	Matriz
Direitos (rights)	Informação sobre os direitos acerca do objeto

3.2.3 Ontop

Segundo (CALVANESE et al., 2015), Ontop é definido como um sistema de acesso a dados baseado em ontologia que permite aos usuários acessar um banco de dados relacional através de uma camada conceitual fornecida por uma ontologia, trata-se de um projeto de código aberto que pode ser usado como: (i) um plug-in para o Protégé, fornecendo uma interface gráfica para edição de mapeamento, execução de consultas SPARQL e verificação de inconsistências. Além disso, o Ontop consegue se conectar com os principais bancos de dados relacionais (PostgreSQL, MySQL, H2, DB2, ORACLE, HSQL e MS SQL Server).

3.2.4 SPARQL

SPARQL (pronunciado como sparkle) denota o Protocolo SPARQL e o RDF QueryLanguage, recomendado pelo World Wide Web Consortium (W3C). O

núcleo do SPARQL é consultas simples na forma de padrões gráficos simples, criados para recuperar e manipular dados no formato RDF. Os dados armazenados em triplas podem ser manipulados pela linguagem SPARQL através dos Endpoints SPARQL, assim como os dados armazenados em bancos de dados relacionais podem ser tratados pela linguagem Structured Query Language (SQL).

3.3 Ferramentas para Desenvolvimento do Protótipo

Neste tópico, estão as tecnologias para o desenvolvimento do protótipo do repositório.

3.3.1 Python

Python é uma linguagem de alto nível, orientada a objeto, tipagem dinâmica e forte, interpretada, interativa, de fácil aprendizado e de código aberto.

A linguagem está se popularizando muito atualmente, pois a produtividade e expressividade aliadas com uma estrutura de dados eficiente, bem como uma abordagem simples porém eficiente de orientação objeto, fizeram com que a linguagem encaixasse de forma satisfatória com as metodologias ágeis Borges (2014).

O Python além de ser usado como a linguagem principal de desenvolvimento, é uma linguagem que também permite criar scripts para automatização de tarefas e adição de novas funcionalidades, bem como, trabalhar facilmente com aplicações de terceiros, tais como: BrOffice, PostgreSQL, GIMP, entre outros Borges (2014).

3.3.2 Django

O Django é um framework para a construção de aplicações Web em Python gratuito e de código aberto. Sua arquitetura consiste no que é denominado pelos desenvolvedores por MTV - isto é: Model-Template-View. A camada View é a forma de processar dados de uma URL, é ela quem descreve qual informação é apresentada, enquanto que na camada Model, é onde está disposta toda a lógica de acesso aos dados. A camada Template por sua vez, é responsável pela forma como as informações são apresentadas Romano, Hillar e Ravindran (2018).

3.3.3 PostgreSQL

Tecnologias de banco de dados dão suporte às empresas no que tange manter os dados, sendo que, a quantidade de dados tende a aumentar com o cres-

cimento das empresas, o que torna inevitável fazer uso de algum tipo de sistema gerenciador de banco de dados (SGBD), pois a principal função destes tipos de sistemas é o de manter a integridade dos dados quando existe alteração nas estruturas das tabelas (CARVALHO, 2017).

O PostgreSQL é um SGBD extremamente eficiente de objeto-relacional de código aberto . O PostgreSQL tem como características (CARVALHO, 2017):

- Usabilidade : Os comando SQL e o do PostgreSQL são semelhantes, tornando-o assim fácil de usar.
- Segurança: Totalmente transacional, ou seja, é possível fazer qualquer tipo de transação com extrema segurança, sendo ela, uma inclusão, alteração ou exclusão.
- Eficência: Suporta diversas tecnologias tais como : JSON, XML, etc.
- Confiabilidade: É muito confiável tanto por parte de quem desenvolve tanto por parte de quem administra o banco de dados.
- Velocidade: Faz uso de indexação e consultas otimizadas para trabalhar com o menor esforço possível.

Algumas limitações do PostgreSQL estão listadas abaixo:

- Tamanho máximo do banco de dados: ilimitado
- Tamanho máximo de uma tabela: 32 TB
- Tamanho máximo de uma linha: 16 TB

Como mencionado o PostgreSQL vai além do que é necessário para o projeto para o nosso protótipo.

3.4 Fase I - Construção da Ontologia

A construção de ontologias exige grande esforço, disciplina e uma abordagem rigorosa; a tarefa de construir uma ontologia deve ser suficientemente formal e requer consenso por parte dos envolvidos no seu desenvolvimento ou daqueles que farão uso. Se feita por uma única pessoa, corre-se o risco de ocorrer um único ponto de vista sobre o domínio.

Conforme definido por Gruber as ontologias são resultado de especificações formais, estas favorecem a viabilização do processamento automático, e ao ser compartilhada expõe o conhecimento consonante acordado sobre o domínio de conhecimento.

A construção da ontologia OntoAAN foi feita a partir do uso de duas metodologias, sendo utilizadas as ferramentas Léxico Ampliado da Linguagem (LAL) (BREITMAN, 2005), alguns conceitos derivados da Método 101 (NOY; MCGUINNNESS, 2001).

As etapas do processo estão balizadas em: Construção do léxico, levantamento da informação, modelagem do LAL e análise do LAL.

3.4.1 Construção do Léxico

O Léxico Ampliado da Linguagem (LAL) é uma linguagem de representação simples. É composta apenas de três entidades básicas: Termo, Noção e Impacto. O LAL é uma linguagem de representação da Engenharia de Requisitos que tem por objetivo mapear o vocabulário utilizado no Universo de Informação (Udi), ou seja, o contexto no qual o software deverá ser desenvolvido e operado. O Udi inclui todas as fontes de informação e todas as pessoas relacionadas ao software. A construção do LAL é realizada nas etapas: Levantamento da Informação e Modelagem.

3.4.1.1 Levantamento de Informação

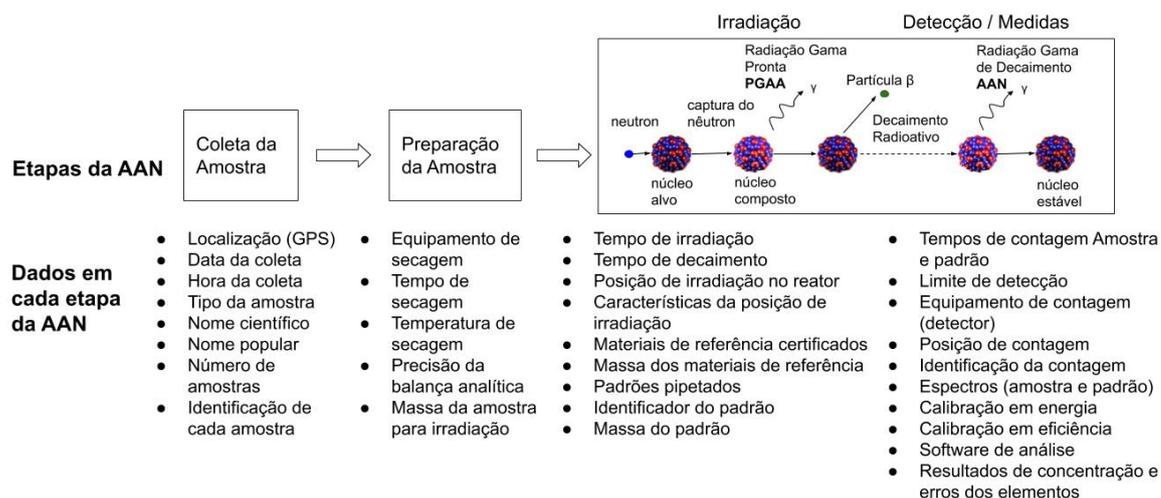
Visando coletar os termos pertinentes dentro do domínio da área de AAN, foi feito um extenso levantamento bibliográfico acerca do tema, foram lidas diversas teses e dissertações bem como entrevistas com especialistas da área. Todo este volume de dados coletado foi organizado de forma lógica resultando em fluxo de trabalho. Os termos coletados podem ser considerados apenas como símbolos e não houve qualquer pretensão em identificar alguma função para eles dentro da aplicação. A tabela 2 ilustra alguns termos coletados nesta etapa. A figura 7 e a tabela 2, mostra alguns termos coletados.

Tabela 2 – Lista de Termos Candidatos

Projeto	Padrão
Pesquisador	Elemento Tóxico
Amostra	Tempo de decaimento
Matriz	Limite de Detecção
Localização	Estrutura da Amostra
Temperatura	Nome Popular
Processo de Extração	Quantidade

A obtenção dos termos para a composição do léxico foi dividida de acordo com as etapas da AAN mostradas na figura 7.

Figura 7 – Dados gerados em cada etapa da AAN



Fonte: Elaborado pelo autor

3.4.1.2 Modelagem do LAL

A representação do LAL requer que para cada termo sejam descritos Noções e Impactos. Noção é o que significa o termo, Impacto descreve os efeitos do uso/ocorrência do termo na aplicação ou o efeito de algo na aplicação sobre o termo (BREITMAN, 2005). O principal objetivo da modelagem do LAL é manter uma estrutura enxuta e indentificar os relacionamentos que cada termo mantém com os outros. A Tabela 3 apresenta as regras usadas para modelar o LAL.

3.4.1.3 Análise do LAL

A análise do LAL ocorreu junto aos pesquisadores (orientadores) do CRPq, bem como dos pesquisadores (orientados) com o objetivo de garantir que a metodologia dos processos de AAN tenha sido bem entendida e de que nenhum termo importante não tenha sido contemplado no levantamento e nem na modelagem. A submissão dos processos de levantamento e modelagem visaram garantir nos certificarmos que tudo foi realizado de acordo com o proposto.

3.4.1.4 Mapeamento do LAL

Para realizar o processo de mapeamento do Léxico, foi feita a listagem dos termos de acordo com o seu tipo (verbo, objeto, sujeito ou estado) separando

Tabela 3 – Regras de formação do LAL

Termo tipo	Noção	Impacto
Sujeito	Quem é o sujeito?	Quais ações são executadas?
Verbo	Quem faz, quando acontece e quais procedimentos estão envolvidos	Quais são os impactos da ação no ambiente (UId), ou seja, quais são as outras ações que estão acontecendo também e quais são os estados resultantes.
Objeto	Defina o objeto e identifique outros objetos e ações relacionados aplicados ao objeto	Ações aplicadas ao objeto
Estado	O que significa e quais ações são responsáveis por esse estado	Identifique outros estados e ações possíveis que podem ser seguidos do estado descrito atual.

cada um em suas respectivas colunas. Como fica demonstrado na tabela **referenciar tabela 3**. A fase da construção consiste em capturar as informações geradas na análise do mapeamento e identificar todas as classes e organizá-las hierarquicamente, bem como os relacionamentos entre elas.

3.5 Fase II - Construção do Protótipo do Repositório

São detalhados nesta fase as etapas do desenvolvimento do protótipo. Previamente são abordadas a descrição dos autores, estes representam aqueles que farão uso do sistema. Em seguida é apresentado o diagrama de caso de uso, finalizando com as funcionalidades e restrições do sistema.

3.5.1 Descrição dos atores

Seguindo o propósito para qual o protótipo irá servir, os seguintes autores terão acesso ao sistema:

- Pesquisador: São todos aqueles que trabalham realizam pesquisas utilizando as técnicas de análise por ativação com nêutrons.
- Administrador: Responsável por manter o sistema atualizado e controlar todos os tipos de permissões.

3.5.2 Requisitos Funcionais

Os requisitos funcionais tem o objetivo de descrever as ações que cada tipo de usuário deve ser capaz de executar no sistema (Sommerville (2011)).

- Cadastro de Usuários: O sistema deve disponibilizar uma área para que o coordenador possa cadastrar pesquisadores.
- Redefinição de Senha: O sistema terá que ter uma área onde somente o administrador possa ter acesso a modificar a senha.
- Autenticação: Disponibilização de uma tela de login para que os usuários cadastrados possam se conectar ao sistema.
- Adição de Dados das Pesquisas: O sistema deve disponibilizar os meios para que o usuário possa adicionar os dados coletados por seus respectivos experimentos.
- Listagem dos Dados Cadastrados: O sistema tem que ser capaz de listar os dados cadastrados pelo usuário.
- Edição dos Dados Cadastrados: O sistema tem que ser capaz de permitir que o usuário possa modificar os dados.
- Remoção dos Dados Cadastrados: O sistema tem que ser capaz de permitir que o usuário possa apagar algum dado não desejado.

3.5.3 Diagrama de Caso de Uso

Os requisitos funcionais descritos anteriormente, serviram como base para modelar as funcionalidades do sistema. Sendo assim, é apresentado na figura 8 o diagrama de casos de uso do sistema.

Figura 8 – Diagrama de Caso de Uso da AAN



Fonte: Elaborado pelo autor

4 RESULTADOS E DISCUSSÃO

4.1 Modelagem da Ontologia

4.1.1 Construção e Análise do LAL

O processo de construção do LAL foi criado através de um levantamento de termos, estes foram classificados em quatro categorias: objeto, sujeito, estado e verbo. O processo de construção serviu para auxiliar a descrever o uso/ocorrências dos termos que farão parte do LAL. A tabela 4 apresenta alguns exemplos do processo de construção do LAL.

Uma vez finalizado o processo de construção, os termos que compunham o LAL passaram por um processo de análise, aos quais foram submetidos a uma curadoria cuidadosa, com o objetivo de validar e selecionar os termos pertinentes para à aplicação.

4.1.2 Mapeamento do LAL

A fase de mapeamento dos termos compreendeu em identificar as classes dos termos classificados como sujeitos e objetos e adicina - lós a elas, bem como estabelecer os relacionamentos entre as classes através das propriedades. A tabela 5 apresenta o mapemanto entre as classes, ou seja, mostra as propriedades e as classes ao qual ela relaciona.

4.2 Construção da Ontologia usando o Framework Protégé

4.2.1 Conceitualização da OntoAAN

O nível conceitual tem um esquema conceitual que descreve a estrutura de toda a base de dados. O esquema conceitual é uma descrição global da base de dados, que omite detalhes da estrutura de armazenamento físico e se concentra na descrição de entidades, tipos de dados, relacionamentos e restrições.

O nível conceitual descreve a estrutura de toda a base da composição da ontologia e se concentra na descrição de entidades e relacionamentos. A figura 9 representa a estrutura conceitual da OntoAAN, onde os retângulos representam as classes e as flechas as propriedades.

Tabela 4 – Léxico da AAN

Projeto/Projetos tipo:sujeito Noção: Planejamento das atividades Impacto: Desenvolvimento da pesquisa	Software tipo:sujeito Noção: Programa Impacto: Programa utilizado para realização das análises
Pesquisador/ Pesquisadores tipo:sujeito Noção: Planejamento das atividades Impacto: Amostra	Localização tipo:sujeito Noção: Lugar Impacto: Identificação de coleta das amostras
Matriz tipo: sujeito Noção: Natureza da amostra Impacto: Inclue todas as amostras	Temperatura tipo:sujeito Noção: Temperatura de secagem de uma amostra tratada Impacto: Aplica-se a todas as amostras tratadas que foram secadas por estufa, mufla, etc.
Quantidade tipo: sujeito Noção: Número de amostras coletadas Impacto: Inclue todas as amostras	Estado Físico tipo:sujeito Noção: Amostras tem estado físico Impacto: Característica física das amostras
Padrao_Pipetado tipo: sujeito Noção: Alíquota do material de referência Impacto: Pode ser irradiado junto com a alíquota no reator	Material_Referencia_Certificado tipo:sujeito Noção: material com concentração dos elementos bem definidos Impacto: Pode ser irradiado junto com a alíquota no reator
Homogeneizacao tipo: sujeito Noção: Tipo de método de tratamento Impacto: Método de tratamento para uniformizar a amostra	Irradiacao tipo:sujeito Noção: Tipo de método de tratamento. Impacto: Método de tratamento para ativar a amostra.
Secagem tipo: sujeito Noção: Tipo de método de tratamento Impacto: Retira a umidade das amostras	Pesagem tipo:sujeito Noção: Tipo de método de tratamento Impacto: Determinação da massa das amostras e alíquotas
Espectrometro tipo: sujeito Noção: Tipo de equipamento Impacto: Faz a análise qualitativa e quantitativa das alíquotas	Reator_Nuclear tipo:sujeito Noção: Tipo de equipamento Impacto: Irradia as alíquotas
Aliquota tipo: sujeito Noção: É parte da amostra já tratada Impacto: parte da amostra já tratada	Amostra_Tratada tipo:sujeito Noção: É parte da amostra Impacto: parte da amostra que passou por métodos de tratamento.

4.2.2 Identificador de Recurso Internacional (IRI)

De acordo com Isotani e Bittencourt (2015), Identificador de Recurso Internacional (ou IRI) é o responsável por identificar de maneira única um recurso

Tabela 5 – Propriedades do objeto

Propriedade	Domínio	Escopo
coletada_em	Amostra	Localidade
coletado_por	Amostra	Pesquisador
contem_amostra	Localidade	Amostra
contem_elemento	PadraoPipetado	Elemento
tem_estado	Amostra	EstadoFisico
tem_estado	AmostraTratada	EstadoFisico
tem_estado	Aliquota	EstadoFisico
passa_por	AmostraTratada	Metodo
usa_reator_nuclear	Irradiacao	ReatorNuclear
usa_espectrometro	Contagem	Espectrometro
usa Equipamento	Metodo	Equipamento
contem_elemento	PadraoPipetado	Elemento
desenvolvido_por	Projeto	Pesquisador
detectada_por	Equipamento	Equipamento
gerada_por	AnaliseEspectral	Software
irradiacao_usa	Irradiacao	ReatorNuclear
tem_software	Espectrometro	Software
tem_pipetado	PadraoPipetadoIrradiado	PadraoPipetado
tem_matriz	Amostra	Matriz
tem_estado_fisico	ObjetoAnalise	EstadoFisico

na Web. O RDF necessita que todo recurso tenha um URI/IRI, então o IRI pode representar um <sujeito>, <predicado> ou <objeto>.

Para isso foi gerada uma IRI <http://www.ipen.br/crpq/inaa> para a OntoAAN, a qual irá associar todos os recursos da ontologia.

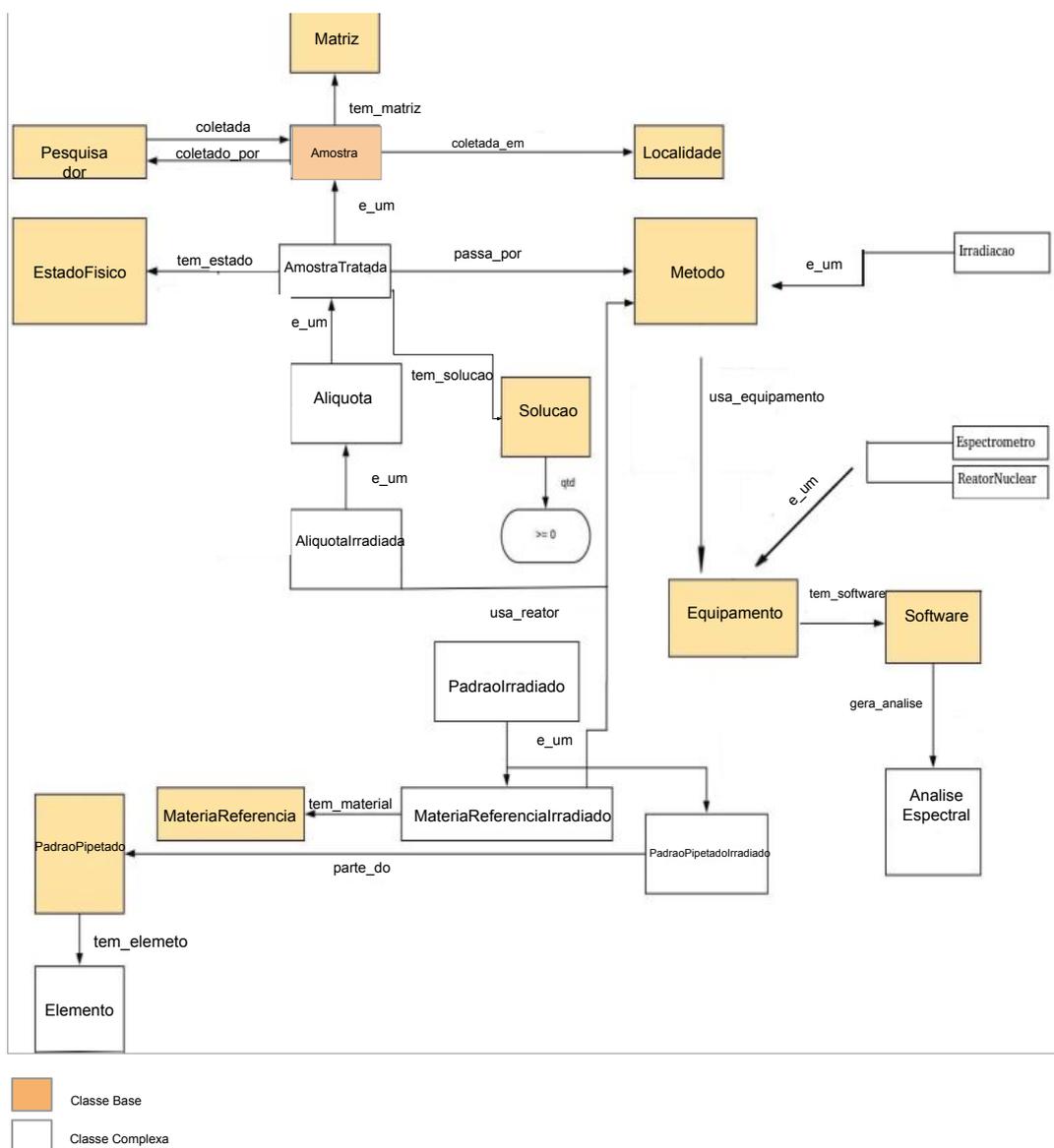
4.2.3 Estrutura de classes e subclasses da OntoAAN

Diferentemente das representações de Dicionários, Vocabulário Controlado, Tesouro ou Lexical. A ontologia é a estrutura na qual as informações são organizadas. Em dicionários, vocabulários controlados, tesouro ou representações lexicais, as informações estão contidas em um plano, enquanto em Ontologias, são organizadas em hierarquias. A tabela 5 ilustra algumas classes que compõem a ontologia, enquanto que a figura 10 ilustra a composição da estrutura de classes no editor Protégé.

4.2.4 Relacionamentos da OntoAAN

As classes por si só não são capazes de responder questões de competência, sendo assim, é necessário criar os relacionamentos (chamados de Object Properties na nomenclatura do Protégé), para que estabeleçam vínculos entre as

Figura 9 – Conceitualização da Ontologia



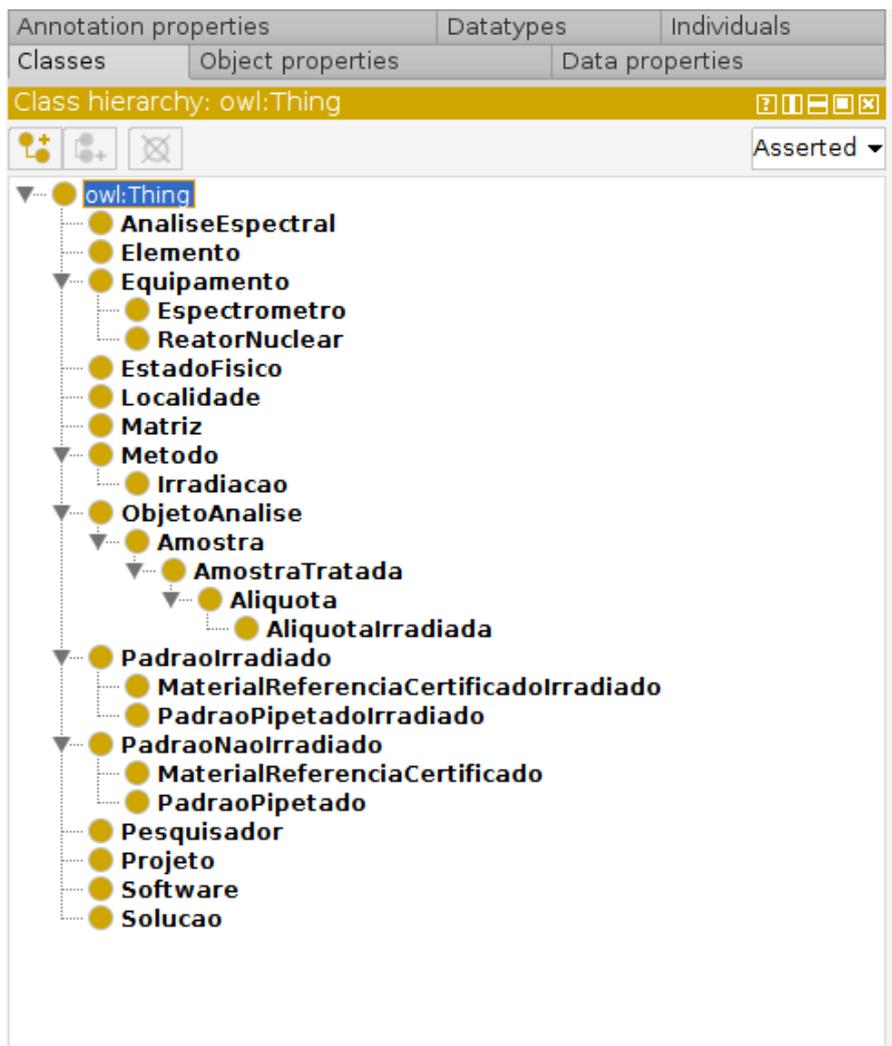
Fonte: autor

classes, para isso, é necessário determinar a classe de Domínio e a classe de Destino dentro da propriedade do objeto para que o vínculo entre elas aconteça. A figura 10 ilustra a composição da estrutura de classes no editor Protégé, onde são compostas por dois tipos de classes as bases e as complexas, enquanto as classes bases independente de outras classes para existirem as complexas dependem inteiramente das classes bases para se relacionarem.

4.2.5 Atributos da OntoAAN

A OntoAAN contém diversos atributos (Slots ou Data Properties na nomenclatura do Protégé). Todos eles são utilizados para representar valores de tipos de dados básicos (inteiro, real, string) para os indivíduos das classes. A figura 12

Figura 10 – Hierarquia de Classes



Fonte: autor

ilustram alguns atributos criados.

4.2.6 OntoAAN

A OntoAAN possui 26 classes, 22 relacionamentos indivíduos. A seguir serão destacadas as classes, as suas propriedades e restrições.

Classe ObjetoAnalise - possui a hierarquia representada na figura 13.

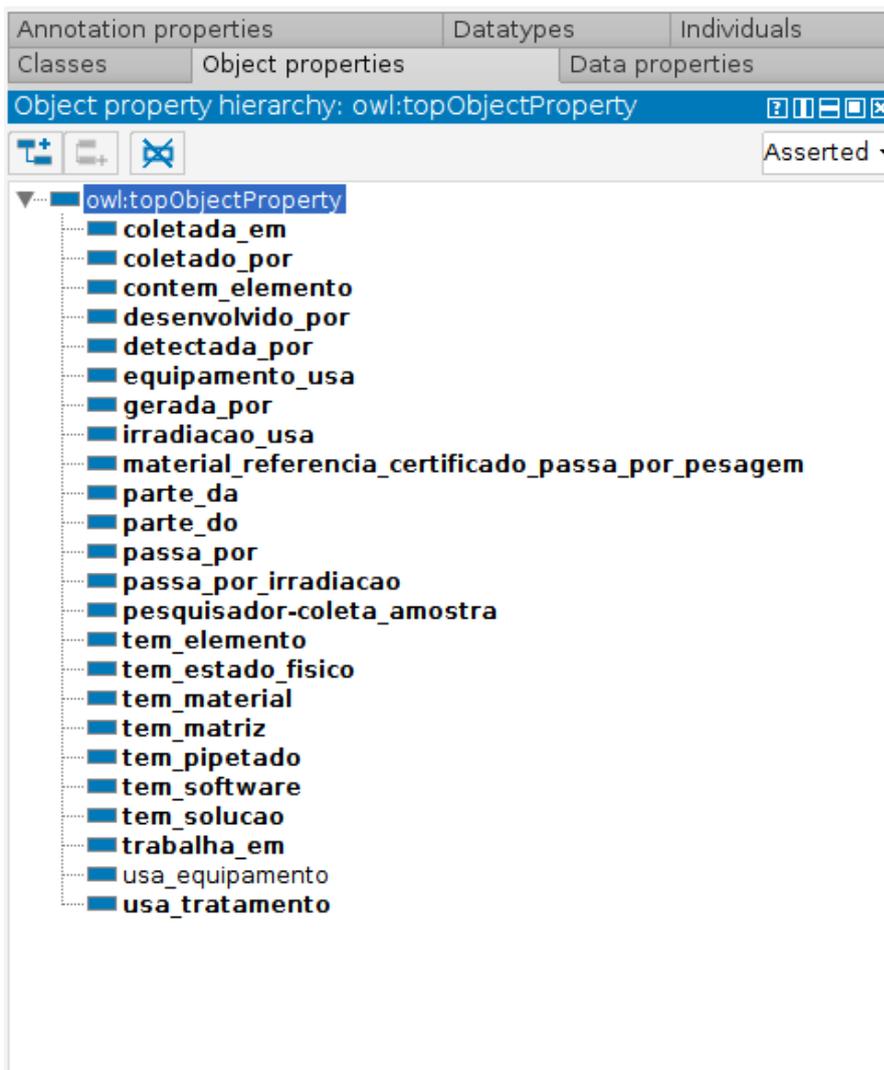
Figura 13 – Hierarquia da classe ObjetoAnalise



Fonte: autor

Classe Amostra - Representa as amostras coletadas por indivíduos da classe Pesquisador. Possui as seguintes propriedades:

Figura 11 – Propriedades da OntoAAN



Fonte: autor

- tem_matriz : identifica o tipo de matriz que a amostra possui.
- coletada_por : identifica o pesquisador que coletou a amostra.

A hierarquia da classe Amostra é mostrada na figura 14.

Figura 14 – Hierarquia da classe Amostra

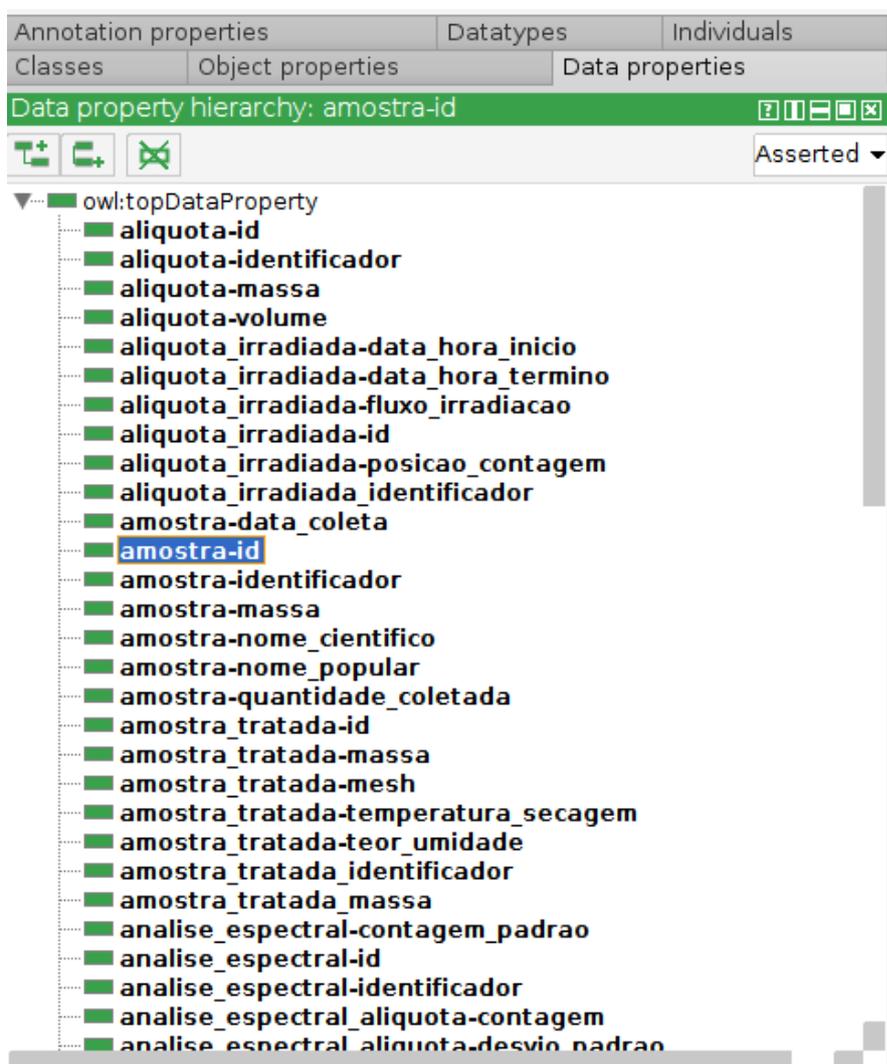


Fonte: autor

Classe AmostraTratada - Representa as amostras que estão associadas a indivíduos da classe Metodo.

- passa_por : identifica o tipo de metodo a qual a amostra foi submetida.

Figura 12 – Atributos da OntoAAN



Fonte: autor

- tem_solucao : identifica se a amostra foi tratada com algum tipo de solução.

Aliquota - Representa uma porção da amostra que já passou por algum método de tratamento e contém a subclasse AliquotaIrradiada. A figura **falta referenciar, o número da figura não aparece** mostra a hierarquia da classe Aliquota.

Figura 15 – Hierarquia da classe Aliquota



Fonte: autor

AliquotaIrradiada - Representa as amostras que estão associadas a indivíduos da classe Metodo 17.

Figura 16 – Hierarquia da classe AliquotalIrradiada



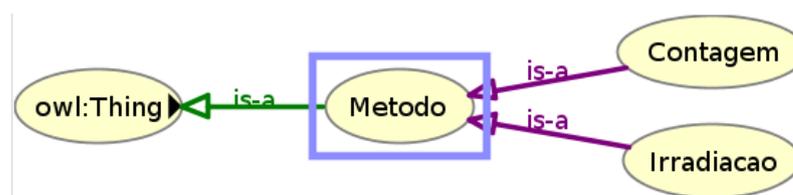
Fonte: autor

Método - Representa os métodos de tratamento a qual uma amostra pode ser submetida.

- `usa_equipamento` : identifica o equipamento a qual o método de tratamento faz uso.

O Método possui também duas subclasses: Contagem e Irradiação.

Figura 17 – Hierarquia da classe Método



Fonte: autor

Contagem - Representa o Método ao qual as amostras irradiadas são submetidas para gerar a contagem dos raios gamas em um espectrômetro, como mostrado na figura 18.

Figura 18 – Hierarquia da classe Contagem



Fonte: autor

- `usa_espectrometro` : relaciona a classe Contagem com a classe Espectrometro.
- `passa_por_contagem`: relaciona a classe AliquotalIrradiada com a classe Contagem.

Irradiação - Representa o Método ao qual as amostras são inseridas dentro de um Reator Nuclear, como mostrado na figura 19.

Figura 19 – Hierarquia da classe Irradiacao

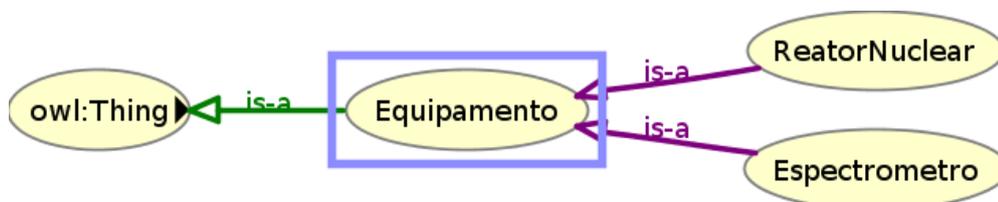


Fonte: autor

- irradiacao_usa : relaciona a classe Contagem com a classe Espectrometro.
- usa_reator_nuclear: relaciona a classe AliquotasIrradiada com a classe Contagem.

Equipamento - Possui a hierarquia da classe Equipamento, representada na figura 20.

Figura 20 – Hierarquia da classe Equipamento



Fonte: autor

ReatorNuclear - Representa o Equipamento onde as alíquotas são inseridas para serem irradiadas, como mostrado na figura 21.

Figura 21 – Hierarquia da classe ReatorNuclear



Fonte: autor

Espectrometro - Representa o Equipamento onde as alíquotas já irradiadas são inseridas para serem medidas e gerarem os dados da análise, como mostrado na figura 23.

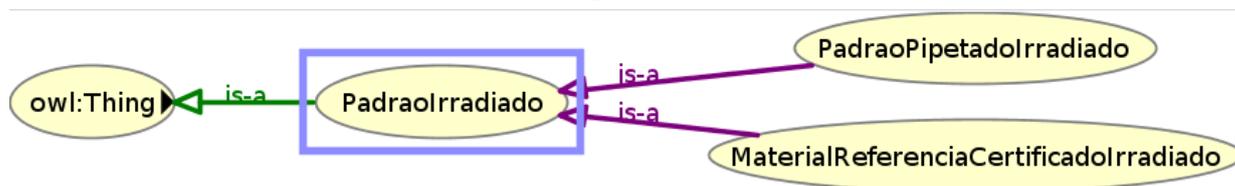
Figura 22 – Hierarquia da classe Espectrometro



Fonte: autor

PadraoNaIrradiado - Representa a hierarquia de classes dos padrões não irradiados, é composta pelas subclasses MaterialReferenciaCertificado e PadraoPipetado.

Figura 23 – Hierarquia da classe do PadraoNaIrrradiado



Fonte: autor

MaterialReferenciaCertificado - Representa os materiais de referência da classe MaterialReferenciaCertificado como mostrado na figura 24

Figura 24 – Classe MaterialReferenciaCertificado



Fonte: autor

PadraoPipetado - Representa a classe Padrão Pipetado da classe PadraoPipetado como mostrado na figura 25

Figura 25 – Classe PadraoPipetado



Fonte: autor

Elemento - Representa os elementos que compõem o padrão pipetado. A classe Elemento é mostrada na figura 26

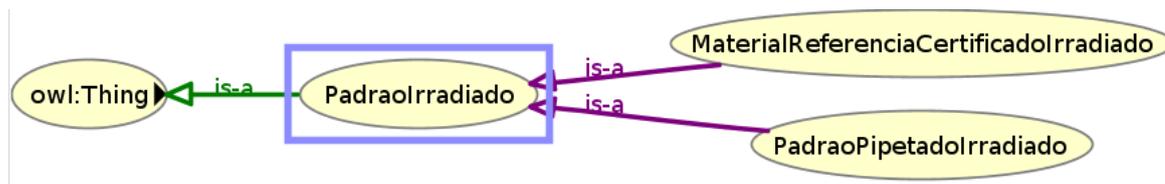
Figura 26 – Classe Elemento



Fonte: autor

PadraoIrrradiado - Representa a hierarquia de classe do PadraoIrrradiado como mostrado na figura 27

Figura 27 – Hierarquia da classe Elemento



Fonte: autor

MaterialReferenciaCertificadoIrradiado - Representa a classe dos Materiais de Referência MaterialReferenciaCertificadoIrradiado como mostrado na figura 28.

Figura 28 – Hierarquia da classe MaterialReferenciaCertificadoIrradiado

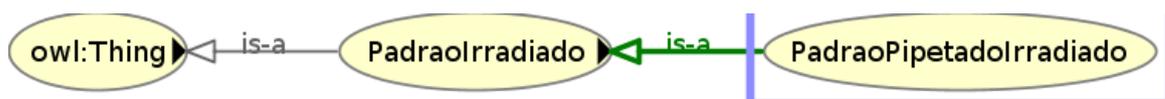


Fonte: autor

- `passa_por_irradiacao` : relaciona a classe MaterialReferenciaCertificadoIrradiado com a classe Irradiacao.

PadraoPipetadoIrradiado - Representa a hierarquia de classe do PadraoPipetadoIrradiado como mostrado na figura 29.

Figura 29 – Hierarquia da classe PadraoPipetadoIrradiado



Fonte: autor

- `passa_por_irradiacao` : relaciona a classe PadraoPipetadoIrradiado com a classe Irradiacao.

AnaliseEspectral -

Figura 30 – Hierarquia da classe AnaliseEspectral



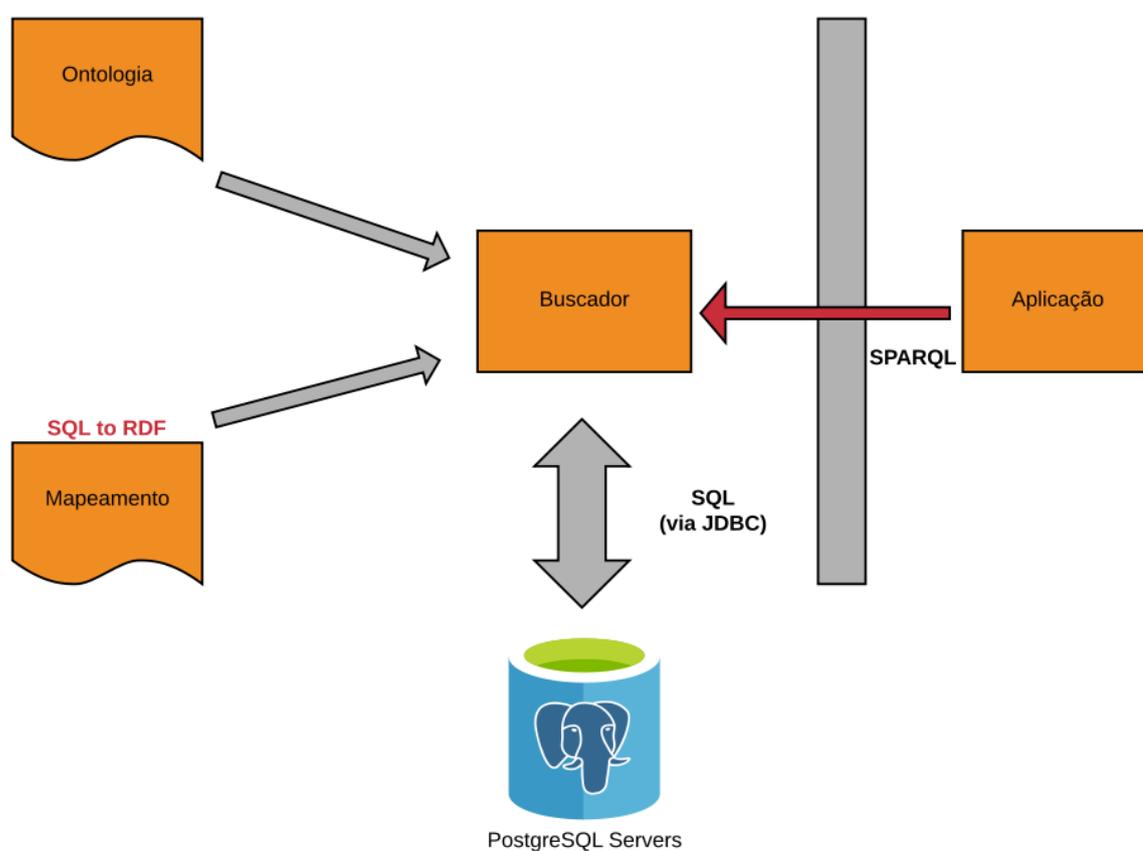
Fonte: autor

- `irradiacao_usa` : relaciona a classe Contagem com a classe AnaliseEspectral.

4.3 Mapeamento e consulta do Protégé com o PostgreSQL

O editor Protégé conta com o plugin Ontop, sendo este responsável por fazer a conexão entre o banco de dados relacional PostgreSQL com a ontologia criada dentro do editor usando o Ontop Mapping, pois este trabalha modelo OBDA. As fontes de dados em ontop são identificadas por um URI e são definidas por meio das propriedades da conexão JDBC. A figura 31

Figura 31 – Arquiteura Ontop com a camada semântica



Fonte: autor

O processo de mapeamento consistiu em vincular as consultas SQL com as propriedades criadas em na janela Data Property.

O processo de mapeamento da base de dados relacional para a camada semântica mostrou - se um trabalho bastante meticuloso, pois mesmo todo o desenvolvimento dentro do editor Protégé, a tarefa de mapear os dados exigiu minucia em vincular os tipos de dados da ontologia com os campos da base de dados.

A seguir é apresentado na figura 32 o mapeamentos dos campos referente a classe Amostra em R2RML, especificando sujeito, predicado e objeto na camdada semântica:

A figura 33 apresenta os resultados da consulta SPARQL dentro do editor

Figura 32 – Mapeamento da Amostra no formato R2RML

```

<urn:map-amostra_1> a          rr:TriplesMap ;
  rr:logicalTable [ a          rr:R2RMLView ;
    rr:sqlQuery  "SELECT * FROM \"public\".\"portal_amostra\""
  ] ;
  rr:predicateObjectMap [ a      rr:PredicateObjectMap ;
    rr:objectMap [ a      rr:ObjectMap , rr:TermMap ;
      rr:column      "nome_popular" ;
      rr:datatype    xsd:string ;
      rr:termType    rr:Literal
    ] ;
    rr:predicate    :amostra-nome_popular
  ] ;
  rr:predicateObjectMap [ a      rr:PredicateObjectMap ;
    rr:objectMap [ a      rr:ObjectMap , rr:TermMap ;
      rr:column      "nome_cientifico" ;
      rr:datatype    xsd:string ;
      rr:termType    rr:Literal
    ] ;
    rr:predicate    :amostra-nome_cientifico
  ] ;
  rr:predicateObjectMap [ a      rr:PredicateObjectMap ;
    rr:objectMap [ a      rr:TermMap , rr:ObjectMap ;
      rr:column      "identificador" ;
      rr:datatype    xsd:string ;
      rr:termType    rr:Literal
    ] ;
    rr:predicate    :amostra-identificador
  ] ;
  rr:predicateObjectMap [ a      rr:PredicateObjectMap ;
    rr:objectMap [ a      rr:TermMap , rr:ObjectMap ;
      rr:column      "id" ;
      rr:datatype    xsd:integer ;
      rr:termType    rr:Literal
    ] ;
    rr:predicate    :amostra-id
  ] ;

```

Fonte: autor

Protégé em OntoSPARQL, sendo assim, foram listando as amostras da base de dados.

As variáveis SPARQL começam com um caractere “?” e podem ser definidas em qualquer uma das três posições de uma tripla (sujeito, predicado, objeto) no conjunto de dados RDF. Os padrões de triplas da cláusula SELECT têm a mesma forma de triplas normais, exceto que qualquer uma das três partes da tripla pode ser substituída por uma variável. A cláusula SELECT retorna uma tabela de variáveis com os valores que satisfazem a consulta.

O uso de uma ontologias como sistemas de organização do conhecimento para fins de representação, organização e recuperação da informação e comparti-

Figura 33 – Consulta Sparql classe Amostra

```

SELECT DISTINCT ?id ?identificador ?nome_popular ?nome_cientifico
WHERE {
  ?x <http://www.ipen.br/crpq/inaa#amostra-nome_popular> ?amostra-id .
  ?x <http://www.ipen.br/crpq/inaa#amostra-nome_popular> ?amostra-identificador .
  ?x <http://www.ipen.br/crpq/inaa#amostra-nome_popular> ?amostra-nome_popular .
  ?x <http://www.ipen.br/crpq/inaa#amostra-nome_popular> ?amostra-nome_cientifico .

```

id	identificador	nome_popular	nome_cientifico
1	G1	Água	H2O
2	L1	Água	H2O
3	R1	Água	H2O
4	P1	Água	H2O

Fonte: autor

lhamento de conhecimento, é vital, pois ela dá o suporte necessário para resolver alguns dos problemas que cercam a o desenvolvimento de tecnologias que utilizam bases de dados com representação formal.

4.4 Protótipo do Repositório de AAN

O protótipo do repositório foi desenvolvido usando softwares de código aberto. A linguagem de programação escolhida é o Python 3, juntamente com a estrutura dedesenvolvimento web do Django 2.1 e como banco de dados para persistir os dados foi utilizado o PostgreSQL.

No diagrama de casos de uso, podemos começar a pensar sobre as entidades do nosso projeto. As entidades são os modelos que criaremos e estão intimamente relacionados aos dados que nosso aplicativo Django processará. A figura iustra no diagrama de classe do repositório.

O Framework Django faz uso da arquitetura MTV (Model-Template-View) como citado anteriormente, sendo assim, todos os modelos que são subclasses da classe `django.db.models.Model` serão transformados em tabelas de banco de dados. Cada campo é representado por instâncias das subclasses `django.db.models.Field` (núcleo interno do Django) e serão convertido em colunas do banco de dados.

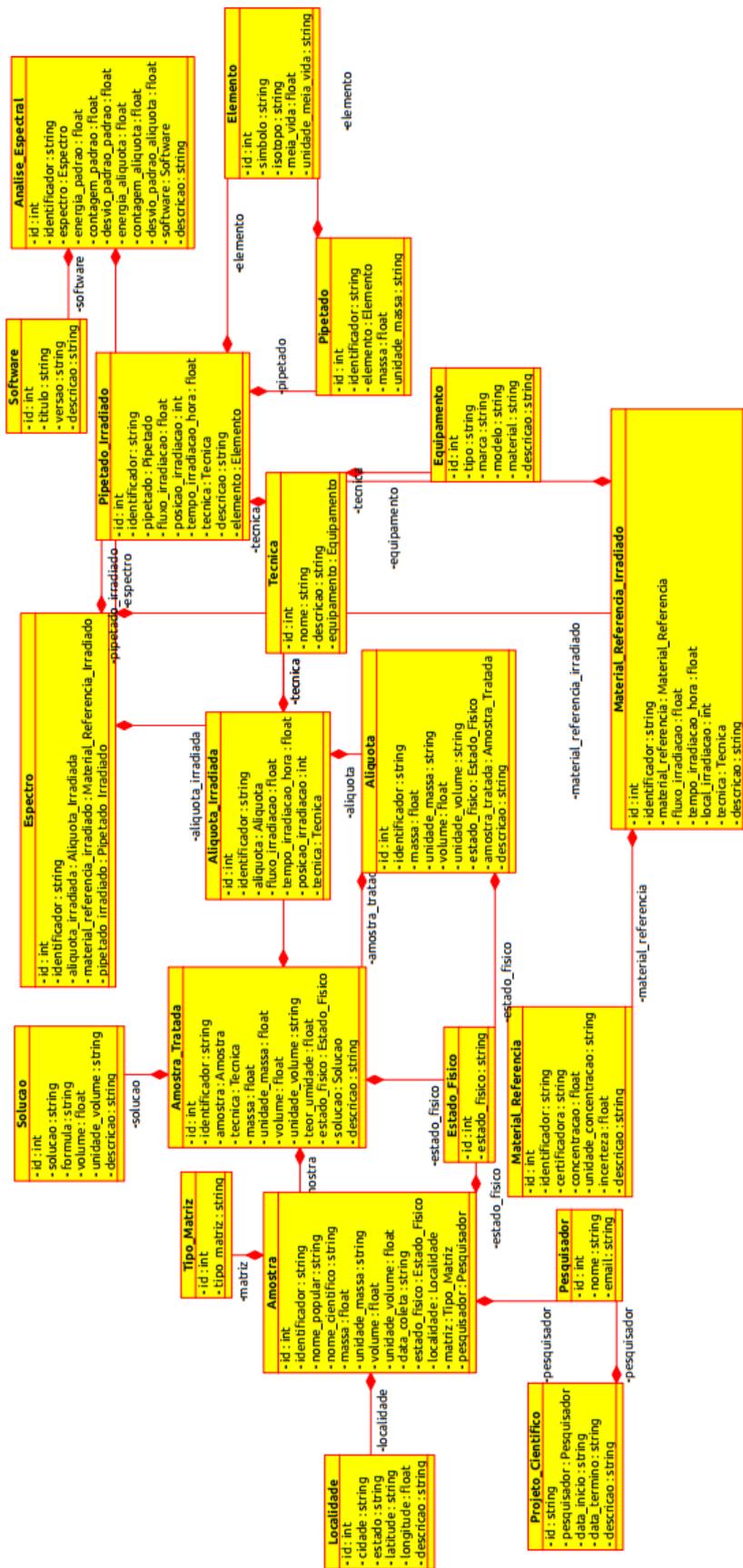
As classes criadas no LAL da OntoAAN serviram como base para a construção do Model, bem como os termos gerados no levantamento dos dados na figura 7 serviram como gerar as colunas em suas respectivas classes. A figura ilustra a construção da classe Amostra dentro de model 35.

Após a construção de todas as classes no Model do Django foi geradas as Views, esta representam a "lógica"da aplicação e ela tem a função de extrair as

informações do model criado e envia - lás aos templates, bem como as URLs para definir o endereço de cada página para que posasm ser acessados através de um navegador. A figura 36 apresenta a tela de listagem das amostras cadastradas.

O protótipo tem o intuito de persistir as informações coletadas pelos pesquisadore e que derivirão como dados para serem a extração de conhecimento usando o plugin Ontop do editor Protégé.

Figura 34 – Classes do Protótipo



Fonte: autor

Figura 35 – Classe Amostra

```

class Amostra(models.Model):
    identificador = models.CharField(max_length=30)
    user = models.ForeignKey(User, on_delete=models.DO_NOTHING)
    nome_popular = models.CharField(max_length=30)
    nome_cientifico = models.CharField(max_length=30)
    massa = models.DecimalField(max_digits=5, decimal_places=2, null=True, blank=True)
    unidade_massa = models.CharField(max_length=5, choices=MASS_CHOICES, null=True, blank=True)
    volume = models.DecimalField(max_digits=5, decimal_places=4, null=True, blank=True)
    unidade_volume = models.CharField(max_length=3, choices=VOLUME_CHOICES, null=True, blank=True)
    data_coleta = models.DateTimeField(auto_now_add=True, null=True, blank=True)
    estado_fisico = models.ForeignKey(Estado_Fisico, on_delete=models.DO_NOTHING, related_name='amostra_estadoFisico')
    matriz = models.ForeignKey(Tipo_Matriz, on_delete=models.DO_NOTHING, related_name='matriz_tipo')

    localidade = models.ForeignKey(
        Localidade,
        on_delete=models.CASCADE,
        related_name='amostra_localidade',
        null=True,
        blank=True
    )

    class Meta:
        verbose_name_plural = 'Amostras'

    def get_absolute_url(self):
        return reverse('amostra_detalhe', kwargs={'pk':self.pk})

    def __str__(self):
        return '%s, %s, %s' % (self.identificador, self.nome_popular, self.nome_cientifico)

```

Fonte: autor

Figura 36 – Interface da listagem das Amostras

← → C 127.0.0.1:8000/amostra

Repositório de AANI

Listagem das Amostras

[Inserir Amostra](#)

Código	Identificador	Nome popular	Nome científico	Massa	Unidade massa	Volume	Unidade volume	Estado físico	
1	G1	Água	H2O	None	None	5.0000	l	Líquido	Editar Excluir
2	L1	Água	H2O	None	None	5.0000	l	Líquido	Editar Excluir
4	P1	Água	H2O	None	None	5.0000	l	Líquido	Editar Excluir
3	R1	Água	H2O	None	None	5.0000	l	Líquido	Editar Excluir

Fonte: autor

5 CONCLUSÃO

Neste projeto foram abordados os conceitos, os critérios de construção, as metodologias utilizadas sendo estas: Metodologia 101 e a ferramenta Lexico Ampliado a Linguagem (LAL), linguagens e ferramentas para a construção tanto das ontologias como do protótipo de repositório.

As Ontologias fornecem mecanismos para inserção de semântica nos documentos para que os agentes possam raciocinar sobre os mesmos. A ideia deste trabalho reside na construção de ontologia de domínio de dados referente a área de Análise por Ativação Neutrônica e a criação de um protótipo de repositório de dados de pesquisa, desenvolvidas no Centro do Reator de Pesquisas - CERPq, do IPEN-CNEN/SP, que abrangem diversas áreas de aplicação, tais como dados relacionados a amostras ambientais, coletadas com indicação de geolocalização cuja ontologia foi integrada, onde agentes de busca poderão determinar o que representa cada material e onde ele se encontra no repositório.

Os resultados apresentados demonstram os ganhos oriundos desta integração e apontaram para vantagens de se ter uma camada semântica em um repositório de dados, incluindo, mas não se limitando, à maior possibilidade de reuso dos dados visto que podem ser melhor entendidos a partir da ontologia que os descreve. A integração da camada semântica com o SGBD PostgreSQL favoreceu a inferência de dados através da linguagem SPARQL, o que favoreceu em eficiência e evita a ambivalência de termos.

A metodologia desenvolvida contribuiu para a criação da ontologia, é detalhada, permitindo sua aplicação a outros domínios correlatos dentro do IPEN-CNEN/SP, bem como, em outras instituições, no qual foram desenvolvidos mecanismos de integração desta camada semântica com a estrutura tradicional de uma aplicação Web, isto é, com bancos de dados relacionais tradicionais, de modo que os benefícios das operações semânticas pudessem ser incorporados a um sistema tradicional. O protótipo de repositório foi desenvolvido utilizando-se um framework de desenvolvimento Web de código aberto – Django.

A integração da ontologia com o protótipo do repositório não foi possível de ser realizado, pois como foi utilizado um banco de dados relacional para armazenamento de dados e não um banco de dados de tripla como o RDF4J (The Eclipse Foundation, 2020) , OpenLink Virtuoso (ERLING; MIKHAILOV, 2010), entre outros, isto implicou em algumas desvantagens, principalmente no processo de mapear

a ontologia owl junto ao banco de dados dentro do protótipo, o que se mostrou uma atividade extremamente complexa, com poucas ferramentas identificadas para realizar tal tarefa em Python, dentre as quais a Owlready2 (LAMY, 2017). Porém, a ausência de uma documentação mais bem detalhada da Owlready2 impediu que esta pudesse ser melhor aproveitada no desenvolvimento da camada de integração. Como sugestão para um trabalho futuro, esta integração da camada semântica com o PostgreSQL junto ao protótipo do repositório pode ser realizada.

REFERÊNCIAS

- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, v. 32, n. 3, p. 7–20, 2003. ISSN 0100-1965.
- BONTCHEVA, K. et al. *Semantic Annotation and Human Language Technology*. John Wiley Sons, Ltd, 2006. 29-50 p. ISBN 9780470030332. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/047003033X.ch3>>.
- BORGES, L. *Python para Desenvolvedores: Aborda Python 3.3*. Novatec Editora, 2014. ISBN 9788575224052. Disponível em: <<https://books.google.com.br/books?id=eZmtBAAQBAJ>>.
- BRAATZ, B.; BRANDT, C. Graph transformations for the resource description framework. 01 2008.
- BREITMAN, K. *Web Semântica - A Internet do Futuro*. 1ª. ed. Rio de Janeiro, RJ: LTC - Livros Técnicos e Científicos Editora S.A., 2005. 190 p. ISBN 85-216-1466-7.
- CALERO CORAL; RUIZ, F. P. M. *Ontologies for Software Engineering and Software Technology*. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN 3540345175.
- CALVANESE, D. et al. How to stay ontop of your data: Databases, ontologies and more. In: . [S.l.: s.n.], 2015. p. 20–25. ISBN 978-3-319-25638-2.
- CARVALHO, V. *PostgreSQL: Banco de dados para aplicações web modernas*. [S.l.]: Casa do Código, 2017. ISBN 9788555192562.
- CASELLAS, N. *Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge*. [S.l.]: Springer Netherlands, 2011. (Law, Governance and Technology Series). ISBN 9789400714977.
- CATHARINO, M. Gabriela Miranda. *Análise de Mercúrio e Selênio em Materiais Biológicos pelo Método de Análise por Ativação com Nêutrons*. Dissertação (Mestrado) — Instituto de Pesquisas Energéticas e Nucleares - IPEN/CNEN-SP, São Paulo, 2002.
- CHIBENI, S. S. O que é ciência? 2013.
- CORDEIRO, D. et al. Da ciência à e-ciência: paradigmas da descoberta do conhecimento. *Revista USP*, n. 97, p. 71–81, maio 2013. Disponível em: <<http://www.revistas.usp.br/revusp/article/view/61867>>.
- DEMCHENKO, Y. et al. Cloud-based infrastructure for data-intensive e-science applications: Requirements and architecture. In: TERZO, O.; MOSSUCCA, L. (Ed.). *Cloud Computing with e-Science Applications*. 1. ed. Boca Raton, FL - USA: CRC Press, 2015. p. 310. ISBN 13-978-1-4665-9116-5.
- ERLING, O.; MIKHAILOV, I. Virtuoso: Rdf support in a native rdbms. In: _____. *Semantic Web Information Management: A Model-Based Perspective*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 501–519. ISBN 978-3-642-04329-1. Disponível em: <https://doi.org/10.1007/978-3-642-04329-1_21>.

FRANK VAN, H. B. S. et al. *OWL Web Ontology Language*. 2003. Disponível em: <<https://www.w3.org/TR/owl-ref/>>.

GRAY, J. The fourth paradigm – data-intensive scientific discovery. In: _____. Redmond, Washington: Microsoft Research, 2009. cap. Jim Gray on e-Science – A Transformed Scientific Method, p. xvii–xxxii.

GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, v. 5, n. 2, p. 199–220, 1993. ISSN 1042-8143. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1042814383710083>>.

GUARINO, N. Formal ontology and information systems. In: *FOIS'98*. Trento, Italy: IOS Press, 1998. p. 3–15.

HORROCKS, I. A. N.; SATTLER, U.; TOBIES, S. Practical Reasoning for Very Expressive Description Logics. *Logic Journal of IGPL*, v. 8, n. 3, p. 239–263, 2000. ISSN 1367-0751.

IAEA. *Guidelines for radioelement mapping using gamma ray spectrometry data*. Viena, Austria, IAEA–TECDOC–1363, July 2003.

ISOTANI, S.; BITTENCOURT, I. *Dados Abertos Conectados: Em busca da Web do Conhecimento*. [S.l.]: Novatec, 2015. ISBN 9788575224496.

KNOLL, G. F. *Radiation Detection and Measurement*. 2nd. ed. Singapore: John Wiley & Sons, 1989.

KUHN, T. S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

LAMY, J.-B. Owlready: Ontology-oriented programming in python with automatic classification and high level constructs for biomedical ontologies. *Artificial intelligence in medicine*, Elsevier, v. 80, p. 11–28, 2017.

LASSILA, O.; MCGUINNESS, D. The role of frame-based representation on the semantic web. *Linköping Electronic Articles in Computer and Information Science*, v. 6, n. 5, p. 10, 2001. ISSN 1401-9841. Disponível em: <<http://www.ep.liu.se/ea/cis/2001/005/cis01005.pdf>>.

MARQUES, F. Ciência transparente. *Revista Fapesp*, n. 218, p. 54–58, 2014.

MCGUINNESS, D. L. Question answering on the semantic web. *IEEE Intelligent Systems*, IEEE Educational Activities Department, USA, v. 19, n. 1, p. 82–85, jan. 2004. ISSN 1541-1672. Disponível em: <<https://doi.org/10.1109/MIS.2004.1265890>>.

MILLER, E. An introduction to the resource description framework. *D-Lib Magazine*, 5 1998.

MOREIRA, E. G. *Aplicação da Análise por Ativação com Nêutrons ao Estudo da Composição Química de Materiais Metálicos*. Dissertação (Mestrado) — Instituto de Pesquisas Energéticas e Nucleares - IPEN/CNEN-SP, São Paulo, 2002.

NOY, N. F.; MCGUINNESS, D. L. Ontology Development 101: A Guide to Creating Your First Ontology. *Stanford Knowledge Systems Laboratory*, p. 25, 2001. ISSN 09333657.

ROMANO, F.; HILLAR, G.; RAVINDRAN, A. *Learn Web Development with Python: Get hands-on with Python Programming and Django web development*. [S.l.]: Packt Publishing, 2018. ISBN 9781789950885.

SCHREIBER, G.; DEAN, M. W3C Recommendation, *OWL Web Ontology Language Reference*. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-owl-ref-20040210/>>.

SOMMERVILLE, I. *Engenharia de software*. [S.l.]: PEARSON BRASIL, 2011. ISBN 9788579361081.

STAAB, S.; STUDER, R. Handbook on ontologies. *Decision Support Systems*, v. 16, p. 654, 2007. ISSN 10744770. Disponível em: <<http://www.gbv.de/du/services/toc/bs/368354474>>.

The Eclipse Foundation. *RDF4J*. 2020. Disponível em: <<https://pythonhosted.org/Owlready2/index.html>>.

YU, L. *A Developer's Guide to the Semantic Web*. [S.l.]: Springer Berlin Heidelberg, 2011. (IT Pro). ISBN 9783642159701.

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Diretoria de Pesquisa, Desenvolvimento e Ensino
Av. Prof. Lineu Prestes, 2242 – Cidade Universitária CEP: 05508-000
Fone/Fax(0XX11) 3133-8908
SÃO PAULO – São Paulo – Brasil
<http://www.ipen.br>

O IPEN é uma Autarquia vinculada à Secretaria de Desenvolvimento, associada à Universidade de São Paulo e gerida técnica e administrativamente pela Comissão Nacional de Energia Nuclear, órgão do Ministério da Ciência, Tecnologia, Inovações e Comunicações.
