UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
INSTITUTO DE QUÍMICA DE SÃO CARLOS

RICARDO ROBERTO DE LIMA

Desenvolvimento de um portal de conhecimento para a tuberculose baseado em
web semântica

São Carlos

2018

RICARDO ROBERTO DE LIMA
VERSÃO CORRIGIDA

Desenvolvimento de um portal de conhecimento para a tuberculose baseado em
web semântica

Dissertação apresentada à Escola de Engenharia de São Carlos, Faculdade de Medicina de Ribeirão Preto e Instituto de Química de São Carlos da Universidade de São Paulo, como requisito para a obtenção do Título de Mestre em Ciências.

Área de concentração: Bioengenharia

Orientador: Prof. Dr. Domingos Alves

São Carlos

2018

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

L732d Lima, Ricardo Roberto de
Desenvolvimento de um portal de conhecimento para a tuberculose baseado em web semântica / Ricardo Roberto de Lima; orientador Domingos Alves. São Carlos, 2018.

Dissertação (Mestrado) - Programa de Pós-Graduação Interunidades em Bioengenharia e Área de Concentração em Bioengenharia -- Escola de Engenharia de São Carlos; Faculdade de Medicina de Ribeirão Preto; Instituto de Química de São Carlos, da Universidade de São Paulo, 2018.

1. Tuberculose. 2. Web semântica. 3. Indicadores de Tuberculose. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE JULGAMENTO

Candidato(a): Ricardo Roberto de Lima

Título: "Desenvolvimento de um portal de conhecimento para a tuberculose baseado em web semântica"

Data da defesa: 26/10/2018

Comissão Julgadora	Assinatura	Resultado
Prof(a). Dr(a). Domingos Alves FMRP/USP		<u>Não votante</u>
Prof(a). Dr(a). Rui Pedro Charters Lopes Rijo Instituto Politécnico de Leiria		<u>Aprovado</u>
Prof(a). Dr(a). Ieda Pelógi Martins Damian FFCLRP/USP		<u>Aprovado</u>
Dr(a). Flávio Barbosa CIAware		<u>Aprovado.</u>

*À minha querida esposa e queridos
filhos pela compreensão, carinho e
apoio incansável.*

AGRADECIMENTOS

Primeiramente, agradecer a Deus pela oportunidade de realizar meu mestrado.

A minha família por estar sempre ao meu lado me apoiando e compreendendo minha ausência em diversos momentos.

Ao Professor Doutor Domingos Alves, que muito me ensinou contribuindo para o meu crescimento científico.

Ao Professor Doutor Rui Pedro Charters Lopes Rijo, pela atenção e conselhos que contribuíram para meu aprendizado.

Aos presentes na banca que se colocaram a disposição e dedicaram uma parte de seu tempo a me ouvir.

“A desobediência é uma virtude necessária
à criatividade”

Raul Seixas (1989)

RESUMO

LIMA, R. R de. **Desenvolvimento de um portal de conhecimento para a tuberculose baseado em web semântica**. 2018. 97 f. Dissertação (Mestrado em Ciências) - Escola de Engenharia de São Carlos, Faculdade de Medicina de Ribeirão Preto e Instituto de Química de São Carlos da Universidade de São Paulo, São Carlos, 2018.

Segundo o relatório de 2017 da Organização Mundial da Saúde (OMS), a tuberculose é a nona causa de mortes no mundo e a primeira causa de mortes entre as doenças infectocontagiosas. O Brasil está entre os 20 países com maior número da doença, quando somados, esses países totalizam 84% dos casos de tuberculose no mundo. O grande volume de pesquisas e trabalhos realizados sobre a tuberculose possui um conhecimento consolidado que poderia ser extraído e aproveitado com o uso da tecnologia da informação para o provimento de informações auxiliando no trabalho de profissionais da saúde, bem como apoiando a criação de políticas e estratégias para o controle da tuberculose. Este estudo propõe o desenvolvimento de um portal *web* com marcação semântica para reunir e disponibilizar conhecimento, bem como indicadores públicos sobre a tuberculose. O software Drupal foi utilizado para a criação do portal com parte de seu conteúdo marcado semanticamente utilizando ontologias hospedadas em repositórios públicos na *web*, e por um repositório virtual de ontologias configurado na nuvem com o auxílio da plataforma D2RQ para padronização de indicadores da tuberculose. Um conjunto de nove indicadores utilizados no tratamento e prevenção da tuberculose foi selecionado para o estudo. A base de dados definida para testes foi a do Sistema de Gerenciamento de Pacientes da Tuberculose utilizado no município de Ribeirão Preto. Como resultado foi gerado um portal *web* de conhecimento reunindo informações e indicadores sobre a tuberculose marcada semanticamente. Também foi criado um servidor virtual de ontologias baseado na base de dados relacional do Sistema de Gerenciamento de Pacientes da Tuberculose. O estudo apresentou a utilização da *Web Semântica* na criação de portais de conteúdo com conhecimento para a tuberculose, objetivando uma experiência de uso enriquecedora aos usuários por meio de um portal inteligente capaz de entregar informações de maneira mais aderente às necessidades dos usuários e também a computadores que por meio de softwares inteligentes possam interpretar e entender seu conteúdo conforme preconizado pela *Web Semântica*.

Palavras-chave: Tuberculose. Web semântica. Indicadores de tuberculose.

ABSTRACT

LIMA, R. R de. **Development of a knowledge portal for tuberculosis based on semantic web**. 2018. 97 f. Thesis (Master in Sciences) – School of Engineering of São Carlos, – Medical School of Ribeirão Preto and Chemistry Institute of São Carlos, University of São Paulo, São Carlos, 2018.

According to the World Health Organization report of 2017, tuberculosis is the ninth leading cause of death in the world and the leading cause of death among infectious diseases. Brazil is among the 20 countries with the highest number of the disease, when added, these countries account for 84% of tuberculosis cases in the world. The large volume of research and work carried out on tuberculosis has a consolidated knowledge that could be extracted and harnessed with the use of information technology to provide information to assist the work of health professionals, as well as supporting the creation of policies and strategies for tuberculosis control. This study proposes the development of a web portal with semantic markup to gather and make available knowledge, as well as public indicators on tuberculosis. Drupal software was used to create the portal with part of its content marked semantically using ontologies hosted in public repositories on the web, and by a virtual repository of ontologies configured in the cloud using D2RQ platform for standardization of tuberculosis indicators. A set of nine indicators used in the treatment and prevention of tuberculosis were selected for the study. The database defined for testing was the Tuberculosis Patient Management System used in the city of Ribeirão Preto. As a result a web portal of knowledge was generated, gathering information and indicators about tuberculosis marked semantically. A virtual ontology server was also created based on the Tuberculosis Patient Management System relational database. The study presented the use of Semantic Web in the creation of content portals with knowledge for tuberculosis, aiming at an experience of enriching use to users through a smart portal capable of delivering information more closely to the needs of users and also to computers which through intelligent software can interpret and understand its content as recommended by the Semantic Web.

Keywords: Tuberculosis, Semantic Web. Tuberculosis Indicators.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo de Grafo RDF	22
Figura 2 – Notação RDF/XML do grafo RDF	22
Figura 3 – Ferramenta online SPARQL Explorer para execução de pesquisa na base DBPedia	24
Figura 4 – Resultado da consulta sobre a TB em SPARQL realizada por meio da SPARQL Explorer	25
Figura 5 – Diagrama OWL gerado no Protégé representando classe e subclasse.....	27
Figura 6 – Código OWL gerado automaticamente pelo Protégé a partir do diagrama..	28
Figura 7 – Sequência de execução das etapas relacionadas a este estudo	33
Figura 8 – Framework de interoperabilidade da TB	37
Figura 9 – Ecossistema proposto para o suporte ao gerenciamento de pacientes de Tuberculose	40
Figura 10 – Quadro kanban com a lista de atividades do projeto	47
Figura 11 – Seção de pesquisa de ontologias do portal Schema.org	48
Figura 12 – Conjunto de algumas propriedades relacionadas à entidade Person	49
Figura 13 – Repositório BioPortal de ontologias biomédicas	50
Figura 14 – Tela de apresentação da instalação do Drupal	51
Figura 15 – Tela de configuração do banco de dados do Drupal	51
Figura 16 – Representação gráfica do conjunto de nome-valor do Microdata	52
Figura 17 – Dado marcado na página por meio do código Microdata	53
Figura 18 – Quadro Kanban representando organização das tarefas do estudo	57
Figura 19 – Layout da tela inicial modificado pelo template qatar zymphonies Theme	58
Figura 20 – Tela principal do Portal TB com menu de manutenção do portal Ativo	60
Figura 21 – Portal TB na visão do usuário	60
Figura 22 – Portal TB na visão do computador	62
Figura 23 – Recorte do trecho do código da página do Portal TB com Microdata	63
Figura 24 – Página de cadastro de artigos com marcação semântica	65
Figura 25 – Proporção de casos de tuberculose testados para HIV	66
Figura 26 – Proporção de coinfeção TB/HIV	67

Figura 27 – Proporção de casos de tuberculose com HIV em andamento	68
Figura 28 – Proporção de casos de tuberculose curados	69
Figura 29 – Trecho de código destacado com a marcação Microdata do indicador “Proporção de casos de tuberculose curados”	70
Figura 30 – Apresentação das informações contidas nas estruturas de dados encontradas na página referentes ao indicador “Proporção de casos de tuberculose curados”	72
Figura 31 – Tela do servidor D2R de ontologias	72
Figura 32 – Vocabulário do indicador Proporção de casos de tuberculose curados	73
Figura 33 – Dados das variáveis utilizadas pelos indicadores	74
Figura 34 – Interface web para consultas SPARQL do D2R	75
Figura 35 – Consulta de listagem de indicadores filtrando o indicador “Proporção de casos de tuberculose curados”	75
Figura 36 – Consulta aos dados das variáveis associadas ao indicador “Proporção de casos de tuberculose curados”	76
Figura 37 – Representação da inclusão do Portal TB no Ecossistema SISTB	77

LISTA DE ABREVIATURAS E SIGLAS

AIDS	–	Síndrome da Imunodeficiência Adquirida
BFO	–	Basic Formal Ontology
CMS	–	Content Management Systems
CRUD	–	Create, Read, Update, Delete
CSS	–	Cascading Style Sheets
DNC	–	Doenças de Notificação Compulsório
DOTS	–	Tratamento Diretamente Observado de Curta
e-SUS AB	–	Atenção Básica do SUS
FTP	–	File Transfer Protocol
GAL	–	Gerenciador de Ambiente Laboratorial
GEOTB	–	Grupo de Estudos Epidemiológico-Operacional em Tuberculose
GPL	–	General Public License
HIV	–	Vírus da Imunodeficiência Humana
HTML5	–	Hyper Text Markup Language 5
HTTP	–	Hypertext Transfer Protocol
Hygia	–	Sistema para Gestão em Saúde Pública
IRIs	–	Internationalized Resource Identifiers
ICD-9	–	International Classification of Disease 9th
ICD-10	–	International Classification of Disease 10th
JSON	–	JavaScript Object Notation
LIS	–	Laboratório de Inteligência em Saúde
MS	–	Ministério da Saúde
ODPs	–	Ontology Design Patterns
OMS	–	Organização Mundial da Saúde
OWL	–	Web Ontology Language
PCT	–	Plano de Controle da Tuberculose
PHP	–	Hypertext Preprocessor
RDF	–	Resource Description Framework
RFDa	–	Resource Description Framework in Attributes
RNTCP	–	Revised National TB Control Programme

SBIS/CFM de Medicina	–	Sociedade Brasileira de Informática em Saúde e do Conselho Federal
SINAN	–	Sistema de Informação de Agravos de Notificação
SISTB	–	Sistema de Gerenciamento de Pacientes da Tuberculose
SPARQL	–	Simple Protocol and RDF Query Language
SQL	–	Structured Query Language
SRES	–	Sistemas de Registros Eletrônicos de Saúde
SVM	–	Support Vector Machine
SWRL	–	Semantic Web Rule Language
TB	–	Tuberculose
TI	–	Tecnologia da Informação
TBWEB	–	Sistema de Notificação e Acompanhamento dos Casos de Tuberculose
URI	–	Uniform Resource Identifier
URL	–	Uniform Resource Locator
XHTML	–	Extensible Hypertext Markup Language
XML	–	Extensive Markup Language
XSLT	–	Extensible Stylesheet Language for Transformation
WHO	–	World Health Organization
W3C	–	World Wide Web Consortium
WWW	–	World Wide Web
WYSIWYG	–	What You See Is What You Get

SUMÁRIO

1 INTRODUÇÃO.....	15
2 OBJETIVOS.....	19
2.1 Objetivo geral.....	19
2.2 Objetivo específico.....	19
3 FUNDAMENTÇÃO TEÓRICA.....	20
3.1 Web semântica.....	20
3.1.1 <i>RDF</i>	21
3.1.2 <i>SPARQL</i>	22
3.1.3 <i>OWL</i>	25
3.2 Revisão bibliográfica: O uso de Web semântica em Tuberculose.....	28
4 MATERIAIS E MÉTODOS.....	32
4.1 Tipo de pesquisa e etapas do estudo.....	32
4.2 Fonte de dados.....	34
4.3 Indicadores TB.....	38
4.4 SISTB – Um ecossistema em saúde.....	39
4.5 Ferramenta de desenvolvimento do portal: CMS Drupal.....	42
4.5.1 <i>CMS</i>	44
4.5.2 <i>Core do Drupal</i>	45
4.5.3 <i>Instituições que utilizam o Drupal</i>	45
4.6 Técnicas de desenvolvimento.....	46
4.7 Configuração de um servidor virtual de ontologias.....	54
5 RESULTADOS E DISCUSSÃO.....	56
6 CONCLUSÃO E TRABALHOS FUTUROS.....	78
REFERÊNCIAS.....	80
APÊNDICE A – Tela principal do portal (<i>Home</i>)	85
APÊNDICE B – Tela sobre do menu tuberculose.....	86
APÊNDICE C – Tela de cadastro de artigos e trabalhos.....	87
APÊNDICE D – Tela listagem de artigos e trabalhos.....	88
APÊNDICE E – Tela exibindo dados do artigo/trabalho e o link do arquivo.....	89
APÊNDICE F – Tela exibindo o indicador de proporção de pacientes curados.....	90
APÊNDICE G – Tarefas gerenciadas pelo Kanbanchi Task & Project Management.....	91

1 INTRODUÇÃO

No âmbito da saúde, de uma maneira geral, as informações e os dados coletados, sistematizados, resultantes das pesquisas experimentais, das pesquisas clínicas, das pesquisas epidemiológicas e da prática profissional em saúde são a base para a construção do conhecimento nessa área que, por sua vez, deve subsidiar o estabelecimento de práticas e políticas adequadas, voltadas para a população (BITTAR et al., 2018). Dessa forma, a capacidade de "executar" a gestão do conhecimento em saúde, também depende de uma inteligência organizacional de saúde que, em primeira instância, deve promover a sinergia entre as atividades de ciência e tecnologia da informação reunindo mecanismos e instrumentos, bem como metodologias e formas de organização, que possam incrementar a capacidade de promover as atividades em redes (BRIDGES et al., 2011).

O desafio a ser enfrentado é integrar e articular as informações em saúde, que facilmente podem subsidiar o conhecimento de ações e serviços nessa área e de seu impacto na situação de saúde da população (BITTAR et al., 2018). Para isso, o armazenamento computacional dos dados relacionados com um determinado evento (individual ou coletivo) em saúde é uma estratégia fundamental que tem como objetivo a melhoria da qualidade e eficiência dos processos de trabalho, automaticamente alimentando a cadeia de informações e eliminando todos os demais instrumentos paralelos de coleta relativos a este evento (ARENS et al., 1993).

Parte essencial para o sucesso desse desafio é manter a pertinência acadêmica, tendo em vista a assistência em saúde enquanto serviço. Todavia é importante frisar que essa estratégia tem de estar imbuída nas necessidades e dificuldades encontradas no fluxo de trabalho diário da assistência em saúde, sem nunca esquecer que as metodologias e produtos propostos estejam alinhados com as possibilidades das instituições parceiras.

A partilha de informação e experiências entre grupos de investigação e/ou entre indivíduos singulares permitem a criação de equipes multidisciplinares para a resolução de problemas e torna efetivamente possível posicionar os utilizadores dos Sistemas de Informação em Saúde (estudantes, profissionais de Tecnologia da Informação (TI), profissionais de saúde e doentes) no centro das soluções tecnológicas e organizacionais necessárias a melhoria dos produtos e serviços (CHOI; PAK, 2006).

Tal intensificação aumenta, então, a necessidade de adquirir novo conhecimento e novas competências, mas diminui a quantidade de tempo disponível para aprender. Assim, a

introdução de tecnologia colaborativa facilita a aprendizagem ao permitir que indivíduos (pesquisadores ou profissionais de saúde) trabalhem em conjunto no desenvolvimento do seu conhecimento e das suas competências, ao melhorar o acesso aos recursos informacionais e ao potenciar a comunicação (STAHL, 2006).

Nesse contexto, é importante desenvolver um ambiente colaborativo livre que permite aos múltiplos utilizadores trabalharem em grupo de forma síncrona ou assíncrona. Um ambiente computacional assim definido tem o potencial de complementar, melhorar e adicionar novas dimensões colaborativas a qualquer base de educação e/ou serviço de investigação atual. A sua arquitetura de participação deve oferecer novas formas de aprender, que estão muito mais em linha com as necessidades de aprendizagem atuais, tanto em contextos académicos, como fora destes (processo de aprendizagem informal), promovidos ao longo da vida.

Aliar essas características únicas às potencialidades da *Web Semântica* permite a representação de associações entre as partes das informações outrora desconectadas, o desenvolvimento de novos mecanismos de pesquisa, a recuperação estruturada de informação, a reutilização automática de conteúdos e a visualização da informação como nas bases de dados.

Vale destacar que em 2001, em um artigo na revista *Scientific American*, foram lançadas as bases da *Web Semântica* (BERNERS-LEE; HENDLER; LASSILA, 2001). Desde então, a *Web* de Documentos, a que utiliza a Internet para a consulta e a atualização de itens de informação (documentos) organizados em uma estrutura hipertextual, pode evoluir para uma *Web* Programável, a que executa códigos de programas e, a partir do processamento desses códigos, retorna um conteúdo gerado dinamicamente por esses programas (WORLD WIDE WEB CONSORTIUM, 2018a). Assim, inicia-se a crescente oferta de aplicações que executam tarefas de manipulação de dados publicados das mais diversas formas nesse ecossistema de dados na *Web*. Por fim, chegou-se a ideia de agregar semântica aos dados visando facilitar o entendimento e a interoperabilidade dos mesmos nesse universo de informações heterogêneas, publicadas nos mais diferentes formatos e com diferentes protocolos de acesso. Os blocos básicos que definem a *Web Semântica* são:

- Um modelo de dados padrão;
- Um conjunto de vocabulários de referência;
- Um protocolo padrão de consulta.

A *Web Semântica*, portanto, busca facilitar o processo de comunicação entre os diversos participantes do ecossistema, de forma a criar um modelo mental comum,

minimizando a possibilidade de ambiguidades e facilitando, assim, o trabalho necessário para o desenvolvimento de aplicações que manipulem as diversas fontes de dados.

A Tuberculose (TB) é uma doença de emergência mundial que em 2015 tornou-se a principal causa de mortes por doenças infectocontagiosas, ficando a frente do vírus da imunodeficiência humana (HIV/AIDS) (WORLD HEALTH ORGANIZATION, 2016a). Causada pelo *Mycobacterium tuberculosis*, bacilo de Koch, afeta principalmente os pulmões (TB pulmonar), mas em alguns casos podem afetar outros órgãos (TB extrapulmonar). A transmissão da doença acontece através do ar quando uma pessoa com tuberculose pulmonar expele gotículas pela respiração, espirros ou tosse (WORLD HEALTH ORGANIZATION, 2016a). Segundo a Organização Mundial da Saúde (OMS), um terço da população mundial está infectado pelo bacilo de Koch, e em risco de desenvolver a doença. Em 2015, estimou-se 1,4 milhões de mortes e mais 0,4 milhões de óbitos de pessoas com tuberculose e infectadas com HIV soro positivo. A incidência estimada foi de 10,4 milhões de casos novos, incluindo 1,2 milhões de HIV soro positivo. A população masculina apresenta um maior número de indivíduos infectados, totalizando 5,9 milhões de homens, soma-se ainda aos números, 3,5 milhões de mulheres e 1 milhão de crianças (WORLD HEALTH ORGANIZATION, 2016a).

No mesmo ano, o Brasil apresentou uma incidência estimada de 84 mil casos, equivalente a 41 casos por 100 mil habitantes, ocupando a 20ª posição entre os 30 países responsáveis por 87% da carga da doença no mundo, e embora seja tratável e curável, apresentou elevada taxa de abandono (11,8%) e óbito (3,5%), e baixa taxa de cura (71%) (WORLD HEALTH ORGANIZATION, 2016b).

O acompanhamento do tratamento da TB envolve diversos instrumentos padronizados dentre os quais, a Ficha de Notificação, o Livro de Registro e Acompanhamento de Tratamento dos Casos de Tuberculose e a Ficha de Registro do Tratamento Diretamente Observado de Curta Duração (DOTs) (BRASIL, 2011), além do prontuário clínico e o sistema de informação, Sistema para Gestão em Saúde Pública (Hygia), existente no município de Ribeirão Preto. Nesse sentido, as diversas formas de registro podem gerar duplicidade de dados, informações contraditórias e/ou a perda dos dados, não garantindo necessariamente a integralidade da assistência. Portanto, notam-se diferentes maneiras de acompanhamento e coleta de dados de pacientes no âmbito da TB. Neste cenário com diversas ferramentas independentes em utilização, muitas vezes com o mesmo propósito, naturalmente, provoca a pulverização das informações em diferentes repositórios de dados e dificulta o acesso aos dados, que de certa forma, poderiam corroborar na extração e entrega do conhecimento que estes carregam.

Neste sentido a monografia aqui apresentada, discute a construção de um portal de conteúdo que propõe a integração dessas informações, colocando-as à disposição dos usuários e serviços de saúde, por meio de um conteúdo estruturado utilizando as bases da *Web Semântica* e que seja capaz de extrair conhecimento das informações apresentadas (LIMA, 2018a). Para reunir um grande volume de informações, o portal, além de sua própria base de dados, responsável pelo armazenamento dos dados de gerenciamento do portal, propõe o acesso a um conjunto de base de dados estruturados e padronizados semanticamente. Tais bases compreendem a sistemas como o Sistema de Gerenciamento de Pacientes da Tuberculose (SISTB), Sistema de Notificação e Acompanhamento dos Casos de Tuberculose (TBWEB), Sistema de Informação de Agravos de Notificação (SINAN), Gerenciador de Ambiente Laboratorial (GAL) e Hygia, por meio das quais poderão ser obtidos, entre outras informações, indicadores capazes fornecer dados que apoiem políticas e ações no controle da TB.

Uma vez que, as informações do portal estarão estruturadas semanticamente baseadas em ontologias e seus vocabulários, os resultados das buscas por informações dentro do portal, recurso que será oferecido por padrão, naturalmente, apresentará o conteúdo filtrado e organizado com uma precisão maior, buscando entregar ao usuário um conteúdo com conhecimento agregado. Diferentemente de buscadores como Google e outros motores de busca, um portal estruturado semanticamente sabe exatamente o que mostrar de acordo com as buscas dos usuários, reduzindo a quantidade de conteúdo não relacionado ao assunto pesquisado e consequentemente diminuindo a necessidade de aplicações de diferentes filtros pelos usuários.

Por fim, alinhado aos princípios da *Web Semântica*, por meio da marcação dos dados com as respectivas ontologias, o portal viabilizará que o conhecimento e o significado de seu conteúdo, sejam entendíveis e interpretados igualmente tanto por humanos quanto por computadores.

2 OBJETIVOS

2.1 Objetivo geral

Desenvolver um portal *web* de conhecimento baseado nos conceitos e tecnologias da *Web Semântica* com a finalidade de reunir informações relacionadas à TB provenientes de trabalhos e pesquisas, bem como um conjunto de indicadores que poderão ser extraídos das bases de dados acessadas pelo portal.

2.2 Objetivo específico

1. Criar um portal baseado no conhecimento inteligente existente no contexto da TB com um conteúdo organizado e estruturado por meio das tecnologias da *Web Semântica*.
2. Estabelecer um método que apresente os passos necessários à criação de um portal com dados estruturados semanticamente.
3. Avaliar a viabilidade de utilização de um software gerenciador de conteúdo, Drupal, para a geração de conteúdo marcado semanticamente quanto aos recursos disponibilizados pela ferramenta.
4. Disponibilizar indicadores associados às ontologias, restritos à cidade de Ribeirão Preto, sobre a TB baseados no Manual de Recomendações para o Controle da Tuberculose no Brasil (MINISTÉRIO DA SAÚDE, 2011).

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Web semântica

A atual *Web* de Documentos, ou simplesmente *Web*, refere-se a um conjunto de arquivos de diferentes formatos (texto, imagem, vídeos e sons) que juntos compõem o documento com o conteúdo de uma página *web*. Estes documentos são armazenados em servidores *web* e podem ser requisitados pelo usuário por meio de aplicações clientes como os navegadores (LEE, 1990). No entanto, esses documentos *web*, hospedados em servidores ao redor do mundo, na maioria dos casos não se comunicam, criando um universo de informações repetidas, fragmentadas e desconectadas. Portanto, a *Web Semântica*, além de utilizar os elementos e recursos da *Web* de Documentos, tem por objetivo conectar repositórios de dados desconectados e espalhados pela *web* permitindo a construção de sistemas computadorizados que possam interoperar e reaproveitar não somente os dados, mas o conhecimento existente nestes (SINACI; ERTURKMEN, 2016). A *Web Semântica* viabiliza a representação do conhecimento por meio da criação de modelos de dados estruturados que representam um determinado contexto estudado e pela criação de vocabulários formais, por meio das ontologias, que contém definições individuais dos componentes dos modelos (BERNERS-LEE; HENDLER; LASSILA, 2001; WORLD WIDE WEB CONSORTIUM, 2018a). Desta forma, torna-se possível que todo o conhecimento humano existente em um dado domínio, também possa ser estendido e entendido por computadores.

Existe um conjunto de tecnologias que juntas formam os pilares da *Web Semântica*. Neste conjunto está o *Resource Description Framework* (RDF) que representa o relacionamento entre os dados baseados no modelo de grafos. O *Simple Protocol and RDF Query Language* (SPARQL) é a linguagem utilizada para pesquisa em repositórios RDF e a *Web Ontology Language* (OWL) para definição das ontologias e vocabulários responsáveis por carregar o conhecimento dentro de um determinado domínio mapeado.

Um dos mais conhecidos portais que pode ser citado como exemplo de utilização da *Web Semântica* é o portal DBPedia. O DBPedia é um portal de conteúdo diversificado mantido pela comunidade, que através da utilização das tecnologias bases da *Web Semântica* estruturou e organizou o conhecimento da enorme base de dados do Wikipédia. Atualmente, existem quase cinco milhões de itens mapeados no portal DBPedia por meio de ontologias (DBPEDIA, 2018a).

3.1.1 RDF

O RDF é um *framework* utilizado para representar a informação na *Web*. É um padrão utilizado para descrição e representação de como os dados estão interconectados. É recomendado pelo *World Wide Web Consortium* e, atualmente, encontra-se em na versão 1.1. Sua sintaxe abstrata tem duas estruturas chaves: os grafos RDF e os *datasets* RDF. Os grafos, também chamados de triplas, são conjuntos formados por sujeito-predicado-objeto (*subject*, *predicate*, *object*), estes elementos do grafo podem ser *Internationalized Resource Identifiers* (IRIs), nós em branco e literais. Os *datasets* são os conjuntos de dados RDF utilizados para organizar coleções de grafos RDF (WORLD WIDE WEB CONSORTIUM, 2018b).

Inicialmente, a WWW foi desenvolvida para o consumo de dados por pessoas. No entanto, computadores ou máquinas também podem consumir dados da *web*, porém não conseguem entendê-los. Com a finalidade de permitir que máquinas fossem capazes de entender os dados que consomem, foi proposta a utilização de metadados para descrever os dados na *web*. Os metadados são dados que carregam informações sobre um determinado dado.

Na *Web Semântica*, o RDF é a base para o processamento de metadados, através dele as aplicações podem trocar dados entre as máquinas. É importante destacar, que o modelo RDF de um determinado contexto seja passível de reutilização. A ideia defendida é que um modelo seja estendível por meio dos metadados, podendo ser incrementado para reutilização em diferentes cenários. O RDF assemelhasse-se a um sistema de classes organizadas numa hierarquia de classes e subclasses. Esta organização de classes é chamada de *schema*. Os *schemas* RDF possuem *namespaces* que os identificam univocamente por meio de um *Uniform Resource Identifier* (URI), e também carregam um conjunto de propriedades que definem o vocabulário dos elementos inseridos no contexto representado (WORLD WIDE WEB CONSORTIUM, 2018b).

Tais propriedades podem ser consideradas como atributos de recursos através de pares de valores (atributo-valor), bem como representação dos relacionamentos entre os recursos do modelo. O modelo básico do RDF assemelha-se a um diagrama de entidade-relacionamento e consiste em três tipos de objetos:

- Sujeito ou Recurso – Todas as coisas representadas no RDF são recursos. Por exemplo, um recurso pode ser uma página *web*, um computador, uma pessoa ou animal. Sempre são identificados de maneira unívoca através de URIs;

- Predicado ou Propriedade – São características específicas ou relações utilizadas para descrever o recurso. Cada propriedade tem um significado específico, valores permitidos, descreve tipos de recursos e o relacionamento com demais propriedades;
- Objeto ou Literal – O valor da propriedade pode ser outro recurso (especificado por um URI) ou um literal (como uma *string* ou outro tipo primitivo de dados definido por *Extensive Markup Language* (XML));

Os modelos RDF representados por grafos (Figura 1) são representados textualmente pela notação RDF/XML (Figura 2).

Figura 1 – Modelo de grafo RDF

Fonte – Própria do autor

Figura 2 – Notação RDF/XML do grafo RDF

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://w3.org/TR/1999/PR-rdf-syntax-19990105#"
  xmlns:sch="http://repositorio.org/schema/">
  <rdf:Description about="http://www.repositorio.org/medico">
    <sch:CRM>numero_crm</sch:CRM>
    <sch:Prescreve>
      <rdf:Description about="http://www.repositorio.org/medicamento">
      </sch:Prescreve>
    </rdf:Description>
  </rdf:RDF>
```

Fonte – Própria do autor

3.1.2 SPARQL

A recuperação de informações dos repositórios que armazenam dados baseados no padrão RDF é possível por meio da utilização da linguagem de pesquisa SPARQL. Recomendada pelo *World Wide Web Consortium*, sua versão atual é a 1.1. SPARQL tem a

capacidade de extrair informações de diferentes fontes de dados RDF. Seus códigos e estrutura assemelham-se a *Structured Query Language* (SQL), linguagem de pesquisa padrão das bases relacionais, oferecendo suporte a recursos como *join*, agregação de dados, subqueries, entre outros, bem como alguns comandos similares como *select*, *where*, *union* entre outros (WORLD WIDE WEB CONSORTIUM, 2018c). Utilizando a SPARQL é possível:

- Efetuar busca de dados através de diversos repositórios, seja localmente ou espalhados pela *web*;
- Realizar pesquisas através de padrões de grafos, obrigatórios ou opcionais, aplicando junções entre os diversos modelos de representação de informações;
- Aplicação de testes de valores;
- Aplicação de restrições de pesquisas;
- Representação dos resultados de pesquisas em formato de conjunto de dados ou em grafos;

Ferramentas *online* apresentam interface *web* que permitem a execução e testes de sentenças SPARQL, como o Virtuoso SPARQL *Query Editor* e o SNORQL/SPARQL *Explorer for* <http://dbpedia.org/sparql>. Tais ferramentas conectam-se a diversos repositórios de ontologias espalhados pela *web* e são capazes de acessarem e extraírem informações das estruturas RDF desses repositórios. A similaridade da SPARQL com a SQL, pode tornar seu aprendizado uma tarefa menos árdua para os desenvolvedores. A ferramenta SNORQL/SPARQL *Explorer*, apresenta uma interface simplificada que exhibe os repositórios a qual a ferramenta se conecta, uma área para entrada da consulta de SPARQL e uma área de resultados. Por exemplo, uma sentença SPARQL é apresentada abaixo buscando algumas informações relacionadas à TB apontando para diferentes repositórios (Figura 3).

Figura 3 – Ferramenta online SPARQL *Explorer* para execução de pesquisa na base DBPedia

SPARQL Explorer for http://dbpedia.org/sparql

SPARQL:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

```
select ?name ?desc ?cid

where {?resource a dbo:Disease .
       ?resource foaf:name ?name .
       ?resource dbo:abstract ?desc .
       ?resource dbo:icd10 ?cid

filter (lcase(str(?name)) = "tuberculosis")
filter (langMatches(lang(?desc), "EN"))}
```

Results:

Fonte – SPARQL Explorer (2018b)

Acessada por meio de um navegador web, a Figura 3 apresenta os endereços (*Prefix*) dos repositórios conectados, a consulta SPARQL e uma opção *Results* que permite ao usuário informar a saída do resultado em diferentes formatos, neste exemplo, foi selecionado o formato *browse* que apresenta a saída de resultado no próprio navegador. Também é possível produzir resultado em JavaScript *Object Notation* (JSON), XML e XML + *Extensible Stylesheet Language for Transformation* (XSLT).

A consulta relaciona os dados nome (*name*), descrição (*desc*) e cid (*cid*) e filtra o resultado da pesquisa no idioma Inglês (EN). As fontes de dados relacionadas neste exemplo, possuem um conjunto maior de informações definidas a respeito da TB, no entanto, apenas para ilustração, a área de resultados apresenta apenas os três campos conforme definido na consulta (Figura 4).

Figura 4 - Resultado da consulta sobre a TB em SPARQL realizada por meio da SPARQL Explorer

SPARQL results:		
name	desc	cid
"Tuberculosis"@en	"Tuberculosis (TB) is an infectious disease caused by the bacterium <i>Mycobacterium tuberculosis</i> (MTB). Tuberculosis generally affects the lungs, but can also affect other parts of the body. Most infections do not have symptoms, known as latent tuberculosis. About 10% of latent infections progress to active disease which, if left untreated, kills about half of those infected. The classic symptoms of active TB are a chronic cough with blood-containing sputum, fever, night sweats, and weight loss. The historical term "consumption" came about due to the weight loss. Infection of other organs can cause a wide range of symptoms. Tuberculosis is spread through the air when people who have active TB in their lungs cough, spit, speak, or sneeze. People with latent TB do not spread the disease. Active infection occurs more often in people with HIV/AIDS and in those who smoke. Diagnosis of active TB is based on chest X-rays, as well as microscopic examination and culture of body fluids. Diagnosis of latent TB relies on the tuberculin skin test (TST) or blood tests. Prevention of TB involves screening those at high risk, early detection and treatment of cases, and vaccination with the bacillus Calmette-Guérin vaccine. Those at high risk include household, workplace, and social contacts of people with active TB. Treatment requires the use of multiple antibiotics over a long period of time. Antibiotic resistance is a growing problem with increasing rates of multiple drug-resistant tuberculosis (MDR-TB). One-third of the world's population is thought to be infected with TB. New infections occur in about 1% of the population each year. In 2014, there were 9.6 million cases of active TB which resulted in 1.5 million deaths. More than 95% of deaths occurred in developing countries. The number of new cases each year has decreased since 2000. About 80% of people in many Asian and African countries test positive while 5–10% of people in the United States population tests positive by the tuberculin test. Tuberculosis has been present in humans since ancient times."@en	"A15–A19"

Fonte – SPARQL Explorer (2018b)

3.1.3 OWL

A OWL é a linguagem lógica da *Web Semântica* que representa o conhecimento sobre as coisas e as relações que existem entre elas. Utilizada para criação de ontologias, é possível realizar o mapeamento do conhecimento de um contexto estudado e criar documentos OWL que armazenam as ontologias que tornam o conhecimento passível de entendimento pelos computadores. Desenvolvida e mantida pelo *World Wide Web Consortium*, atualmente está na versão 2 (WORLD WIDE WEB CONSORTIUM, 2018d).

Ao longo do tempo, a *Web Semântica* tem se desenvolvido com base em um conjunto de tecnologias, tais como XML, RDF e RDF *Schema* (RDFS), cada qual com suas finalidades e limitações. Contudo, adotou-se a OWL como a principal tecnologia em função de sua facilidade de expressão e construção semântica, além da possibilidade de construção de vocabulários com maior riqueza de detalhes. A OWL divide-se em três outras sublinguagens, OWL *Lite*, OWL *DL* e OWL *Full*. Na respectiva ordem, cada sublinguagem é uma extensão de sua predecessora, ou seja, as especificações da OWL *Lite* são contempladas pela OWL *DL*, conseqüentemente, as especificações da OWL *Lite* juntamente com a OWL *DL* são contempladas na OWL *Full*. Contudo, é importante observar que esta regra não se aplica na ordem inversa (WORLD WIDE WEB CONSORTIUM, 2018d).

A OWL oferece recursos que permitem o mapeamento lógico e a formalização de domínios do conhecimento por meio das ontologias. Adicionalmente à definição de classes e propriedades, a OWL trata dos relacionamentos entre classes, equivalências de classes e propriedades, gerência de múltiplas classes que estão representando as mesmas coisas (recurso *sameAs*, em português, *mesmoQue*) entre outros recursos que complementam a formalização de ontologias. RDF e OWL são diferentes. A primeira tecnologia é estritamente a descrição estrutural dos dados representados, a segunda é um pouco mais completa, pois, também define o vocabulário específico do conhecimento mapeado. Soma-se a este vocabulário todo conjunto de regras, restrições e relações com outras ontologias de outros domínios de conhecimento, seja no mesmo ou em repositórios diferentes (WORLD WIDE WEB CONSORTIUM, 2018e; TEIXEIRA, 2016).

A criação de ontologias envolve o processo de definição de vocabulários completos e complexos de maneira formal de um conjunto de termos utilizados dentro de um domínio de conhecimento. Não existe um tamanho padrão para estes vocabulários, isto depende do domínio representado, podendo ter poucos ou milhares de termos definidos. Posteriormente, estes vocabulários são utilizados para realização de inferências. São finalidades também atribuídas às ontologias, definir regras, restrições, relacionamentos e descrição de contextos, integração de dados e eliminação de ambiguidades entre repositórios diferentes. Ao aplicar estas funcionalidades, por exemplo, considerando dois repositórios de ontologias, um sobre medicina com o mapeamento de tratamentos relacionados às doenças infectocontagiosas e outro repositório sobre medicamentos, por meio de programas computacionais é possível consumir estes repositórios integrando informações sobre as formas de tratamento das doenças e os medicamentos padronizados que podem ser prescritos, com respectivas posologias, frequências, interações medicamentosas entre outras informações relevantes condizentes às melhores práticas e cuidados associados ao tratamento de uma doença específica (WORLD WIDE WEB CONSORTIUM, 2018f).

Ontologias são elementos de representação da semântica de diferentes domínios, uma vez definidas, são publicadas em repositórios espalhados pela *web*, podendo ser consumidas de maneira inteligente por meio de computadores, *tablets*, *smartphones* e demais equipamentos eletrônicos (WORLD WIDE WEB CONSORTIUM, 2018e; TEIXEIRA, 2016).

Segundo Lima e Carvalho (2005), o elemento básico de uma ontologia é a classe, esta representa de uma maneira geral um grupo de objetos, indivíduos ou coisas que compartilham das mesmas características, como por exemplo, a classe Homem e a classe Mulher são

representados pela superclasse Pessoa. As ontologias são definidas por uma hierarquia de classes que podem ser chamadas de superclasses ou subclasses. Para qualquer ontologia e qualquer domínio representado, a classe *Thing*, coisa em Inglês, é a superclasse que sempre estará no topo de todas as classes (WORLD WIDE WEB CONSORTIUM, 2018e). A diferenciação dessas classes ocorre em função das características que as definem, tais características definidas pelos elementos chamados de propriedades. O Protégé, desenvolvido pelo Centro de Pesquisa de Informática Biomédica da Universidade de Stanford, é uma das mais conhecidas e utilizadas ferramentas para criação de ontologias para sistemas inteligentes. Por meio de sua utilização, é possível criar diagramas de ontologias com suas classes, subclasses e suas respectivas propriedades, bem como os vocabulários, relacionamentos e demais recursos (Figura 5).

Figura 5 – Diagrama OWL gerado no Protégé representando classe e subclasse

Fonte – Própria do autor

A Figura 5 mostra uma simples modelagem que tem a classe *Thing*, base de todos diagramas OWL, em seguida, a classe Pessoa de onde derivam as subclasses Mulher e Homem, e por último a classe Humano que está relacionada com a classe Pessoa. O conhecimento representado no diagrama pode ser interpretado da seguinte maneira: homem e mulher são pessoas, pessoas são humanos, logo mulher e homem também são humanos. O Protégé oferece a funcionalidade de exportação de diagramas para a linguagem OWL (Figura 6).

O diagrama que é transformado em código OWL pelo Protégé, conforme no exemplo da Figura 5, revela além das classes e subclasses, as propriedades associadas e seus relacionamentos uma com as outras através de suas URIs.

Figura 6 – Código OWL gerado automaticamente pelo Protégé a partir do diagrama

```

@prefix : <http://purl.obolibrary.org/obo/TEMP#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@base <http://purl.obolibrary.org/obo/TEMP> .
<http://purl.obolibrary.org/obo/TEMP> rdf:type owl:Ontology .

#####
#   Object Properties
#####

### http://purl.obolibrary.org/obo/TEMP#TemBoca
:TemBoca rdf:type owl:ObjectProperty ;
        rdfs:domain :Pessoa .
### http://purl.obolibrary.org/obo/TEMP#TemCabelo
:TemCabelo rdf:type owl:ObjectProperty ;
        rdfs:domain :Pessoa .
### http://purl.obolibrary.org/obo/TEMP#TemMembros
:TemMembros rdf:type owl:ObjectProperty ;
        rdfs:domain :Pessoa .
### http://purl.obolibrary.org/obo/TEMP#TemNariz
:TemNariz rdf:type owl:ObjectProperty ;
        rdfs:domain :Pessoa .
### http://purl.obolibrary.org/obo/TEMP#TemOlhos
:TemOlhos rdf:type owl:ObjectProperty ;
        rdfs:domain :Pessoa .
### http://purl.obolibrary.org/obo/TEMP#TemOrelhas
:TemOrelhas rdf:type owl:ObjectProperty ;
        rdfs:domain :Pessoa .

#####
#   Classes
#####

### http://purl.obolibrary.org/obo/TEMP#Homem
:Homem rdf:type owl:Class ;
        rdfs:subClassOf :Pessoa .
### http://purl.obolibrary.org/obo/TEMP#Humano
:Humano rdf:type owl:Class ;
        owl:equivalentClass :Pessoa .
### http://purl.obolibrary.org/obo/TEMP#Mulher
:Mulher rdf:type owl:Class ;
        rdfs:subClassOf :Pessoa .
### http://purl.obolibrary.org/obo/TEMP#Pessoa
:Pessoa rdf:type owl:Class .

```

Fonte – Própria do autor

A partir do código, sistemas inteligentes são capazes de interpretar e processar o conhecimento mapeado pelas ontologias.

3.2 Revisão bibliográfica: o uso de web semântica em tuberculose

Conforme apontado anteriormente a *Web Semântica* não é um conceito novo de tecnologia. Desde o seu lançamento em 2001, seu enorme potencial para a construção de

bases de conhecimento ainda foi muito pouco explorado. Sua utilização pode ser aplicada em qualquer área de conhecimento, e para o desenvolvimento deste trabalho cujo foco é a área da saúde, especificamente, tuberculose, foram realizadas buscas de trabalhos referentes à tuberculose que fazem uso da *Web Semântica* ou das tecnologias a compõe. A combinação de palavras chaves utilizadas na busca por trabalhos científicos foram: *Semantic Web*, *Ontologies*, *Ontology and Tuberculosis*. Também foram estudados os conceitos e tecnologias da *Web semântica*.

Abhishek e Singh (2016), dedicaram seus esforços na construção de ontologias para dar suporte à tomada de decisão ao controle e gestão da tuberculose do programa nacional indiano *Revised National TB Control Programme* (RNTCP). Em 1997, o governo indiano estabeleceu o RNTCP que implementa as recomendações do *WHO STOP TB Strategy* da OMS (WORLD HEALTH ORGANIZATION, 2006). O programa abrange os cinco níveis de atenção ao paciente na Índia e cada nível tem seus responsáveis para por em prática as estratégias recomendadas. Apesar de seu pleno funcionamento, a estrutura do RNTCP é complexa e dificulta a obtenção, integração, troca e reuso das informações dos pacientes acompanhados. Diante deste cenário, utilizando ontologias, um dos pilares da *Web Semântica*, os autores desenvolveram um *framework* para criar um ambiente integrado de informações dentro do contexto do programa de combate a Tuberculose que auxilie os profissionais indianos na tomada de decisão. Tal *framework*, fundamentado na representação da *Basic Formal Ontology* (BFO) (BASIC FORMAL ONTOLOGY, 2018), mapeia na forma de ontologias todos os níveis de atenção do RNCTP, desde o nível gestor ao nível de atendimento. Também foram criadas e relacionadas ontologias com sinais e sintomas característicos da Tuberculose que permite a identificação da doença no paciente e no caso deste estar infectado, o classifica dentro de uma das categorias definidas por este sistema de apoio. Com auxílio de algoritmos criados a partir da linguagem *Semantic Web Rule Language* (SWRL), é possível aplicar regras e fazer inferências na base de conhecimento para identificar se o paciente é portador da TB e indicar o tratamento adequado correspondente à categoria a qual está enquadrado. O sistema criado ainda tem seu uso restrito, pois não está dentro de uma interface de uso adequada, exigindo do profissional que fará uso, um conhecimento mais técnico referente às tecnologias utilizadas.

Pandyan et al. (2016) desenvolveram um trabalho de utilização de ontologias com *Support Vector Machine* (SVM) para auxílio em análises estatísticas e diagnósticos da Tuberculose. Baseado na dificuldade de diagnosticar a doença através de alguns métodos mais antigos que ainda estão em uso por muitos laboratórios, imprecisão de diagnósticos e que

novos métodos e tecnologias para a detecção da doença ainda são caras e não acessíveis a muitos países, desenvolveu este estudo que utiliza como base de conhecimento ontologias de teste de laboratórios da Tuberculose, dados geográficos e características da doença que com suporte da SVM, que através de algoritmos de aprendizagem, permite um diagnóstico mais rápido e preciso da doença. Defende os autores deste trabalho o baixo custo deste método de identificação da doença que poderá ser disponibilizado principalmente aos países que não possuem recursos financeiros para ter acesso às tecnologias contemporâneas. Outra aplicação proposta pelo trabalho é que através do conjunto de dados plotados na SVM, seja possível uma previsão futura de início da doença e emitir alertas avisando sobre este possível desenvolvimento da doença.

Harjito, Cahyani e Doewes (2016) propuseram a criação bilíngue de ontologias, nos idiomas Indonésio e Inglês, de maneira automática através da análise de conteúdos textos a respeito da Tuberculose, tais como artigos científicos, que aliada ao uso de modelos padrões existentes de criação de ontologias, denominados *Ontology Design Patterns* (ODPs). A gestão e o conhecimento dos colaboradores no combate a Tuberculose pode ser melhorado através das diversas fontes de informações que, atualmente, estão espalhadas em revistas científicas, livros, *web sites* entre outras fontes. Reforçando a ideia que as ontologias representam o domínio do conhecimento, tais fontes acima citadas servem de base para extração de informações para definição das ontologias dentro do domínio da Tuberculose. A criação de ontologias é um processo com alto grau de complexidade e necessita do envolvimento de diversos profissionais para validá-las. Aos autores deste trabalho afirmam que utilizando ODPs foram criadas ontologias sem a ajuda de profissionais especializados e com a ajuda de algoritmos para validação chegou-se a uma precisão de 71% quando comparado a outro modelo base existente de ontologias TB. Mesmo mostrando-se um processo ágil de definição de ontologias, ainda são necessárias melhorias que permitam atingir um maior grau de precisão destas.

Hitzler e Janowicz (2010), no trabalho intitulado “*Web Semântica – Interoperabilidade, Usabilidade, Aplicabilidade*”, destacam a *Web Semântica* como tendência atualmente. A *Web Semântica* é uma tecnologia multidisciplinar e heterogênea podendo ser aplicada a diversas áreas de conhecimento, mundialmente nota-se um número cada vez maior de eventos e pesquisadores das mais variadas disciplinas no desenvolvimento de trabalhos direcionados a esta tecnologia. Os autores destacam o *Semantic Web Journal*, um portal que concentra contribuições em diversos aspectos da *Web Semântica* que reúne relatórios, ferramentas, sistemas, aplicações e ontologias disponíveis ao público. Com um conceito

aberto de conteúdo, o portal apoia a alta qualidade de produção de conteúdo através de um processo de controle e revisão de trabalhos de forma pública e transparente, uma vez que, todos os trabalhos submetidos podem receber contribuições e comentários publicamente de revisores associados.

Valle et al. (2005) no trabalho denominado “A Necessidade de Serviço de *Web Semântica* no modelo eSaúde” destaca a importância da *Web Semântica* no processo de interoperabilidade entre sistemas de informação em saúde. Adotado universalmente como um dos padrões amplamente utilizados no mercado, estruturas baseadas na *Extensive Markup Language* (XML) e *web services* promovem com eficiência a interoperabilidade de sistemas informatizados heterogêneos. Contudo, tais tecnologias promovem apenas uma interação sintática desprovida do conhecimento proposto pelo nível semântico. O relacionamento dos elementos dentro estrutura semântica que carregam o conhecimento são determinados pelas ontologias que permitem que o entendimento humano sobre um determinado domínio seja estendido para os computadores. Atualmente, sistemas informatizados de saúde ainda são desenvolvidos de maneira proprietária ficando restrito seu uso apenas dentro da própria empresa, e apesar dos esforços para criação de padrões de troca de dados, cada padrão possui seu mecanismo de integração o que implica em criação de interfaces que permitam a comunicação entre eles. Tal dificuldade pode ser suprida através da *Web Semântica*, uma vez que, definida as ontologias de um determinado domínio de conhecimento, estas poderão ser consumidas pelos mais variados sistemas informatizados sem a necessidade de uma interface de comunicação específica e com a certeza de que todos estes sistemas receberão as mesmas informações.

4 MATERIAIS E MÉTODOS

4.1 Tipo de pesquisa e etapas do estudo

A metodologia de pesquisa aplicada a este projeto é a Pesquisa-ação (Action research). A Pesquisa-ação é a busca contínua de criação de melhorias dentro de determinado contexto observado. Podendo ser aplicada em diferentes áreas de pesquisa, a Pesquisa-ação é um tipo de investigação da prática de processos que ao mesmo tempo atua efetivamente dentro do processo buscando sua melhoria (TRIPP, 2005). O caráter investigativo e prático proposto pela metodologia adequa-se a este projeto, uma vez que, há o desenvolvimento da pesquisa e, no aspecto prático, a criação do portal da TB marcado semanticamente.

Para a melhor organização deste estudo, seu desenvolvimento foi definido em etapas que ajudaram a nortear o fluxo de trabalho, permitindo uma melhor compreensão do cenário estudado, bem como, conceitos, tecnologias, ferramentas, trabalhos e soluções já desenvolvidos e propostos no âmbito da *Web Semântica*. O conjunto de etapas realizadas contou com o "Levantamento bibliográfico", para exploração de artigos, trabalhos e estudo dos conceito e tecnologias da *Web Semântica*. Na etapa de "Seleção de Repositórios de Dados da TB", foram buscadas bases de dados da TB e, verificou-se a existência de grande volume de informações desconectadas baseadas em práticas, trabalhos e pesquisas que poderiam ser utilizadas e reutilizadas no fornecimento do conhecimento a ser modelado pela *Web Semântica*. Além do conteúdo de TB abordado de uma forma geral, este trabalho procurou enriquecer o conteúdo do portal por meio da disponibilização de indicadores públicos da TB, portanto, de acordo com o Manual de Recomendações para o Controle da Tuberculose no Brasil, conforme apresentado no "Anexo H – Indicadores de monitoramento e avaliações das ações de controle da TB" que apresenta um total de 24 indicadores. Foram selecionados nove indicadores possíveis de serem extraídos de acordo com os dados disponíveis para acesso por meio do portal (MS). A relação de indicadores está relacionada na seção de resultados. Para acrescentar conhecimento aos dados, na etapa de "Seleção de Ontologias", foram pesquisados repositórios públicos de ontologias, bem como a criação de um pequeno repositório de ontologias para os indicadores do portal, para efeito de elaboração de testes. Por último, em "Desenvolvimento do Portal", o *Content Management Systems* (CMS) Drupal foi estudado por meio de seu manual, disponibilizado no próprio repositório *web* do Drupal, e outros materiais relacionados com livros, *e-books*, vídeos e artigos. As etapas determinadas no

estudo estão esquematizadas conforme abaixo (Figura 7), e estão representadas numa visão ampla do contexto deste estudo.

Figura 7 - Sequência de execução das etapas relacionadas a este estudo

Fonte – Própria do autor

Essas etapas são adaptadas do trabalho de Abhishek e Singh (2016), onde os autores aliam o conceito de Sistema de Apoio à Decisão (do inglês *Decision Support System (DSS)*) e *Web Semântica* ou ainda a *Linguagem de Ontologia da Web*.

Particularmente, como vamos destacar a seguir o DSS inclui sistemas baseados em conhecimento que ajudam na tomada de decisões e na solução de problemas. A Taxonomia de DSS consiste em:

- DSS orientado por modelo: auxilia na tomada de decisões, analisando dados e parâmetros fornecidos pelo do utilizador;
- DSS orientado por dados: auxilia na tomada de decisões recuperando e manipulando dados;
- DSS orientado por documentos: ajuda na tomada de decisões analisando documentos não estruturados;
- DSS orientado pelo conhecimento: auxilia na tomada de decisão real com a ajuda de fatos, regras, procedimentos etc.

4.2 Fonte de dados

As informações do portal são compostas por dados relacionados à TB em um contexto amplo e geral, procurando concentrar de maneira, estruturada e organizada, um grande volume de informações relevantes ao tema, proporcionando ao usuário a experiência rica em conteúdo fundamentado no conhecimento existente ao assunto. Somam-se ao conteúdo do portal, a exibição de dados provenientes do acesso proposto às bases de dados de um conjunto de Sistemas de Registro Eletrônico de Saúde (SRES), atualmente utilizados no auxílio à gestão dos pacientes com TB. Estes dados serão apresentados no portal no formato de indicadores de saúde para a TB conforme comentado anteriormente. Tais sistemas utilizados nas diferentes esferas de governo, Municipal, Estadual e Federal, possuem repositórios de dados específicos, portanto, são independentes, não provendo a comunicação e interoperabilidade necessária de seus dados, que de alguma maneira, poderiam ser reutilizados e corroborarem nas melhorias de um conjunto de práticas e ações no combate à TB. As principais bases de dados responsáveis pela geração de informações que norteiam as ações de profissionais da saúde e pesquisadores da TB, são alimentadas pelos seguintes sistemas: SISTB, TBWEB, SINAN, GAL e Hygia.

O SISTB é um sistema que auxilia no controle e assistência dos pacientes da TB no município de Ribeirão Preto que entre outras finalidades, corrobora com o gerenciamento e monitoração do tratamento dos pacientes. Suas informações estão estruturadas com base nos dados dos pacientes, evolução sintomática da doença, processos do serviço de notificação, acompanhamento no programa de TB, exames realizados, administração dos medicamentos e

internações realizadas no início e durante o tratamento. Possui base de dados relacional que utiliza o MySQL como sistema de gerenciamento de banco de dados que armazena as informações centralizadas no paciente e no tratamento. A estrutura do banco torna-se possível interoperar com o sistema informatizado de saúde do programa de tuberculose do estado de São Paulo, TBWEB, a partir da compatibilidade semântica dos campos entre os dois sistemas. Dentre os repositórios de dados da TB relacionados, para o desenvolvimento deste estudo, o SISTB foi elencado como fornecedor de dados para a geração de indicadores apresentados pelo portal.

O TBWEB é o sistema do Estado de São Paulo que tem como objetivo a vigilância epidemiológica da TB no Estado de São Paulo e o monitoramento dos casos da doença. Por meio dele é possível cadastrar os pacientes e acompanhar a evolução dos tratamentos mediante o histórico de registros. Sua criação buscou solucionar a necessidade do Estado de centralizar em um único banco de dados os casos de TB, permitindo um melhor acompanhamento do paciente e eliminando a duplicidade de cadastros.

O SINAN é o Sistema de Informação de Agravos de Notificação, com abrangência em todo território brasileiro, de casos de doenças e agravos que constam na lista nacional de notificação compulsória, conforme a Portaria de Consolidação Nº 4, de 28 de Setembro de 2017, anexo V – Capítulo I, além de permitir a estados e municípios incluírem outras doenças regionais que são consideradas importantes. Sua utilização permite identificar a realidade epidemiológica das áreas investigadas, permitindo dessa maneira, o planejamento correto de intervenções junto à população.

Hygia é o sistema de gestão em saúde municipal da cidade de Ribeirão Preto. Está presente nas unidades de saúde do município e entre outras funções apresenta controle de agendamento de pacientes, regulação de encaminhamentos de pacientes, atendimentos ambulatoriais, internações de urgência, controle de vacinas, etc.

GAL é o sistema Gerenciador de Ambiente Laboratorial em âmbito nacional responsável por gerenciar o Sistema Nacional de Laboratórios de Saúde Pública das Redes Nacionais de Laboratórios de Vigilância Epidemiológica e Vigilância em Saúde Ambiental. Entre suas funcionalidades, permite o acompanhamento das etapas de realização de exames laboratoriais, permitindo o envio dos resultados de casos suspeitos ou confirmados, sejam eles, positivos ou negativos, das Doenças de Notificação Compulsória (DNC) ao SINAN.

Respeitadas as particularidades de cada sistema, todos poderiam em alguns aspectos, compartilhar informações semelhantes por meio da interoperabilidade entre sistemas. A interoperabilidade de sistemas possibilitando a troca de informações é um requisito desejável

nos sistemas atualmente alinhados com as evoluções de softwares e padrões tecnológicos (LIMA, 2018b). Este trabalho propõe a interoperabilidade utilizando a semântica dos dados, viabilizando a troca de conhecimento, eliminando a necessidade de recadastrar informações e consequentemente diminuir a redundância e a divergência de conteúdo. Dados que poderiam ser reaproveitados, como informações demográficas dos pacientes, alguns indicadores e dados de notificação da doença são apenas alguns exemplos. Em relação ao ponto de vista da *Web Semântica*, é possível criar ontologias a fim de mapear o conhecimento existente nas informações constantes nessas bases de dados. Uma vez criadas às ontologias, o passo seguinte é a marcação dos dados semanticamente utilizados no portal, permitindo o processamento inteligente das informações por meio de sistemas capazes de interpretar informações estruturadas. A ontologia padroniza a informações que podem ser trocadas e reaproveitadas por diferentes sistemas e portais.

Propõe-se neste trabalho que as bases de dados dos sistemas acima citados sejam as principais fontes de dados para a geração de indicadores ao portal. Para possibilitar o acesso aos dados das bases, foram identificadas duas vias possíveis: acesso direto ao banco e acesso intermediado por uma plataforma de interoperabilidade proposta (Figura 8).

Figura 8 - Framework de interoperabilidade da TB

Fonte – Adaptado de Pellison (2017)

O acesso direto ao banco é realizado por meio de usuários e senhas disponibilizados por cada uma das respectivas bases utilizadas dentro do contexto. Outra possibilidade de acesso é por meio da utilização de uma camada intermediária de interoperabilidade proposta que abstrai o acesso aos dados das diferentes fontes de dados relacionadas neste estudo. Esta camada é um *framework* que entre outras coisas, gerencia as conexões às bases de dados, e integra-se a um conjunto de sistemas que estão dentro de um ecossistema denominado ecossistema da tuberculose. O framework tem como principais características acesso seguro e controlado aos dados mediante utilização de *web services* e padronização semântica dos dados tornando-os passíveis de reutilização e integração entre os diferentes sistemas conforme apresenta a Figura 8 (PELLISON, 2017).

Particularmente, esta fase de integração de dados a partir de fontes heterogêneas é denominada como a construção de um *DSS orientado por dados*.

A Figura 8 acima apresenta o *framework* de integração entre vários sistemas, a partir do fluxo do atendimento do paciente com TB ou com suspeita de TB e a guarda das informações desses pacientes nos repositórios de dados. Também é apresentado a camada de interoperabilidade que abstrai as informações em comum entre as bases e as padroniza mediante a utilização de ontologias. A ontologias definidas utilizam o *framework* BFO na padronização de seu desenvolvimento (BASIC FORMAL ONTOLOGY, 2018). A camada de segurança e acesso aos dados controlam os acessos de sistemas que solicitam informações ao *framework* de interoperabilidade.

4.3 Indicadores TB

Para apoio a tomada de decisões é necessário o acesso a dados confiáveis que apoiem as instituições de saúde na criação de políticas e ações que corroborem com a melhoria dos serviços prestados à população, bem como na gestão das próprias instituições. Uma vez que os dados são registrados, somados aos avanços no controle das doenças infecciosas, compreensão do conceito de saúde e de seus determinantes sociais, considerando outras dimensões de avaliação da saúde, como por exemplo o acesso a serviços, qualidade da atenção, condições de vida entre outros, foram estabelecidos indicadores de saúde conforme Manual de Recomendações para o Controle da Tuberculose no Brasil para uma melhor avaliação das condições de saúde no país (RONDELLI, 2017).

Conforme mencionado anteriormente, o Manual de Recomendações para o Controle da Tuberculose no Brasil dispõe de 24 indicadores utilizados no apoio ao monitoramento e avaliações das ações de controle da TB. Tais indicadores podem ser utilizados para fornecer informações que permitam a avaliação das ações e programas adotados pelas instituições de saúde no controle da TB dos municípios. Como recomendação do manual, o primeiro passo é definir um responsável pelas ações do controle da TB no município que fará entre outras coisas, um levantamento da situação epidemiológica, análise da organização geral do sistema de saúde observando a rede de serviços existentes como unidades básicas de saúde, suporte à saúde das famílias, rede de referência de laboratórios e hospitais, situação socioeconômica e inserção do município no Plano de Controle da Tuberculose (PCT) (MNISTÉRIO DA SAÚDE, 2011).

A avaliação dos indicadores deve considerar o contexto onde foram obtidos, pois, diferentes fatores podem interferir na doença e nos programas de controle, como por exemplo, é importante levar em consideração as condições estruturais para a realização das

atividades, os recursos humanos, insumos e equipamentos a disposição, bem como a organização das atividades. A observação desse conjunto de condições viabiliza uma definição mais adequada de indicadores para acompanhamento dos programas adotados e metas propostas (MNISTÉRIO DA SAÚDE, 2011). Os indicadores podem ser classificados em:

- Indicadores de Processo: medem efeitos imediatos das ações em um curto espaço de tempo;
- Indicadores de Resultado: relacionados diretamente aos objetivos específicos, podendo ser avaliados em um espaço mediano de tempo;
- Indicadores de Impacto: relacionados ao objetivo geral, devendo ser avaliados em um longo espaço de tempo, no mínimo cinco anos.

Os indicadores selecionados para a realização deste trabalho foram mapeados em ontologias onde foram definidas estruturas de triplas juntamente com os metadados que definem o conhecimento a respeito de cada indicador. O processo de mapeamento será apresentado mais adiante na seção 4.7 deste trabalho. Os critérios para a escolha dos indicadores limitaram-se na capacidade de geração de dados pelas fontes de dados utilizadas pelo portal e na adaptação do trabalho de Rondelli (2017) que desenvolveu um estudo para geração de indicadores para a TB, no entanto, o autor não considerou a utilização de *web semântica*.

Dessa maneira, esta fase de extração de conhecimento a partir documentos normativos e *guidelines* esta relacionada à construção de *DSS orientado por documentos* e *DSS orientado por conhecimento*.

4.4 SISTB – Um ecossistema em saúde

Particularmente, um ecossistema em saúde pode ser visto como um conjunto de sistemas de informação integrados que tem como principal objetivo fornecer as tarefas de gerenciamento de suporte, gestão e intercâmbio de informações. Este modelo também permite incorporar facilmente várias ferramentas de apoio e estruturas de dados, bem como atender às necessidades dos utilizadores como profissionais da saúde, estudantes e pacientes.

Especificamente, o SISTB é um sistema de monitoramento para o tratamento da TB que foi criado para ser utilizado durante as visitas dos profissionais de saúde ao domicílio dos pacientes com TB, domiciliados na cidade de Ribeirão Preto, com a finalidade de registrar e tornar acessível os dados do acompanhamento dos pacientes de uma maneira mais rápida e

fácil possibilitando a geração de relatórios de apoio às ações dos profissionais. Diferentemente dos sistemas para a TB citados anteriormente, que são focados na notificação de doenças, resultados de exames e acompanhamento do seguimento do tratamento por meio da compilação de informações, o SISTB, acrescenta o registro diário da tomada da medicação pelo paciente. Este acompanhamento é preconizado pela OMS por meio do DOTs, que é a observação diária da tomada de medicação pelo paciente. O tratamento dura na média de 6 meses e a sua não aderência, é um dos fatores de aumento de óbito por TB. As funcionalidades aqui apresentadas para o SISTB, o colocam como uma plataforma para processamento de informações de pacientes da TB, em que sistemas podem se integrar com a finalidade de acrescentar ou apenas consumir seus dados. Portanto, nesta perspectiva, o SISTB se posiciona como um ecossistema que comporta diferentes sistemas com finalidades distintas, que juntos integram uma plataforma de processamento de informações com o objetivo de auxiliar no combate e controle da TB, denominado ecossistema SISTB. Os sistemas componentes do ecossistema são independentes, podendo funcionar separadamente ou em conjunto (Figura 9).

Figura 9 - Ecossistema proposto para o suporte ao gerenciamento de pacientes de Tuberculose

Fonte – Própria do autor

Abaixo é apresentado de maneira esquemática os outros elementos de toda infraestrutura computacional observada na Figura 9 acima:

- Softwares
 - SISTB Mobile é a extensão do SISTB, é utilizada pelos profissionais de saúde nas visitas domiciliares aos pacientes que permite o registro do tratamento dos como por exemplo, anotação do DOTs. Utilizado por dispositivos *Android*, é possível sua utilização em modo *off-line* e posteriormente, quando conectado à internet, permite que os dados coletados sejam descarregados na base de dados;
 - GuiaTB é um guia digital criado para dispositivos *Android*, que permite aos usuários realizarem pesquisas a protocolos e procedimentos de tratamentos da TB. Possui um conteúdo estruturado e organizado, além de possuir ferramentas adicionais como calculadora de dosagens de drogas, por exemplo. Sua organização segue as recomendações do Guia Rápido para Profissionais da Saúde - Recomendações para o Controle da Tuberculose do MS. Esta ferramenta é baseada em *guidelines* que ajudam a simplificar o acesso aos dados desejados e ajudam no processo de tomada de decisão dos profissionais da saúde por meio de um processo rápido, prático e intuitivo (LIMA et al., 2017);
 - WebDOT é uma plataforma de vídeo que permite o envio de vídeos gravados em *smartphones* pelos pacientes e sua posterior visualização e validação pelos profissionais da saúde;
 - Assistente Virtual para o Tratamento TB tem como objetivo oferecer suporte aos pacientes por meio do monitoramento remoto. Possui recursos de inteligência artificial agregados que habilita o paciente a responder perguntas, buscar conselhos e informações de tratamento por meio de conversação com um *chatbot* utilizando linguagem natural. Dessa maneira é possível fornecer cuidados complementares aos pacientes a qualquer momento.
- Representação e gestão de informações
 - Camada de Suporte a Decisão é uma camada para apoiar os profissionais de saúde na tomada de decisões. O módulo será desenvolvido utilizando o conhecimento existente em banco de dados e heurísticas por meio das informações fornecidas pelos profissionais da saúde. Baseado em casos para apoiar a tomada de decisão em vários aspectos, tem como principal objetivo identificar o abandono precoce do paciente em tratamento.

- Camada de Interoperabilidade é uma camada de melhoria proposta, baseada no paradigma da *Web Semântica*, a ser incorporada ao SISTB para viabilizar troca de informações entre o SISTB e outros sistemas, como o TBWEB e o software de gerenciamento de Atenção Básica do SUS (e-SUS AB). Neste ponto, far-se-á a criação de ontologias da tuberculose de domínio específico e, posteriormente, a marcação e associação de dados recebidos e armazenados com a respectiva ontologia, garantindo dessa maneira, a interoperabilidade semântica dos dados. A extração de conhecimento por esta camada, pode ser melhorada por meio da utilização de ontologias de outras fontes, bem como pela associação de dados marcados por diferentes ontologias;
- Camada de Segurança é a camada responsável pela proteção dos dados sensíveis dos pacientes, permitindo a troca segura de dados entre os aplicativos por meio da internet.
- Repositório
 - Repositório Tuberculose, domínio de conhecimento sobre a TB com um conjunto completo de protocolos e procedimentos (gestão inicial de TB, tipos de tratamento, reinfecções, resistência a antibióticos etc.). Consolida as informações provenientes de diferentes fontes (pré-hospitalar, hospitalar, diagnóstico, tratamento, etc.);
 - Ontologias, vocabulários padronizados por meio de ontologias que contém os códigos de classificação e codificação da TB e uma grande variedade de sinais, sintomas, exames anormais, queixas e circunstâncias sociais;
 - *Web Semântica*, repositório semântico de dados;
 - Terminologias, repositório com códigos e classificações utilizadas por profissionais de saúde envolvidos no tratamento da TB;

Em concreto todo projeto visa explorar a componente de disseminação de informação sobre a TB a partir de um portal de conteúdo que faz parte de todo o ecossistema em saúde definido como SISTB.

4.5 Ferramenta de desenvolvimento do portal: CMS Drupal

A internet tornou-se um dos principais meios de comunicação ao redor do mundo, onde pessoas compartilham diariamente milhões de informações em diferentes formatos, provenientes de diversas fontes e diferentes assuntos (MOREIRA, 2012). Tais informações

são compostas por textos, imagens, sons, vídeos e arquivos diversos chamados de hipertexto, padrão de dados utilizado pela rede mundial de computadores.

No princípio, portais e páginas *web* eram compostos basicamente por textos estáticos sem atualizações constantes. Eram páginas exclusivamente de caráter informacional e desprovidas de design adequado (MOREIRA, 2012).

Aos poucos, novos formatos de conteúdos foram incorporados às páginas *web*, imagens, sons, vídeos entre outros tipos. O design foi aprimorado e o conteúdo apresentado passou a ser dinâmico por meio de atualizações com maior frequência. Em consequência disso, houve o aumento do volume de dados, e juntamente, veio à dificuldade de gerir a quantidade de informações geradas, o que obrigou as instituições a repensarem e adotarem estratégias para organizar a gestão das informações.

Inicialmente, a gestão do conteúdo das páginas, dependia exclusivamente da ajuda de um profissional técnico. Com a finalidade de diminuir essa dependência e auxiliar na gestão do ciclo de vida das informações, permitindo o gerenciamento de conteúdo por pessoas técnicas e não técnicas, foram criadas ferramentas para gestão de conteúdos de sites denominadas Sistemas Gerenciadores de Conteúdo, em Inglês, CMS. Com um conjunto de regras e processos bem definidos, tais ferramentas permitem aos usuários criar, editar e publicar uma variedade de conteúdos utilizando diferentes formatos de dados como texto, imagens, vídeos, sons entre outros (RAWTANI; CHIDAMBARAM, 2009).

O Drupal é um software CMS que auxilia os usuários na publicação, organização e gestão de conteúdos publicados em *web sites* de uma maneira mais rápida e fácil. Ele possui um conjunto de layouts prontos que permite aos usuários criarem e utilizarem diferentes tipos de serviços *web* como fóruns de discussão, sites corporativos, intranets, e-commerce e blogs.

É um *software open source* licenciado de acordo com a *General Public License* (GPL) e mantido por uma comunidade de desenvolvedores ao redor do mundo. É uma ferramenta *free* que pode ser utilizada sem custo por pessoas ou organizações. Baseado na linguagem de programação *Hypertext Preprocessor* (PHP), oferece suporte nativo para o banco de dados MySQL e PostgreSQL, porém, é possível configurá-lo para trabalhar com outros bancos de dados (RAWTANI; CHIDAMBARAM, 2009).

Conforme define Byron et al. (2009), o Drupal é um CMS *open source* utilizado por milhares de organizações e pessoas para construir *web sites* com conteúdos mais ricos. Destaca ainda que a construção de *web sites* no Drupal é feita por meio da combinação de uma variedade de blocos com finalidades específicas de acordo com a necessidade dos usuários.

Após a finalização dos sites, a manutenção é realizada por meio de diferentes formulários sem a necessidade de utilizar codificação manualmente. Para os desenvolvedores, o Drupal oferece recursos de personalização por meio da instalação de módulos *plug-in* que estendem seus recursos e poder de desenvolvimento.

Para Tomlinson (2015), o Drupal é um CMS *free e open source* desenvolvido em PHP e distribuído sob a GPL. Pessoas, times e comunidades aproveitam das facilidades do Drupal para publicar, gerir e organizar conteúdo nos mais variados *web sites* que vão desde *blogs* pessoais, grandes empresas e sites governamentais.

Todd também destaca a capacidade de personalização e extensão das capacidades da ferramenta através da instalação de novos módulos.

Segundo Vandyck (2009), o Drupal é um *framework* modular de código aberto para gerenciamento de conteúdo de *web sites*.

Além das muitas funcionalidades que possui por padrão, é possível adicionar novas funcionalidade através da habilitação de módulos internos e instalação de módulos de terceiros.

4.5.1 CMS

As ferramentas de CMS foram uma alternativa adotada por empresas e pessoas como uma das estratégias para gestão do volume de informações produzidas diariamente. Os avanços tecnológicos que proporcionaram velocidades cada vez mais rápidas de internet contribuíram para um aumento significativo de conteúdo produzido para a *web*. Tal aumento, e a produção despadronizada e desorganizada de conteúdo, tornou-se um problema na década de 1990 quando surgiram os primeiros *softwares* de CMS.

O CMS veio para facilitar o processo de criação e gestão de conteúdo na *web*. Por ser uma ferramenta flexível, Algumas de suas tarefas, dependendo do nível de complexidade, como por exemplo, mudança de layout ou extração de dados do banco para construção de relatórios, podem demandar o trabalho de um profissional técnico especializado. No entanto, quando as tarefas se concentram na utilização básica para criação e administração de conteúdo para o portal, a utilização do CMS dispensa a necessidade de um profissional técnico especializado, passando a ser mantido pelos próprios usuários autorizados.

Além do Drupal, existem outros CMSs para gerenciamento de conteúdo de *web sites*. Outros dois famosos gerenciadores de conteúdo são o Joomla e o Wordpress.

4.5.2 Core do Drupal

Ao ser instalado, o Drupal instala o que comumente é chamado de Drupal *core*. O *core* nada mais é que o núcleo onde se encontra toda a engenharia e lógica de funcionamento da ferramenta (TOMLINSON, 2015).

Abaixo são relacionados os componentes que inicialmente são disponibilizados no Drupal *core*:

- Conteúdo;
- *Uploads / downloads* de arquivos;
- Menus;
- Contas de usuários;
- Perfis e permissões de usuários;
- Taxonomia;
- Fóruns de discussão
- Visualizações para extração e exibição de conteúdo em formulários, listas e tabelas;
- Editor de conteúdo *what you see is what you get* (WYSIWYG);
- Ferramentas de pesquisa;
- Suporte a multi-idíomas;
- *Login* de usuário;
- Relatórios de erros;

4.5.3 Instituições que utilizam o Drupal

Ao redor do mundo, diversas instituições no âmbito público e privado adotaram o Drupal como CMS para criar e manter seus *web sites*. Na tabela 1 abaixo estão relacionadas algumas destas organizações.

Tabela 1 – Relação de instituições que utilizam o Drupal

Instituição	Web Site	País
McDonalds	https://mcdonalds.com.au/	Austrália
UNICAMP	http://www.unicamp.br	Brasil
UFBA	https://www.ufba.br/	Brasil
University College – The University of Toronto	http://www.uc.utoronto.ca	Canadá
University of Northern British Columbia	https://www.unbc.ca/	Canadá
Cancer Care Ontario	https://www.cancercareontario.ca/en	Canadá
WhiteHouse	https://www.whitehouse.gov/	Estados Unidos
Nasa	https://www.nasa.gov/	Estados Unidos
BBC	http://www.bbc.com/	Estados Unidos
Berkeley Institute of Data Science	https://bids.berkeley.edu/	Estados Unidos
The Institute for Quantitative Social Science	https://www.iq.harvard.edu/	Estados Unidos
Stanford School of Engineering	https://engineering.stanford.edu/	Estados Unidos
Mit School of Science	http://science.mit.edu	Estados Unidos
Children's Hospital of Los Angeles	https://www.chla.org/	Estados Unidos
London	https://www.london.gov.uk/	Inglaterra
Greenpeace	https://www.greenpeace.org.uk/	Reino Unido
Universidade de Oxford	http://www.ox.ac.uk/	Reino Unido

Fonte – (MOREIRA, 2012); (DRUPAL, 2018)

4.6 Técnicas de desenvolvimento

O desenvolvimento do projeto assentou-se na metodologia *Agile* Scrumban que combina as melhores qualidades dos métodos Scrum e Kaban utilizados no processo de desenvolvimento ágil de *software*.

Figura 10 – Quadro kanban com a lista de atividades do projeto

Fonte – Própria do autor

Por meio da metodologia ágil é possível melhorar a interação entre os participantes do projeto, possibilitar que os envolvidos no projeto possam acompanhar suas tarefas e as dos demais em cada etapa, possibilitar a construção do projeto de maneira incremental, definição dos *sprints* e seus respectivos tempos de realização, entre outros benefícios (LADAS, 2008). O quadro Kanban (Figura 10), entre outros objetivos, ajuda na organização das tarefas de cada participante, dividindo-se por padrão em 3 listas de tarefas: *To Do* (Para Fazer); *Doing* (Fazendo); *Done* (Feito).

Conforme apresentado na Figura 10 acima, foi utilizado o quadro Kanban da ferramenta Kanbanchi *Task & Project Management* na versão *trial*. Esta ferramenta apresenta uma interface simples dividida em três colunas que organizam as tarefas do projeto. O acesso ao projeto é concedido a cada participante do projeto por meio do cadastro de seu e-mail. Além da organização natural das tarefas, é possível associá-las aos participantes, cada participante pode estar envolvido em mais de uma tarefa, definir níveis de prioridade, indicar a previsão de início e fim das tarefas, indicar dependência entre as tarefas, anexar arquivos entre outros recursos.

É importante reservar este parágrafo e ressaltar que a utilização da metodologia ágil e da ferramenta Kanbanchi *Task* não é mandatória quanto ao método estabelecido para a construção de um portal marcado semanticamente. No entanto, foram utilizadas na condução do projeto como uma recomendação de boas práticas no desenvolvimento

do trabalho estruturado em tarefas. Por existir uma sequência de etapas a serem perseguidas, a utilização desses recursos ajudou na melhor organização das ideias traduzidas em tarefas a serem realizadas almejando minimização de erros e redução do tempo de desenvolvimento deste trabalho.

Antes do início do desenvolvimento, foram realizadas buscas por repositórios públicos de ontologias na *web*. Nem todas as informações utilizadas no portal estão mapeadas com suas respectivas ontologias, no entanto, esta etapa procurou conhecer e elencar quais repositórios poderiam ser utilizados como fonte de dados estruturados semanticamente. Foram definidos e posteriormente utilizados, repositórios com ontologias em um contexto genérico, abordando diferentes assuntos, bem como repositórios direcionados com ontologias em saúde e no âmbito da TB. Os repositórios definidos para a marcação semântica dos dados deste portal foram: Schema.org; DBpedia; BioPortal PURL; disease-ontology.org; obofoundry.org.

O Schema.org é um repositório de ontologias mantido de maneira colaborativa pela comunidade com o objetivo de criar e manter conjuntos de *schemas* diversos que estruturam os dados na internet (Figura 11, 12). Fundado por grandes empresas do segmento de tecnologia como Google, Microsoft, Yahoo e Yandex, seus vocabulários representando diferentes entidades com seus relacionamentos e ações, são utilizados por mais de 10 milhões de sites em suas páginas (SCHEMA, 2018). O portal Schema.org possui uma seção de pesquisa que permite aos usuários localizarem as ontologias e suas respectivas propriedades. Na Figura 12, observa-se um recorte da ontologia Person utilizada no portal da TB para representar dados relacionados às pessoas como *givenName* (Nome), *familyName* (Sobrenome), *gender* (Gênero) e *birthDate* (Data de nascimento) entre outros, por exemplo.

Figura 11 – Seção de pesquisa de ontologias do portal Schema.org

Fonte – Schema.org (2018)

Figura 12 – Conjunto de algumas propriedades relacionadas à entidade *Person*

https://schema.org/Person		
birthPlace	Place	The place where the person was born.
brand	Brand or Organization	The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person.
children	Person	A child of the person.
colleague	Person or URL	A colleague of the person. Supersedes colleagues .
contactPoint	ContactPoint	A contact point for a person or organization. Supersedes contactPoints .
deathDate	Date	Date of death.
deathPlace	Place	The place where the person died.
duns	Text	The Dun & Bradstreet DUNS number for identifying an organization or business person.
email	Text	Email address.
familyName	Text	Family name. In the U.S., the last name of an Person. This can be used along with givenName instead of the name property.
faxNumber	Text	The fax number.
follows	Person	The most generic uni-directional social relation.
funder	Organization or Person	A person or organization that supports (sponsors) something through some kind of financial contribution.
gender	GenderType or Text	Gender of the person. While http://schema.org/Male and http://schema.org/Female may be used, text strings are also acceptable for people who do not identify as a binary gender.
givenName	Text	Given name. In the U.S., the first name of a Person. This can be used along with familyName instead of the name property.

Fonte – Schema.org (2018)

No campo de ontologias relacionadas às doenças humanas, o site BioPortal é o maior repositório de ontologias biomédicas (BIOPORTAL, 2018). Nele estão definidos vocabulários representando dados estruturados de um grande número de doenças que apresentam propriedades como *Preferred Name* (Nome), *Definitions* (Definição), ICD-9 (CID-9) e ICD-10 (CID-10), por exemplo. O portal também apresenta o recurso de busca, onde podem ser pesquisadas as entidades com suas respectivas propriedades (Figura 13). Neste portal da TB foi utilizada a ontologia que contém mapeada as classes da TB, portanto, as páginas que possuem alguma referência à tuberculose, foram mapeadas de acordo com as propriedades definidas em cada classe.

A versão do CMS Drupal utilizada foi a 8.3.7 obtida no próprio portal da ferramenta. O Drupal foi desenvolvido em PHP, portanto, independente de sistema operacional, para sua instalação são necessários um servidor de banco de dados e um servidor *web*. O banco de dados é responsável por armazenar toda a estrutura e configuração do portal, além do seu conteúdo. O servidor *web* é responsável por hospedar e gerenciar os recursos do CMS, além de interpretar as páginas em PHP. O gerenciador de banco de dados utilizado foi o MySQL

em conjunto com o servidor *web* Apache. Com intuito de facilitar e agilizar o processo de configuração do ambiente de desenvolvimento, foi utilizado o software gerenciador Xampp na versão 3.2.2. Este software facilita o processo de instalação e configuração das ferramentas de desenvolvimento em sistemas operacionais Windows, que entre outros recursos, instala o MySQL e o Apache.

Figura 13 - Repositório BioPortal de ontologias biomédicas

Fonte – BioPortal (2018)

O processo de instalação do Drupal é fácil e intuitivo, com o CMS fazendo a maior parte do trabalho de configuração para o usuário. Uma vez que o ambiente está com servidor de banco de dados e servidor *web* executando, por meio da utilização de um navegador inicia-se a instalação que apresenta uma interface gráfica que solicitará os dados de acesso ao banco de dados e o nome do *web site*, e com poucos cliques o site está pronto para ser configurado e alimentando com conteúdo (Figura 14, 15).

A Figura 14 apresenta a tela inicial do Drupal que apresenta a versão de instalação e por padrão o idioma nativo do usuário. No entanto, é possível escolher a instalação em diferentes idiomas.

Seguindo a proposta de fácil instalação por meio de uma interface amigável, poucas informações são solicitadas, como por exemplo, as informações de acesso ao banco de dados conforme apresenta a Figura 15. O banco de dados utilizado neste projeto foi o MySQL,

portanto, foram solicitadas algumas informações como nome do banco de dados, usuário, senha, servidor, porta de comunicação e opcionalmente pode-se informar um prefixo de identificação para as tabelas utilizadas pelo CMS.

Figura 14 – Tela de apresentação da instalação do Drupal

Fonte – Própria do autor

Figura 15 – Tela de configuração do banco de dados do Drupal

Fonte – Própria do autor

O Drupal é um CMS extensível que permite a adição de módulos que oferecem diferentes recursos na criação de *sites*. Para encontrar estes módulos, bem como *templates*, no próprio *site* do Drupal é disponibilizado um repositório de acesso *free* e mantido pela comunidade de desenvolvedores onde podem baixados. Por meio desse repositório foi encontrado o *template* utilizado neste projeto, além dos módulos adicionais necessários.

Uma vez que foi definida estrutura padrão do portal da TB, na sequência foi realizado o trabalho com marcação de dados por meio de Microdata. Todas as tecnologias relacionadas à *Web Semântica* até aqui descritas, RDF, OWL e SPARQL, são tecnologias que viabilizam a definição a organização e estruturação dos dados, de vocabulários e consulta em repositórios semânticos na *web* que podem ser consumidos por diferentes sistemas capazes de processarem tais estruturas. No entanto, para criar páginas com conteúdo marcado e atrelado às estruturas desses repositórios, é necessário marcar semanticamente o dado apresentado, que pode ser feito pela utilização do padrão de representação de informações Microdata. O Microdata é um padrão que estende as potencialidades semânticas no *Hyper Text Markup Language 5* (HTML5), semelhante aos *Microformats Extensible Hypertext Markup Language* (XHTML) e *Resource Description Framework in Attributes* (RFDa), para rotular o conteúdo das páginas utilizando itens descritos por um conjunto de pares de nome-valor (*item-value*) que permite processamento de conteúdo (Figura 16) (WORLD WIDE WEB CONSORTIUM, 2018g). Desta maneira, torna-se possível o entendimento do conteúdo das páginas pelos mecanismos de buscas que conseguem identificar tais estruturas, além de permitir que sistemas localizem e utilizem estas informações nas páginas (Figura 17).

Figura 16 - Representação gráfica do conjunto de nome-valor do Microdata

Fonte – Própria do autor

A Figura 16 acima apresenta um recorte de uma estrutura de grafo que representa graficamente uma estrutura item-valor utilizada pelo código Microdata.

Figura 17 – Dado marcado na página por meio do código Microdata

```
<section itemscope itemtype="https://schema.org/Person">
  <p>Nome: <span itemprop="givenName">Ricardo</span></p>
  <p>Sobrenome: <span itemprop="lastName">Roberto de Lima</span>
</section>
```

Fonte – Própria do autor

Na Figura 17 acima representando a marcação por meio do Microdata, observa-se que o item referenciado é o *Person*, mapeado em *itemscope*, a *property* são nome e sobrenome, mapeados por *givenName* e *lastName*, e *value*, é o valor expresso pelas literais Ricardo e Roberto de Lima. O trecho de código na Figura 17, é baseado na classe *Person* do repositório *schema.org* para a marcação semântica do dado.

A marcação semântica dos dados no Drupal utilizando Microdata é realizada por meio de dois caminhos. Um desses caminhos é por meio da codificação direta na página, onde é possível, uma vez que conhecida a ontologia e suas classes, definir diretamente no código de criação da página, as *tags* de marcação aos dados desejados. No entanto, este caminho não é uma tarefa simples que poderia ser realizada pelo usuário sem conhecimento técnico. Este é um trabalho para profissionais técnicos com conhecimento em desenvolvimento de páginas *web*, uma vez que, são necessárias a criação de rotinas de código mediante utilização da linguagem de programação PHP. Para utilização de códigos PHP no Drupal, é necessário estender suas funcionalidades, que foi realizada neste projeto instalando o módulo PHP *Filter module*. Neste processo de marcação de dados, além da PHP, também faz-se necessária a utilização de bibliotecas específicas para manipulação de estruturas RDF e consultas SPARQL. A biblioteca habilitada junto ao *core* do PHP foi a ARC, criada para ser utilizada com esta linguagem de programação, é uma biblioteca free e open-source, compatível a partir da versão 5.3 da PHP, específica para a manipulação de estruturas RDF (ARC2, 2018).

A construção de páginas que traz com algumas informações pertinentes à TB, foi criada manualmente, por meio de codificação HTML5, CSS, PHP e ARC2 para busca dinâmica das informações. A recuperação dos dados foi realizada mediante programação PHP utilizando a biblioteca ARC2 para execução de consultas SPARQL em repositórios de ontologias. Os dados recuperados foram armazenados em um *array* de dados, estrutura lógica de programação e, posteriormente, o *array* foi utilizado na construção dinâmica da página fornecendo os dados a serem marcados por meio do padrão Microdata pelas tags *section*, *itemscope*, *itemtype* e *itemprop*.

Outro caminho possível de marcação de conteúdo é por meio da utilização de formulários que realizam automaticamente a marcação semântica sem a intervenção do usuário no código. *Softwares* de CMS se apoiam na ideia de facilitar a criação e gestão de conteúdo, seguindo este princípio, o Drupal também oferece recursos mediante utilização de formulários com interfaces amigáveis que proveem a marcação por meio do Microdata. É necessário habilitar a utilização desse recurso no Drupal utilizando o recurso de instalação de novos módulos. O próprio repositório do Drupal oferece uma extensa gama de módulos para as mais variadas finalidades, incluindo módulos capazes de trabalhar com Microdata. Baseado na facilidade instalação e utilização, foi efetuado o *download* do módulo Schema.org *configuration tool* (RDF UI) 8.x-1.0-beta1. A ativação deste módulo, habilita os recursos de marcação semântica por meio da utilização de formulários que estão diretamente integrados ao repositório Schema.org. Dessa maneira, o usuário, primeiramente, escolhe qual entidade deseja utilizar para que assim possa ter acesso às suas propriedades para posterior marcação dos dados da página.

4.7 Configuração de um servidor virtual de ontologias

Este estudo é direcionado à utilização de ontologias para marcação de conteúdo em portais, que neste contexto é o portal de conhecimento sobre a TB. A construção de ontologias, portanto, não é o principal objetivo, visto que, o processo de criação de ontologias demanda certa complexidade, uma vez que é baseado em estudos prévios e são necessários profissionais com conhecimento específicos para mapear o contexto desejado. No entanto, para um melhor aproveitamento acadêmico e enriquecimento deste estudo, por meio da plataforma D2RQ, na versão 0.8.1, foram criadas ontologias baseadas no conteúdo do banco de dados do SISTB.

A D2RQ é um sistema *open source* publicado sob a *Apache License* que transforma bases de dados relacionais em uma estrutura virtual de grafos RDF somente leitura que oferece acesso por RDF ao conteúdo relacional do banco de dados. A plataforma permite a execução de pesquisas SPARQL, acesso aos dados por meio da internet como dados conectados (*Linked Data*), *dumps* de bases RDF e utiliza a api Apache Jena API para permitir a manipulação de dados por outras aplicativos. Suporta conexões JDBC para diversos bancos de dados e também permite conexões ODBC.

A plataforma D2RQ é composta por três sistemas: D2RQ *Mapping Language*, responsável por descrever a relação entre as ontologias e o banco de dados relacional; D2RQ

Engine, *plug-in* para ferramenta de Jena *Web Semântica*; D2R Server, servidor HTTP. A instalação da ferramenta é simples e sua configuração pode ser seguida por meio de manuais no próprio *site* da ferramenta.

Para implementar a plataforma, foi utilizado servidor na nuvem da Google cloud com licença gratuita por um ano. Nesta estrutura, foi configurado um banco de dados MySQL para criação do repositório do SISTB e criado um servidor *Windows* 2016 para receber a instalação da plataforma e servir como servidor *Hypertext Transfer Protocol* (HTTP) do repositório virtual de grafos RDF. Foram utilizados também como recursos necessário à instalação da D2RQ, o Java JDK 8 e o servidor *web* Tomcat 8.5.33.

Uma versão de testes do portal foi migrada para um servidor baseado em nuvem da *Cloud* USP rodando o sistema operacional Ubuntu na versão 16.04.4 LTS. O servidor de banco de dados em produção é o MySQL 5.7.22 que recebeu toda estrutura de tabelas do CMS, e o servidor *web* é o Apache 2.4.18 que recebeu todos os arquivos do portal. Como ferramentas de apoio ainda foram utilizadas o phpMyAdmin 4.7.6 para auxiliar na gestão da base de dados, para a transferência de arquivos via File Transfer Protocol (FTP) utilizou-se o FileZilla 3.7.3 e o cliente SSH Putty 0.70 para conectar-se ao console do servidor.

5 RESULTADOS E DISCUSSÃO

A proposta inicial deste trabalho foi a criação de um portal de conhecimento para o campo da TB com conteúdo marcado semanticamente por meio de ontologias. Conforme apresentado na Figura 7, quanto ao esquema metodológico estabelecido para execução deste estudo, a figura pode ser interpretada sob duas perspectivas, uma de contextual e outra detalhada. Sob a perspectiva contextual, este trabalho apresentou uma avaliação teórica dos conceitos e tecnologias da que viabilizam o uso da *web* semântica, bem como o desenvolvimento de um portal de conhecimento sobre a TB marcado semanticamente. Na ótica detalhada, as etapas maiores representadas pelo esquema metodológico, foram divididas em tarefas que permitiram o aprofundamento necessário ao estudo.

Para atingir o objetivo proposto por este trabalho, foram realizadas e documentadas um conjunto de etapas, contemplando a fundamentação teórica com o levantamento bibliográfico e abordagem das tecnologias da *web* semântica, a pesquisa de repositórios de dados sobre a TB para geração de dados ao portal, a seleção de indicadores de saúde no contexto da TB baseando-se no Manual de Recomendação para o Controle da Tuberculose, estudo de repositórios de ontologias com suas classes e propriedades e a configuração do ambiente de trabalho para o desenvolvimento do portal contemplado não somente a geração de conteúdo, mas também a marcação semântica das páginas desenvolvidas utilizando o CMS Drupal. A definição do esquema metodológico para o desenvolvimento deste estudo, configura-se como um caminho possível a ser percorrido no desenvolvimento de um portal semântico, podendo ser replicável e melhorado em outros trabalhos semelhantes.

As etapas teóricas e práticas foram divididas em pequenas tarefas necessárias no decorrer do estudo, e foram organizadas por meio da abordagem ágil de projetos documentadas com a utilização da ferramenta *Kanbanchi Task & Project Management* (Figura 18).

Figura 18 – Quadro Kanban representando organização das tarefas do estudo

Fonte – Própria do autor

Disponibilizada *online*, a versão *trial* fornece as funções essenciais para gerenciamento de projetos, exigindo dos participantes uma conta de e-mail para cadastro como colaborador da equipe. No quadro apresentado pela Figura 18, a primeira coluna apresenta o conjunto de tarefas a serem iniciadas, na segunda coluna são apresentadas todas as tarefas em desenvolvimento e a terceira coluna apresenta todas as tarefas já concluídas.

A extração de conhecimento por meio de uma *web* programável que permita a criação de portais *web* providos de conhecimento e que possam enriquecer a experiência de navegação dos usuários e possibilitar que tanto humanos quanto computadores tenham mesmo entendimento sobre as coisas, é o pilar sob o qual a *Web Semântica* foi concebida. Seguindo este princípio, foi desenvolvido um portal *web* com conhecimento sobre a TB com parte de seu conteúdo marcado semanticamente com dados estruturados mapeados por ontologias. Sendo assim, o portal adquire a capacidade de entregar conhecimento alinhando o entendimento tanto na visão humana quanto na visão de máquina com softwares capazes de identificar e interpretar as estruturas de dados marcados por meio da utilização do padrão Microdata.

Foi utilizado CMS Drupal com a base de dados MySQL para a criação e gerenciamento da estrutura do portal, bem como do conteúdo gerado pelos usuários. A instalação padrão do CMS entrega um layout com poucos recursos e visualmente pobre em estilo. É importante nesta etapa, antes de iniciar o processo de criação do portal, encontrar

um *template* que já implemente alguns recursos, como banner rotativo, menus *dropdown*, formulários de contato, controle de usuário entre outros, pois tornará o processo de criação mais produtivo, além de produzir um site organizado que ofereça uma melhor experiência de uso aos usuários. Foram testados alguns templates, todos *free*. Os itens avaliados para a escolha do template foram: oferta de recursos como comentado anteriormente, desempenho quando sua escalabilidade e velocidade, cores e organização do layout. Foi utilizado neste projeto o template *free qatar zymphonies theme* que oferece um estilo *Cascading Style Sheets* (CSS) agradável e organizado (Figura 19).

Figura 19 – Layout da tela inicial modificado pelo *template qatar zymphonies theme*

Fonte – Própria do autor

A Figura 19 acima apresenta a tela principal do Portal TB. Hospedado em um servidor *Cloud* da USP sob os cuidados do LIS, este servidor foi utilizado como ambiente de testes no processo de desenvolvimento da pesquisa e desenvolvimento do portal, que está em sua primeira versão. Este portal foi totalmente desenvolvido em Drupal, e priorizando as recomendações de usabilidade que recomenda que qualquer informação deve estar a dois cliques do usuário, foi aplicado um *template* com um visual limpo, com poucas informações de entrada, mas, que disponibiliza informações organizadas por meio de menus (KRUG, 2006). Além da barra de menus, a tela principal ainda conta com um recurso de pesquisa de conteúdo na base de dados do portal, a opção de escolha de idiomas de visualização em Português e Inglês, um banner rotativo para notícias em destaque e um *login* de usuários, uma

vez que, para qualquer usuário cadastrar e efetuar manutenções no conteúdo do portal, ele precisa ter efetuado o login.

Os menus foram inicialmente organizados da seguinte maneira:

- SOBRE - Informações referente ao portal e ao seu conteúdo. É uma carta de apresentação ao usuário;
- TUBERCULOSE - Menu *dropdown* que disponibiliza informações a respeito da TB. O que é, principais ações e programas de combate a doença;
- PARCEIROS – Relaciona os parceiros produtores de conteúdo, ações e pesquisas no campo da TB;
- PRODUÇÃO - Conteúdo produzido pela equipe do LIS e outros usuários terceiros autorizados a gerar conteúdo para o site. Entre as produções, encontram-se softwares, trabalhos, pesquisas e artigos produzidos;
- CONTATO - Formulário para contato entre usuários e os responsáveis pelo portal.

Este padrão de visualização do portal é comum a todos os usuários ao acessarem o portal. No entanto, para usuários cadastrados, após efetuarem login, são disponibilizados os menus de manutenção (Figura 20).

É possível observar na Figura 20 uma barra adicional ativa destacada na cor preta com um conjunto de opções para gerenciamento do conteúdo do portal. As opções disponibilizadas podem variar de acordo com o nível de acesso concedido aos usuários. Após efetuar login é possível entre outras coisas, criar novas notícias, realizar manutenção nas notícias existentes, criar novos usuários, criar páginas com ou sem marcação semântica, instalar novos recursos no Drupal entre outras funcionalidades.

Figura 20 – Tela principal do Portal TB com menu de manutenção do portal ativo

Fonte – Própria do autor

Aos olhos do usuário, não existem diferenças aparentes entre um site comum, sem marcação semântica, de um site com marcação semântica (Figura 21).

Figura 21 – Portal TB na visão do usuário

Fonte – Própria do autor

Visualmente são iguais. O que realmente faz a diferença é o que o usuário não consegue ver, a estrutura de dados marcados que referenciam os repositórios de ontologias e seus vocabulários. Imperceptível aos olhos, no entanto, o usuário sem perceber é beneficiado pela organização da informação estruturada e padronizada por meio de vocabulários que mapeiam o conhecimento com base no senso comum que, independentemente, do site que as referenciam, sempre apresentará o mesmo conteúdo, evitando dessa maneira, a divergência de informações e otimizando a busca de conteúdo sobre o assunto pesquisado.

Na Figura 21, a página acessada por meio do menu TUBERCULOSE ->SOBRE, que resume informações sobre a TB, representa a página na visão dos usuários. Apesar de estar marcada semanticamente, referenciando os dados do repositório de ontologias BioPortal, e as informações estarem baseada em padrões definidos internacionalmente, como por exemplo, CID-9 e CID-10, na visão do usuário é apenas uma página como qualquer outra sem marcação. No entanto, ela traz um pequeno resumo organizado e padronizado por ontologias que facilitam o acesso a informação confiável pelo usuário, este tipo de organização, reduz a necessidade dos usuários de realizarem diferentes buscas em diferentes sites, otimizando seu processo por busca de informações.

Para os computadores, a utilização dessas estruturas semânticas pelos *web sites* (Figura 22), carregam conjuntos de informações por meio dos metadados que permitem o processamento inteligente dos dados pelos motores de busca como Google, Bing, Yahoo e outros, bem como por outros softwares capazes de interpretar estruturas semânticas, que neste caso está representada por meio de Microdata.

Figura 22 – Portal TB na visão do computador

The screenshot shows a web browser window with the address bar displaying `http://200.144.254.4:8080/portaltb/pt-br/tuberculose`. The page is titled "Tuberculose | Portal TB". The left pane shows the raw HTML code, including XML namespaces and a detailed definition of Tuberculosis. The right pane shows a structured view of the data, with a table listing properties like @type, label, definition, and database_cross_reference.

Property	Value
@type	http://purl.obolibrary.org/obo/DOID_2957
label	Tuberculose Pulmonar
definition	E a tuberculose localizada nos pulmões. A infecção tem febre alta, sintomas de tosse, sintoma de dificuldade respiratória, infiltrações inflamatórias, formação de tubérculos, caseação, derrame pleural e fibrose sincrônica.
database_cross_reference	CID9 - 011.90
database_cross_reference	CID10 - A15.0

Fonte – Própria do autor

No entanto, na visão do computador, apresentada na imagem acima, é possível detectar e interpretar as estruturas semânticas sinalizadas por Microdata. Por meio da “Ferramenta de teste de dados estruturados” disponibilizada pela Google, é possível validar a estrutura de uma página que possui dados estruturados. Conforme é possível observar no resultado do teste, ao lado esquerdo é apresentada toda codificação da página HTML, CSS e de scripts da página, ao lado direito estão destacadas as informações sobre a classe do repositório das ontologias e suas respectivas propriedades, que no caso são o nome da doença (*label*), definição da doença (*definition*) e referência de CIDs (*database_cross_reference*). Esta é a visão sob a perspectiva da máquina, que alinhado ao conceito estabelecido pela *web* semântica, entende o significado do conteúdo da mesma maneira como é entendido pelos humanos, por exemplo, a definição da doença (propriedade *definition*) que apresenta a definição a respeito da TB.

A construção da página que traz com algumas informações pertinentes à TB, foi criada manualmente, por meio de codificação HTML5, CSS e PHP para busca dinâmica das informações. A recuperação dos dados foi realizada mediante programação PHP, que utilizando a biblioteca ARC2 permite a execução de consultas SPARQL em repositórios de ontologias. Neste exemplo foram utilizadas três ontologias do repositório BioPortal:

URI ontologia da TB - `http://purl.obolibrary.org/obo/DOID_399` ;

URI ontologia da TB pulmonar - `http://purl.obolibrary.org/obo/DOID_2957` ;

URI ontologia da TB multirresistente - http://purl.obolibrary.org/obo/DOID_401 ;

Os dados recuperados dos repositórios foram armazenados em um *array* de dados, estrutura lógica de programação e, posteriormente, seu conteúdo foi utilizado na construção dinâmica da página que foi marcada por meio do padrão Microdata que fornece as *tags* de marcação *section*, *itemscope*, *itemtype* e *itemprop* (Figura 23).

Figura 23 – Recorte do trecho do código da página do Portal TB com Microdata

```
#Tuberculose
SELECT DISTINCT ?property ?hasValue
WHERE {
  {<http://purl.obolibrary.org/obo/DOID_399> ?property ?hasValue .}
  filter (?property = <http://purl.obolibrary.org/obo/def>)
  filter (?property = <http://www.w3.org/2000/01/rdf-schema#label>)
}
```

Consulta SPARQL recuperando as informações
sobre a tuberculose no repositório BioPortal

property	hasValue
<http://purl.obolibrary.org/obo/def>	"A primary bacterial infectious disease that is located_in lungs, located_in lymph nodes, located_in pericardium, located_in brain, located_in pleura or located_in gastrointestinal tract, has_material_basis_in Mycobacterium tuberculosis, which is transmitted_by droplets released into the air when an infected person coughs or sneezes."
<http://www.w3.org/2000/01/rdf-schema#label>	"tuberculosis"
<http://www.w3.org/2000/01/rdf-schema#label>	"Tuberculosis"

```
1 <?php
2 /* ARC2 static class inclusion */
3 include_once('semsol/ARC2.php');
4
5 $dbpconfig = array(
6   "remote_store_endpoint" => "http://sparql.bioontology.org/";
7   $store = ARC2::getRemoteStore($dbpconfig);
8
9   if ($err = $store->getErrors()) {
10    echo "<?php getRemoteStore error<?php>";
11   }
12
13 $query = '
14 PREFIX owl: <http://www.w3.org/2002/07/owl#>
15 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
16 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
17 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
18 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
19 PREFIX dc: <http://purl.org/dc/elements/1.1/>
20 PREFIX : <http://dbpedia.org/resource/>
21 PREFIX dbpedia2: <http://dbpedia.org/property/>
22 PREFIX dbpedia: <http://dbpedia.org/>
23 PREFIX dbpprop: <http://dbpedia.org/property/>
24
```

```
25 SELECT DISTINCT ?property ?hasValue
26 WHERE {
27   {<http://purl.obolibrary.org/obo/DOID_399> ?property ?hasValue .}
28   filter (?property = <http://purl.obolibrary.org/obo/def>)
29   filter (?property = <http://www.w3.org/2000/01/rdf-schema#label>) ;
30 }
31
32 /* execute the query */
33 $rows = $store->query($query, 'rows');
34 if (empty($rows)){
35   echo 'Não encontrado';
36 }
37 ?>
```

Utilização da biblioteca ARC2 com PHP para acessar o
motor SPARQL do repositório BioPortal

Conteúdo traduzido e marcado utilizando Microdata

```
<section itemscope="" itemtype="http://purl.obolibrary.org/obo/DOID_399">
<p itemprop="label"><strong>Tuberculose (Mycobacterium tuberculosis)</strong></p>

<p itemprop="definition">É uma doença infecciosa bacteriana localizada nos pulmões, nos gânglios linfáticos, pericárdio, encéfalo, pleura ou localizada no trato gastrointestinal. Seu agente infeccioso é a bactéria Mycobacterium tuberculosis, que é transmitida por gotículas liberadas no ar quando uma pessoa infectada tosse ou espirra.</p>
</section>
```

Fonte – Própria do autor

Na Figura 23 é apresentada uma sequência de passos representando o processo de extração e marcação de conteúdo por meio das tecnologias da *web* semântica. No primeiro trecho de código, uma consulta SPARQL executada diretamente na *interface* de consulta do repositório, extrai o conhecimento sobre TB documentado no BioPortal. Para montar o código SPARQL, foi realizado o acesso ao repositório, busca da ontologia e análise das informações

e possíveis filtros para realização da consulta. No segundo trecho de código, dentro do código PHP da página do Portal TB, foi importada a biblioteca ARC2 no código e a mesma consulta SPARQL do primeiro trecho, foi associada à biblioteca que é encarregada de processar as consultas aos repositórios de ontologias. No terceiro trecho é apresentado o código HTML marcando o conteúdo da página semanticamente já traduzido para o idioma Português.

O BioPortal oferece um motor de busca baseado em SPARQL (<http://sparql.bioontology.org/>) que, por meio da URI de cada ontologia é possível construir sentenças SPARQL para recuperação das propriedades com seus respectivos valores das classes das ontologias. Utilizando as URIs apresentadas acima, foram criadas consultas para recuperar o conteúdo das ontologias do repositório. As propriedades recuperadas das classes ontológicas utilizadas foram a *label* (define um literal que identifica a doença) e *definition* (definição da doença) que são propriedades comuns a todas e, para ontologia da tuberculose pulmonar também foi recuperado o CID-9 e CID-10 (classificação internacional da doença).

É importante ressaltar que, o BioPortal apresenta seu conteúdo no idioma Inglês, no entanto para este estudo, para uma melhor apresentação na página, o conteúdo recuperado do repositório foi traduzido para o idioma Português. Isso fortalece a necessidade da criação de repositórios de ontologias que atendam o cenário brasileiro na utilização da *web* semântica.

Para utilizar as facilidades de gestão de conteúdo propostas pelos CMSs, é possível utilizar a marcação de dados pela utilização de módulos que se integram com repositórios específicos de ontologias e oferecem interfaces gráficas de fácil utilização. Essas interfaces permitem a criação de páginas marcadas sem a necessidade de utilização direta de linguagens de programação, e é por meio destas interfaces que usuários sem conhecimento técnico podem criar páginas facilmente marcadas semanticamente. Neste estudo foi utilizado o módulo RDF UI que se conecta diretamente ao repositório Schema.org e gera marcação semântica automática dos dados, bastando apenas, apontar a ontologia desejada e suas propriedades. Portanto, foi criada uma página no Portal TB para o cadastro de artigos relacionados à TB em que é possível informar o nome do autor, a instituição que o autor pertence, seu vínculo com a instituição e o evento que o artigo foi apresentado. Todos esses campos foram marcados com propriedades vinculadas à entidade *Person* existente no repositório schema.org (Figura 24).

Figura 24 - Página de cadastro de artigos com marcação semântica

Fonte – Própria do autor

Nota-se na imagem acima, destacado a esquerda, que é apresentada ao usuário uma tela de cadastro comum de entrada de dados, no entanto, ao cadastrar o conteúdo, a tela automaticamente marca o conteúdo com Microdata sem a necessidade da intervenção do usuário. Ao lado direito, apenas para ilustração, é apresentada a seção Schema.org *Mappings*, habilitada após a instalação do módulo RDF UI, que liga os campos do formulário à ontologia *Person*. Ao usuário é apresentado apenas os campos de cadastro à esquerda, uma vez que, a ligação do formulário ao repositório Schema.org conforme apresentada, foi configurada em um momento anterior na criação da tela de cadastro, portanto, é transparente ao usuário.

Foi criada uma seção de indicadores da TB conforme recomendado pelo Manual de Recomendações para o Controle da Tuberculose no Brasil (MINISTÉRIO DA SAÚDE, 2011). Os indicadores são utilizados como recursos para apoio à tomada de decisão que creditam as estratégias traçadas pelas instituições de saúde, uma vez que, quando analisados devem apresentar o estado da saúde no contexto temporal analisado, expondo situações sanitárias da população e fornecendo dados confiáveis para a atuação da vigilância da saúde.

O acesso aos indicadores foi disponibilizado no Portal TB por meio do menu (PRODUÇÃO - INDICADORES). Um conjunto de nove indicadores relacionados ao tratamento e prevenção da TB foram disponibilizados. A página de indicadores foi adequada com base em um trabalho anteriormente desenvolvido na Faculdade de Filosofia, Ciências e

Letras de Ribeirão Preto e Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (USP) (RONDELLI, 2017). Inicialmente, a base de dados principal utilizada nos testes realizados foi a do SISTB, uma vez que, esta base, bem como o Portal TB são componentes do ecossistema da TB proposto e, portanto, facilita o acesso aos dados necessário à geração dos indicadores.

A página de indicadores apresenta um layout simples, apresentando filtros com dois campos para informar as datas de início e fim do período desejado, opção da escolha da base de dados que o usuário deseja buscar os dados, e a seleção do indicador desejado.

O resultado da página é composto por uma seção de descrição detalhada do indicador, variáveis de cálculo e o método de cálculo. Também é apresentado em resultados um gráfico que representa os valores do indicador selecionado.

O indicador Proporção de casos de tuberculose testados para HIV representa o quantitativo dos casos de tuberculose que foram testados para HIV (Figura 25).

Em função de a tuberculose ser a primeira causa de morte em pacientes portadores da aids, a identificação precoce dos casos HIV positivo torna-se importante para que um resultado satisfatório no desfecho de tratamento possa ser alcançado. Seu método de cálculo é baseado número total de pacientes notificados com teste HIV realizado dividido pelo número total de pacientes notificados.

Figura 25 – Proporção de casos de tuberculose testados para HIV

Fonte – Própria do autor

O indicador Proporção de coinfecção TB/HIV representa a magnitude da coinfecção TB/HIV (Figura 26). Seu método de cálculo é baseado número total de casos novos notificados com teste HIV positivo dividido pelo número total de casos novos notificados.

Figura 26 – Proporção de coinfecção TB/HIV

Fonte – Própria do autor

O indicador abaixo representa a Proporção de casos de tuberculose com HIV em andamento (Figura 27). É uma recomendação nacional que todos os casos novos de TB façam o exame anti-HIV. Seu método de cálculo é baseado número casos novos de TB com resultado do HIV em andamento dividido pelo número de casos novos de TB notificados com resultado HIV realizado.

Figura 27 – Proporção de casos de tuberculose com HIV em andamento

Fonte – Própria do autor

O indicador Proporção de casos de tuberculose curados representa a efetividade do tratamento (Figura 28). O alcance das metas pactuadas para este indicador tem como objetivo a redução da transmissão para novos pacientes, diminuindo a ocorrência e casos novos. Seu método de cálculo é baseado número de casos encerrados por cura dividido pelo número de casos de TB notificados.

Figura 28 - Proporção de casos de tuberculose curados

Fonte – Própria do autor

Somam-se ainda aos indicadores acima representados pelas figuras outros cinco indicadores:

- Proporção de casos de tuberculose que abandonaram o tratamento
 - Representa o quantitativo de pacientes que, conforme os parâmetros avaliados, abandonaram o tratamento. Sendo aceitável uma proporção de 5% de abandono;
 - Seu método de cálculo é baseado número de casos encerrados por abandono de tratamento dividido número de casos de tuberculose notificados.
- Proporção dos casos de tuberculose com encerramento em óbito
 - Quantitativo de óbitos em decorrência da TB e outras causas
 - Seu método de cálculo é baseado no número de casos de TB encerrados por óbito dividido pelo número de casos de TB notificados;
- Proporção de casos de tuberculose que realizaram o tratamento diretamente observado
 - Quantitativo de pacientes são acompanhados por meio do DOTS. No numerador devem ser incluídos somente os casos que atingiram 24 tomadas diretamente observadas nos dois primeiros meses e 36 nos quatro últimos meses do esquema básico;

- Seu método de cálculo é baseado no número de casos de TB que realizaram o tratamento diretamente observado dividido pelo número de casos novos de TB notificados;
- Proporção de casos de retratamento que realizaram o exame de cultura
 - Quantitativo de casos de retratamento que já realizaram o exame de cultura de escarro que possibilita identificar a presença do *Mycobacterium tuberculosis*;
 - Seu método de cálculo é baseado no número de casos de retratamento de TB que realizaram o exame de cultura dividido pelo número de casos novos de retratamento de TB notificados;
- Proporção de casos de retratamento de tuberculose
 - Quantitativo de casos que precisaram de retratamento, o que pode expressar a má qualidade do tratamento;
 - Seu método de cálculo é baseado no número de casos de retratamento de TB dividido pelo número de casos novos de TB notificados;

As telas de indicadores apresentadas nas Figuras 25, 26, 27, 28 dispõem da mesma estrutura sendo modificadas apenas as informações associadas ao indicador selecionado pelo usuário.

Figura 29 – Trecho de código destacado com a marcação Microdata do indicador “Proporção de casos de tuberculose curados”

The screenshot displays a web application interface for TB indicators. On the left, a sidebar contains a code editor with HTML and Microdata code. The code includes a form for 'Resultado' and a section for 'Detalhes do Indicador' with Microdata items for 'tuberculose' and 'tuberculose curados'. The main content area on the right shows a 'Detectado' section with a table of indicators. The table has columns for 'ERROS', 'AVISOS', and 'ITEMS'. The first row shows '0 ERROS', '0 AVISOS', and '3 ITENS'. The second row shows '0 ERROS', '0 AVISOS', and '1 ITEM'. The third row shows '0 ERROS', '0 AVISOS', and '1 ITEM'. A play button icon is visible in the bottom right corner of the sidebar.

Fonte – Própria do autor

O que diferencia esta página de resultados de outras páginas *web*, é a estruturação das informações que estão marcadas pelo recurso de Microdata, além de extrair as informações pertinentes aos indicadores de um repositório de ontologias baseado no banco de dados do SISTB.

Os dados são extraídos das ontologias utilizando o recurso da tecnologia de pesquisa de dados em repositórios RDF, SPARQL. Por meio das consultas SPARQL, é possível recuperar as informações das propriedades associadas aos indicadores. Para este estudo, os dados padronizados para os indicadores utilizados nas marcações semânticas das páginas, referem-se ao nome, definição, nome e descrição das variáveis de cálculo (Figura 29).

Na Figura 29, por meio da utilização da “Ferramenta de teste de dados estruturados” da Google, é possível identificar as propriedades marcadas referentes ao indicador “Proporção de casos de tuberculose curados”. Ao lado esquerdo é apresentado em destaque um trecho do código da página analisada pela ferramenta indicando no código a utilização das estruturas de Microdata. As informações marcadas foram recuperadas por meio de consultas SPARQL de um servidor de ontologias de indicadores criadas para este estudo.

Ao lado direito da imagem, são apresentadas as informações retiradas dos dados marcados na página. Com a utilização de um repositório de ontologias, como apresentado neste estudo, permite que, independente da fonte de origem dos dados (SISTB, TBWEB e outros), pode-se recuperar as mesmas informações referentes aos indicadores por meio do compartilhamento da ontologia. Isso significa que, por exemplo, o conhecimento atribuído ao método de cálculo de um indicador, seja apresentado de forma padronizada para diferentes aplicações e usuários.

A figura seguinte (Figura 30) apresenta novamente a tela analisado pela “Ferramenta de teste de dados estruturados”, no entanto, são apresentadas as informações expandidas de cada estrutura de dados encontradas que são referentes à descrição das variáveis utilizadas no cálculo e a descrição do indicador “Proporção de casos de tuberculose curados”.

Figura 30 - Apresentação das informações contidas nas estruturas de dados encontradas na página referentes ao indicador “Proporção de casos de tuberculose curados”

The screenshot shows a web application interface. On the left, there is a form with a button labeled "Resultado" and a table with three rows of data. The table has columns for @type, tbl_variaveis_tipo_variavel, and tbl_variaveis_descricao_variavel. The data rows are:

@type	tbl_variaveis_tipo_variavel	tbl_variaveis_descricao_variavel
http://35.198.44.239:8080/d2rq/resource/tbl_variaveis/7	Numerador	Número de casos de tuberculose encerrados por cura por data de diagnóstico.
http://35.198.44.239:8080/d2rq/resource/tbl_variaveis/2	Denominador	Número de casos de tuberculose notificados em determinado período
http://35.198.44.239:8080/d2rq/resource/tbl_indicadores/4		

The right side of the screenshot shows a table with three rows of data, each with a header row and a description row. The headers are:

- tbl_indicadores_descricao_indicador
- tbl_variaveis_descricao_variavel
- tbl_variaveis_tipo_variavel

The descriptions are:

- Expressa a efetividade do tratamento. O alcance das metas pactuadas para este indicador visa à redução da transmissão para novos pacientes, diminuindo a ocorrência de casos novos.
- Número de casos de tuberculose encerrados por cura por data de diagnóstico.
- Número de casos de tuberculose notificados em determinado período

Fonte – Própria do autor

Figura 31 – Tela do servidor D2R de ontologias

The screenshot shows the My D2R Server web application. The page has a navigation bar with links: Home, atualizacao, cache, config, contato, controle, mensais, importacao, internacao, medicamento, paciente, profissional, supervisionamento, **tbl_indicadores**, **tbl_variaveis**, tratamento, and unidade. The main content area has three sections:

- 1. HTML View**: You can use the navigation links at the top of this page to explore the database.
- 2. RDF View**: You can also explore this database with Semantic Web browsers like [Disco](#) or [Marbles](#). To start browsing, open this entry point URL in your Semantic Web browser: <http://35.198.44.239:8080/d2rq/all>
- 3. SPARQL Endpoint**: SPARQL clients can query the database at this SPARQL endpoint: <http://35.198.44.239:8080/d2rq/sparql>. The database can also be explored using [this AJAX-based SPARQL Explorer](#).

Fonte – Própria do autor

O servidor de ontologias baseado no banco de dados do SISTB foi implementado, seu acesso é possível mediante utilização de navegadores *web* por meio da *Uniform Resource Locator* (URL) do servidor D2R. De uma maneira geral, os repositórios públicos de ontologias sobre a TB, contemplam dados de uma maneira mais ampla para atingir um número maior de usuários e aplicações. O contexto dos indicadores da TB é específico,

portanto, esta etapa tornou-se necessária para viabilizar a marcação semântica dos dados dos indicadores disponibilizados no portal (Figura 31).

A Figura 31 apresenta a tela inicial do servidor D2R. A partir desta tela, o usuário pode navegar na estrutura RDF do SISTB acessando os links disponíveis no topo da página. Cada link corresponde a uma tabela do banco de dados relacional, o acesso ao link conduz o usuário aos dados estruturados mapeados por meio de sua respectiva ontologia. Cada dado é identificado univocamente pela sua URIs, conforme define a *web* semântica. Nos links de acesso às tabelas, foi destacado na imagem dois links referentes a duas tabelas adicionais criadas no SISTB para o desenvolvimento deste estudo. Os links correspondem às tabelas “tbl_variaveis” e “tbl_indicadores”, destinadas a armazenar dados pertinentes aos indicadores como nome, descrição e variáveis de cálculo.

Uma vez que estes indicadores estão estruturados, diferentes aplicações que implementam os recursos da *web* semântica podem compartilhar o mesmo indicador acessando-o por meio de sua URI. Também é possível marcar dados em páginas *web*, bem com extrair o conhecimento armazenado e padronizado pelas ontologias que representam os indicadores (Figura 32).

Figura 32 – Vocabulário do indicador Proporção de casos de tuberculose curados

Property	Value
rdfs:label	tbl_indicadores #4
vocab:tbl_indicadores_cod_indicador	4 (xsd:integer)
vocab:tbl_indicadores_cod_variavel_denominador	<http://35.198.44.239:8080/d2rq/resource/tbl_variaveis/2>
vocab:tbl_indicadores_cod_variavel_numerador	<http://35.198.44.239:8080/d2rq/resource/tbl_variaveis/7>
vocab:tbl_indicadores_descricao_indicador	Expressa a efetividade do tratamento. O alcance das metas pactuadas para este indicador visa à redução da transmissão para novos pacientes, diminuindo a ocorrência de casos novos. Apresenta restrição no uso sempre que ocorra elevada proporção de registros sem informação do desfecho de tratamento ou grande volume de transferências não confirmadas pela vinculação de registros.
vocab:tbl_indicadores_nome_indicador	Proporção de casos de tuberculose curados
rdf:type	vocab:tbl_indicadores

The server is configured to display only a limited number of values (limit per property bridge: 50).

Metadata

<http://35.198.44.239:8080/d2rq/data/tbl_indicadores/4>	
dc:date	2018-08-29T08:27:10.046Z
prov:containedBy	<http://35.198.44.239:8080/d2rq/dataset>
void:inDataset	<http://35.198.44.239:8080/d2rq/dataset>
rdf:type	prov:DataItem
rdf:type	foaf:Document

Generated by D2R Browser

Fonte – Própria do autor

Na tela do servidor D2R apresentada na Figura 32, é mostrada uma instância da classe *tbl_indicadores* pertencente à ontologia que mapeia os indicadores. A tela apresenta todo o conjunto de informações que compõe o conhecimento a respeito do indicador “Proporção de

casos de tuberculose curados”. É possível visualizar a informação da URI que identifica e disponibiliza o acesso para o indicador, seu vocabulário, metadados e os links que conectam as informações como as que estão relacionadas aos dados das variáveis que compõe o cálculo (Figura 33).

Figura 33 – Dados das variáveis utilizadas pelos indicadores

Home All tbl_variaveis	
Property	Value
<code>rdfs:label</code>	tbl_variaveis #2
<code>is vocab:tbl_indicadores_cod_variavel_denominador of</code>	< http://35.198.44.239:8080/d2rq/resource/tbl_indicadores/1 >
<code>is vocab:tbl_indicadores_cod_variavel_denominador of</code>	< http://35.198.44.239:8080/d2rq/resource/tbl_indicadores/4 >
<code>is vocab:tbl_indicadores_cod_variavel_denominador of</code>	< http://35.198.44.239:8080/d2rq/resource/tbl_indicadores/5 >
<code>is vocab:tbl_indicadores_cod_variavel_denominador of</code>	< http://35.198.44.239:8080/d2rq/resource/tbl_indicadores/6 >
<code>is vocab:tbl_indicadores_cod_variavel_denominador of</code>	< http://35.198.44.239:8080/d2rq/resource/tbl_indicadores/8 >
<code>vocab:tbl_variaveis_cod_variavel</code>	2 (xsd:integer)
<code>vocab:tbl_variaveis_descricao_variavel</code>	Número de casos de tuberculose notificados em determinado período
<code>vocab:tbl_variaveis_nome_variavel</code>	totalPacNotificados
<code>vocab:tbl_variaveis_tipo_variavel</code>	denominador
<code>rdf:type</code>	vocab:tbl_variaveis
Property	Value
<code>rdfs:label</code>	tbl_variaveis #7
<code>is vocab:tbl_indicadores_cod_variavel_numerador of</code>	< http://35.198.44.239:8080/d2rq/resource/tbl_indicadores/4 >
<code>vocab:tbl_variaveis_cod_variavel</code>	7 (xsd:integer)
<code>vocab:tbl_variaveis_descricao_variavel</code>	Número de casos de tuberculose encerrados por cura por data de diagnóstico
<code>vocab:tbl_variaveis_nome_variavel</code>	nrCasosCurados
<code>vocab:tbl_variaveis_tipo_variavel</code>	numerador
<code>rdf:type</code>	vocab:tbl_variaveis
The server is configured to display only a limited number of values (limit per property bridge: 50).	

Fonte – Própria do autor

A Figura 33 apresenta as instâncias da classe de variáveis que apresentam por meio de um conjunto de informações o conhecimento associado às variáveis utilizadas (nrCasosCurados e totalPacNotificados) no cálculo do indicador “Proporção de casos de tuberculose curados”.

O servidor D2R disponibiliza o recurso de utilização de consultas SPARQL, seja diretamente pelo código implementado na página, ou por interface *web* (Figura 34). Na interface *web* apresenta-se com um layout simplificado, onde é exibido o endereço do servidor, os prefixos dos repositórios, a tela para a edição da consulta e a tela de resultado.

A Figura 34 apresenta uma consulta que lista os indicadores cadastrados na estrutura RDF montada pelo servidor D2R. A consulta apresenta o código, nome, descrição, numerador

e denominador associados. Na mesma consulta, foi aplicado um filtro para retornar apenas os dados referentes ao indicador “Proporção de casos de tuberculose curados” (Figura 35).

Figura 34 – Interface *web* para consultas SPARQL do D2R

Snorql: Exploring <http://35.198.44.239:8080/d2rq/sparql>

SPARQL:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX d2r: <http://sites.wiwiw.de/berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX meta: <http://www.wiwiw.de/berlin.de/bizer/d2r-server/metadata#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX vocab: <http://35.198.44.239:8080/d2rq/resource/vocab/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX map: <http://35.198.44.239:8080/d2rq/resource/#>
PREFIX db: <http://35.198.44.239:8080/d2rq/resource/>

select distinct ?codigo ?nome ?descricao ?numerador ?denominador
where {
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_cod_indicador> ?codigo .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_nome_indicador> ?nome .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_descricao_indicador> ?descricao .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_cod_variavel_numerador> ?numerador .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_cod_variavel_denominador> ?denominador .
}
```

Results:

SPARQL results:

codigo	nome	descricao	numerador	denominador
1	"Proporção de casos de tuberculose testados para HIV"	"Reflete o quantitativo dos casos de tuberculose que foram testados para HIV. Devido ao fato de a tuberculose ser a primeira causa de morte em pacientes portadores da aids, a identificação precoce dos casos HIV positivo torna-se importante para que um resultado satisfatório no desfecho de tratamento possa ser alcançado. Espora-se que 100% dos casos de tuberculose sejam testados para HIV."	db:tbl_variaveis/1	db:tbl_variaveis/2
2	"Proporção de coinfecção TB/HIV"	"Reflete a magnitude da coinfecção TB/HIV, além do dimensionamento e planejamento das atividades conjuntas com o Programa de aids. Apresenta restrição no uso sempre que ocorra elevada proporção do resultado do exame "em andamento", "não realizado" ou "sem informação"."	db:tbl_variaveis/3	db:tbl_variaveis/4
3	"Proporção de casos de tuberculose com HIV em andamento"	"Recomenda-se nacionalmente que todos os casos de tuberculose façam o teste anti-HIV. Atualmente, muitos exames são solicitados, mas as informações referentes ao seu resultado não são inseridas no sistema informatizado."	db:tbl_variaveis/5	db:tbl_variaveis/6
4	"Proporção de casos de tuberculose curados"	"Expressa a efetividade do tratamento. O alcance das metas pactuadas para este indicador visa à redução da transmissão para novos pacientes, diminuindo a ocorrência de casos novos. Apresenta restrição no uso sempre que ocorra elevada proporção de registros sem informação do desfecho de tratamento ou grande volume de transferências não confirmadas pela vinculação de registros."	db:tbl_variaveis/7	db:tbl_variaveis/2
5	"Proporção de casos de tuberculose que abandonaram o tratamento"	"Expressa a efetividade do tratamento. O alcance das metas pactuadas para este indicador visa a um melhor prognóstico do tratamento, assim como à redução do risco de ocorrência de resistência a drogas de primeira linha. Considera-se encerrado por abandono o caso em que o doente fica 30 dias ou mais sem tomar a medicação, para os casos em tratamento supervisionado, ou 30 dias após a última consulta agendada, quando o tratamento é autoadministrado. Apresenta restrição no uso sempre que ocorra eleva"	db:tbl_variaveis/8	db:tbl_variaveis/2
6	"Proporção dos casos de tuberculose com encerramento em óbito"	"Expressa a efetividade do tratamento. Apresenta restrição no uso sempre que ocorra elevada proporção de registros sem informação do desfecho de tratamento ou grande volume de transferências não confirmadas pela vinculação de registros. Locais com alta prevalência de infecção pelo HIV costumam apresentar altas taxas de óbito, o que prejudica a obtenção de meta de cura. No sistema de informação os óbitos podem ser classificados em óbitos por TB e óbitos por outras causas."	db:tbl_variaveis/9	db:tbl_variaveis/2

Fonte – Própria do autor

Figura 35 – Consulta de listagem de indicadores filtrando o indicador “Proporção de casos de tuberculose curados”

Snorql: Exploring <http://35.198.44.239:8080/d2rq/sparql>

SPARQL:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX d2r: <http://sites.wiwiw.de/berlin.de/suhl/bizer/d2r-server/config.rdf#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX meta: <http://www.wiwiw.de/berlin.de/bizer/d2r-server/metadata#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX vocab: <http://35.198.44.239:8080/d2rq/resource/vocab/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX map: <http://35.198.44.239:8080/d2rq/resource/#>
PREFIX db: <http://35.198.44.239:8080/d2rq/resource/>

select distinct ?codigo ?nome ?descricao ?numerador ?denominador
where {
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_cod_indicador> ?codigo .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_nome_indicador> ?nome .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_descricao_indicador> ?descricao .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_cod_variavel_numerador> ?numerador .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_indicadores_cod_variavel_denominador> ?denominador .
}

filter (?codigo = 4)
```

Results:

SPARQL results:

codigo	nome	descricao	numerador	denominador
4	"Proporção de casos de tuberculose curados"	"Expressa a efetividade do tratamento. O alcance das metas pactuadas para este indicador visa à redução da transmissão para novos pacientes, diminuindo a ocorrência de casos novos. Apresenta restrição no uso sempre que ocorra elevada proporção de registros sem informação do desfecho de tratamento ou grande volume de transferências não confirmadas pela vinculação de registros."	db:tbl_variaveis/7	db:tbl_variaveis/2

Fonte – Própria do autor

As variáveis associadas ao indicador de exemplo utilizados as consultas anteriores, entre outras informações, apresentam código, nome, descrição e tipo (numerador ou denominador) (Figura 36).

Figura 36 – Consulta aos dados das variáveis associadas ao indicador “Proporção de casos de tuberculose curados”

The screenshot shows a SPARQL query interface with the title "Snorql: Exploring http://35.198.44.239:8080/d2rq/sparql". The query is as follows:

```
SPARQL:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX d2r: <http://sites.wiwiiss.fu-berlin.de/sub1/bizer/d2r-server/config.rdf#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX meta: <http://www.wiwiiss.fu-berlin.de/bizer/d2r-server/metadata#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX vocab: <http://35.198.44.239:8080/d2rq/resource/vocab/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX nqi: <http://35.198.44.239:8080/d2rq/resource/#>
PREFIX db: <http://35.198.44.239:8080/d2rq/resource/>

SELECT DISTINCT ?codigo ?nome ?descricao ?tipo
WHERE {
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_variaveis_cod_variavel> ?codigo .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_variaveis_nome_variavel> ?nome .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_variaveis_descricao_variavel> ?descricao .
  ?resource <http://35.198.44.239:8080/d2rq/resource/vocab/tbl_variaveis_tipo_variavel> ?tipo .

  filter (?codigo in (2, 7))
}
```

The results are displayed in a table with the following data:

codigo	nome	descricao	tipo
2	"totalPacNotificados"	"Número de casos de tuberculose notificados em determinado período"	"Denominador"
7	"nrCasosCurados"	"Número de casos de tuberculose encerrados por cura por data de diagnóstico"	"Numerador"

Fonte – Própria do autor

A utilização do CMS Drupal para o desenvolvimento de um portal com marcação semântica mostrou-se totalmente viável. O CMS, mesmo entregando uma estrutura já pré-definida ao desenvolvimento de um *site* ou portal, permite que sejam feitas alterações em sua estrutura quando necessárias. Além das páginas que seguem um layout padrão de acordo com o *template* definido, também é possível agregar páginas codificadas independentemente ao *template*. O Drupal, é uma ferramenta altamente modular, o que facilita as alterações dos *templates* e módulos, pois, diminui o impacto das mudanças na estrutura como um todo.

A marcação semântica dos dados em parte do conteúdo do portal pode ser realizada de duas maneiras: a primeira via codificação mediante o uso de códigos e bibliotecas específicas, exigindo o conhecimento de um profissional técnico e, a segunda maneira foi por meio da utilização de módulos que podem ser facilmente habilitados no CMS, além de serem de fácil utilização permitindo que usuários não técnicos possam cuidar da marcação do conteúdo do portal. O Drupal tem integração nativa em seu *kernel* para a utilização dos recursos da web semântica (MOREIRA, 2012). Uma vez habilitada, formulários de publicação podem

ser preparados pelos próprios usuários a mapearem seu conteúdo a repositórios de ontologias. No exemplo deste portal, foi criada uma página de publicação de artigos que consome a ontologia Person do repositório Schema.org.

Os indicadores de avaliação e monitoramento das ações de combate a TB foram selecionados de acordo com a identificação da capacidade de geração de dados da fonte de dados utilizada nesse estudo, SISTB. Tais indicadores também foram adaptados de acordo com o estudo de Rondelli (2017). Os indicadores são particulares às necessidades do cenário brasileiro, portanto, não foram encontrados repositórios públicos de ontologias que contemplassem o conhecimento dos mesmos. Portanto, conforme já exposto anteriormente, a criação de um servidor próprio de ontologias utilizando a plataforma D2RQ atendeu a esta demanda.

Este portal compõe o conjunto de softwares existentes dentro do ecossistema de sistemas de gerenciamento integrado de pacientes de Tuberculose proposto. Seu objetivo principal no ecossistema é facilitar o acesso às informações com conhecimento por meio da *Web Semântica* atendendo às necessidades dos utilizadores como profissionais da saúde, estudantes, pacientes, entre outros, além de permitir que motores de busca e demais softwares inteligentes possam acessá-lo e processar as informações por intermédio das estruturas de dados baseadas em repositórios de ontologias (Figura 37).

Figura 37 - Representação da inclusão do Portal TB no Ecossistema SISTB

Fonte – Própria do autor

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho promoveu um estudo da utilização da *web* semântica para construção de portais de conteúdo com conhecimento para aplicação no contexto da TB. Considerando que a TB é uma doença milenar e que, portanto, existe um vasto conhecimento fundamento em pesquisas e práticas profissionais, este trabalho procurou explorar e buscar este conhecimento mapeado em diferentes repositórios públicos de ontologias que foram utilizados na construção do portal. Contudo, é importante pontuar que foram encontradas limitações para utilização das ontologias no contexto da TB quanto às práticas de controle e cuidados. Nota-se uma grande quantidade de informações definindo especificamente a doença, mas não as recomendações e processos de combate à doença. Essa condição dificultou a geração de um maior conteúdo no portal. Este estudo identificou, portanto, que há a necessidade de um grande esforço a ser realizado na criação de repositórios ou simplesmente de ontologias que promova um mapeamento com maior completude dos dados da TB para viabilizar de fato extração de um maior conhecimento.

Outro aspecto deste trabalho, foi a avaliação da viabilidade da utilização de uma ferramenta que facilitasse a produção de conteúdo por profissionais não técnicos, permitindo que usuários pudessem alimentar o portal com informações marcadas semanticamente. Portanto, a proposta de utilização do Drupal como ferramenta de CMS foi satisfatória. O Drupal mostrou-se robusto, permite a extensibilidade de suas capacidades, pois é uma ferramenta altamente modular, tem presença em grandes portais no âmbito público e privado e traz por padrão, a *web* semântica integrada em seu core. No entanto, é importante destacar que o Drupal apresenta uma longa curva de aprendizado em função de sua grande capacidade modularização.

O portal apresentou uma seção específica para a consulta de indicadores da TB. No entanto, nenhum repositório público de ontologias foi encontrado com informações destes. Essa dificuldade foi contornada neste trabalho, por meio da criação de um servidor próprio de ontologias mapeando os dados dos indicadores selecionados. Este cenário apresenta uma lacuna a ser preenchida e um longo caminho a ser percorrido para viabilizar a utilização da *web* semântica para a gestão da informação direcionada à TB.

Por fim, a execução deste trabalho em um conjunto de etapas organizadas e representada por um esquema metodológico é passível de reprodução por outros trabalhos.

Portanto, configura-se como caminho no desenvolvimento de portais semânticos baseados fundamentados na utilização de ontologias.

Como trabalhos futuros, pretende-se avaliar a qualidade e propor melhorias ou a criação de novas ontologias para ampliar o mapeamento do conhecimento relacionado à TB e promover uma melhor adequação às necessidades do cenário brasileiro. Não somente mapear dados da doença como nome, definição, tipos, medicamentos, entre outros dados básicos, mas é importante mapear as medidas de controle, políticas recomendadas, DOTs, maior número de indicadores, entre outras informações que venham a corroborar no sentido de aplicar o conceito e a teoria na prática, conectando repositórios, promovendo o desenvolvimento de softwares capazes de processar e tomar decisões baseadas em conhecimento, além da própria difusão do conhecimento. Também como trabalho futuro, propõe-se promover uma maior integração dos repositórios de dados de controle da TB no cenário brasileiro por meio da *web* semântica criando ontologias para padronizar as informações nessas bases possibilitando o compartilhamento de dados e diminuindo possíveis divergências e retrabalhos possibilitando uma maior contribuição na condução e avaliação das políticas públicas de saúde para o controle da TB, visando melhorar o trabalho do profissional, a qualidade de vida de pacientes e reduzir a incidência da doença na população.

7 REFERÊNCIAS

ABHISHEK, K.; SINGH, M. P. An ontology based decision support for tuberculosis management and control in India. **International Journal of Engineering and Technology**, v. 8, n. 6, p.2860–2877, Dec. 2016.

ARC2. **Home**. 2018. Disponível em: <<https://github.com/semsol/arc2/wiki>>. Acesso em: 7 set. 2018.

ARENS, Y. et al. Retrieving and integrating data from multiple information sources. **International Journal of Intelligent and Cooperative Information Systems**, v. 2, n. 2, p. 127–158, Mar. 1993.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, v. 5, n. 284, p. 2837, 2001.

BASIC FORMAL ONTOLOGY. **Home**. Disponível em: <<http://basic-formal-ontology.org/>>. Acesso em: 6 set. 2018.

BIOPORTAL. **Home**. 2018. Disponível em: <<http://bioportal.bioontology.org/>>. Acesso em: 5 set. 2018.

BITTAR, O. J. N. V. Sistemas de informação em saúde e sua complexidade. **Revista de Administração em Saúde**, v.18, n. 70, p. 1-18, jan.-mar. 2018.

BRIDGES, D. R. et al. Interprofessional collaboration: three best practice models of interprofessional education. **Medical Education Online**, v.16, n. 1, p. 1-10, 2011. Disponível em: <[10.3402/meo.v16i0.6035](https://doi.org/10.3402/meo.v16i0.6035)>. Acesso em: 7 set. 2018.

BYRON, A. et al. **Usando o Drupal**. Rio de Janeiro: Alta Books, 2009.

CHOI, B. C. K; PAK, A. W. P. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy:1. Definitions, objectives, and evidence of effectiveness. **Clinical and Investigative Medicine**, v. 29, n. 6, p. 351–364, Ago. 2006.

DBPEDIA. **About**. 2018. Disponível em: <<https://wiki.dbpedia.org/about>>. Acesso em: 7 ago. 2018.

DBPEDIA. **SPARQL Explorer**. 2018. Disponível em: <<http://dbpedia.org/snorql/>>. Acesso em: 7 ago. 2018.

DRUPAL. **Drupal case studies**. Disponível em: <<https://www.drupal.org/case-studies>>. Acesso em: 10 ago. 2018.

HARJITO, B; CAHYANI, D; DOEWES, A. Automatic bilingual ontology construction using text corpus and ontology design patterns (ODPs) in tuberculosis's disease. In: INTERNATIONAL CONFERENCE ON INFORMATICS AND COMPUTING, 1., 2016, Mataram. **Proceedings...** Piscataway: IEEE, 2016.

HITZLER, P; JANOWICZ, K. **Semantic web**: interoperability, usability, applicability. 2010. Disponível em: <<http://www.semantic-web-journal.net/sites/default/files/swj206.pdf>>. Acesso em: 5 set. 2018.

KRUG, S. **Não me faça pensar**. Rio de Janeiro: Alta Books, 2006.

LADAS, C. **Scrumban**: essays on kanban systems for lean software development. Seattle: Modus Cooperandi, 2008.

LEE, T.B. **Information management**: a proposal. 1990. Disponível em: <<http://cds.cern.ch/record/369245/files/dd-89-001.pdf>>. Acesso em: 5 set. 2018.

LIMA, J. C. de; CARVALHO, C. L. **Ontologias OWL (Web Ontology Language)**. Goiás: Instituto de Informática, 2005.

LIMA, R. R de et al. Development of a knowledge portal for tuberculosis based on semantic web. In: INTERNATIONAL CONFERENCE ON HEALTH AND SOCIAL CARE INFORMATION SYSTEMS AND TECHNOLOGIES, 2018, Lisboa. **Proceedings...** Lisboa: [s.n.], 2018a. p. 287-290.

LIMA, R. R de et al. Evaluating a middleware for a terminology service consumption. In: CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE, 2018, Fortaleza: **Anais...** Fortaleza: [s.n.], 2018b.

LIMA, V. C. et al. From guidelines to decision-making: using mobile applications and semantic web in the practical case of guides to support patients. In: INTERNATIONAL CONFERENCE ON HEALTH AND SOCIAL CARE INFORMATION SYSTEMS AND TECHNOLOGIES, 2017, Barcelona. **Anais...** Barcelona: [s.n.], 2017. p. 804-808.

MINISTÉRIO DA SAÚDE. **Manual de recomendações para o controle da tuberculose no Brasil**. Brasília: Ministério da Saúde, 2011. Série A, Normas e Manuais Técnicos.

MOREIRA, L. F. **O uso do CMS Drupal e de seus componentes no desenvolvimento de websites educacionais**: uma revisão bibliográfica. 2012. Disponível em: <https://www.basics.com.br/sites/default/files/artigos/IBTA_SOA25_LESLEY_FERNANDES_MOREIRA.pdf>. Acesso em: 5 set. 2018.

PANDYAN, M et al. Ontology with SVM based diagnosis of tuberculosis and statistical analysis. **International Journal of Engineering and Technology**, v. 3, n. 2, p. 37–43, Mar. 2016.

PELLISON, F.C et al. Development and evaluation of an interoperable system based on the semantic web to enhance the management of patients' tuberculosis data. In: INTERNATIONAL CONFERENCE ON HEALTH AND SOCIAL CARE INFORMATION SYSTEMS AND TECHNOLOGIES, 2017, Barcelona. **Anais...** Barcelona: [s.n.], 2017. P. 791-796.

RAWTANI, M. R.; CHIDAMBARAM, S. S. Drupal: the open source content management system software suit for library with library 2.0 features. In: INTERNATIONAL CALIBER, 7., Kalapet, 2009. **Proceedings...** Kalapet: [s.n.], 2009.

RONDELLI, J. P. P. **Desenvolvimento de uma ferramenta de geração de relatórios para auxiliar o acompanhamento do tratamento de tuberculose**. 2017. Monografia - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2017.

SCHEMA.ORG. **About Schema.org**. 2018. Disponível em: <<https://schema.org/docs/about.html>>. Acesso em: 5 set. 2018.

SINACI, A. A; ERTURKMEN, G. B. L. A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains. **Journal of Biomedical Informatics**, v. 46, n. 5, p.784–794, Oct. 2016.

STAHL, G. **Group cognition**: computer support for building collaborative knowledge. Cambridge: MIT, 2006.

TEIXEIRA, F.O. **Saúde 360**: aplicação dos princípios da *web* semântica para inter-relacionar informação no domínio da saúde. 2016. Tese (Doutorado) - Escola Paulista de Medicina, Universidade de Federal de São Paulo, São Paulo, 2016.

TRIPP, D. Action research: a methodological introduction. **Educação e Pesquisa**, v. 31, n. 3, p. 443-466, set./dez. 2005. DOI: 10.1590/S1517-97022005000300009.

TOMLINSON, T. **Beginning drupal 8**. Califórnia: Apress, 2015.

VALLE, E.D. **The need for semantic web service in the ehealth**. 2005. Disponível em: <<https://pdfs.semanticscholar.org/ce14/e7688120765061c4cf2f4d7328fa2628c6a1.pdf>>. Acesso em: 05 de set. de 2018.

VANDYCK, J. K. **Desenvolvimento profissional com o Drupal**. Rio de Janeiro: Alta Books, 2009.

WORLD HEALTH ORGANIZATION. **The stop TB strategy: building on and enhancing DOTS to meet the TB-related millennium development goals**. Geneva: World Health Organization, 2006.

_____. **Global tuberculosis report**. Geneva: World Health Organization, 2016a.

_____. **Tuberculosis country profiles**. 2016b. Disponível em: <<http://www.who.int/tb/country/data/profiles/en>>. Acesso em: 10 ago. 2018.

WORLD WIDE WEB CONSORTIUM. **Semantic web**. 2018a. Disponível em: <<https://www.w3.org/standards/semanticweb>>. Acesso em: 1 ago. 2018.

_____. **RDF 1.1 concepts and abstract syntax**. 2018b. Disponível em: <<https://www.w3.org/TR/rdf11-concepts/#section-Introduction>>. Acesso em: 7 ago. 2018.

_____. **SPARQL 1.1 query language**. 2018c. Disponível em: <<https://www.w3.org/TR/sparql11-query/>>. Acesso em: 7 ago. 2018.

_____. **Web Ontology Language (OWL)**. 2018d. Disponível em: <<https://www.w3.org/OWL/>>. Acesso em: 7 ago. 2018.

WORLD WIDE WEB CONSORTIUM. **Web ontology language overview**. 2018e. Disponível em: <<https://www.w3.org/TR/2004/REC-owl-features-20040210/>>. Acesso em: 07 de ago. de 2018.

_____. **Vocabularies**. 2018f.

Disponível em: <<https://www.w3.org/standards/semanticweb/ontology>>. Acesso em 07 de ago. de 2018.

_____. **HTML microdata**. 2018g. Disponível em: <<https://www.w3.org/TR/microdata/>>. Acesso em: 07 de set. de 2018.

APÊNDICE A – Tela principal do portal (*Home*)

Fonte – Própria do autor

APÊNDICE B – Tela sobre do menu tuberculose

Fonte – Própria do autor

APÊNDICE D – Tela listagem de artigos e trabalhos

Portal 3

SOBRE TUBERCULOSE PARCEIROS PRODUÇÃO CONTATO

PT-BR

q

Repositório de artigos e trabalhos

Artigo com anexo 04

Submitted by rrlimabr@gmail.com on seg, 07/23/2018 - 23:14
[Read more](#)

Testando a listagem de artigos e trabalhos com o artigo com anexo 04.

Artigo com anexo 03

Submitted by rrlimabr@gmail.com on sex, 07/20/2018 - 14:01
[Read more](#)

Testando a listagem de artigos e trabalhos com o artigo com anexo 03 com marcação semântica de conteúdo com base no repositório Schema.org.

Artigo com anexo 02

Submitted by rrlimabr@gmail.com on sex, 07/20/2018 - 13:10
[Read more](#)

Testando a listagem de artigos e trabalhos com o artigo com anexo 02 com a marcação dos campo Autor, Instituição, Atuação/Vínculo e Evento.

Fonte – Própria do autor

APÊNDICE E – Tela exibindo dados do artigo/trabalho e o link do arquivo

The screenshot displays the 'Portal TB' website interface. At the top, there is a navigation bar with the 'Portal TB' logo, a search bar with a magnifying glass icon, and a language dropdown menu set to 'PT-BR'. Below the navigation bar, a horizontal menu contains links for 'SOBRE', 'TUBERCULOSE', 'PARCEIROS', 'PRODUÇÃO', and 'CONTATO'. The main content area features the title 'Artigo com anexo 04' in a large, bold font. Below the title, a text box contains the submission information: 'Submitted by rrlimabr@gmail.com on seg, 07/23/2018 - 23:14' and 'Testando a listagem de artigos e trabalhos com o artigo com anexo 04.' To the right of this text is a circular icon with a document symbol. Below the submission information, there are several sections with icons and text: 'Arquivo PDF' with a PDF icon and the filename '01079umPT.pdf', 'Autor' with the name 'Ricardo Roberto de Lima', 'Instituição' with 'Universidade São Paulo', 'Atuação/Vínculo' with 'Aluno', and 'Evento' with 'HCist 2018'. At the bottom of the page, a copyright notice states 'Copyright © 2018. All rights reserved.' followed by links for 'Minha conta' and 'Sair'.

Portal TB

PT-BR

SOBRE TUBERCULOSE PARCEIROS PRODUÇÃO CONTATO

Artigo com anexo 04

Submitted by rrlimabr@gmail.com on seg, 07/23/2018 - 23:14
Testando a listagem de artigos e trabalhos com o artigo com anexo 04.

Arquivo PDF
01079umPT.pdf

Autor
Ricardo Roberto de Lima

Instituição
Universidade São Paulo

Atuação/Vínculo
Aluno

Evento
HCist 2018

Copyright © 2018. All rights reserved. [Minha conta](#) [Sair](#)

Fonte – Própria do autor

APÊNDICE F – Tela exibindo o indicador de proporção de pacientes curados

Fonte – Própria do autor

APÊNDICE G – Tarefas gerenciadas pelo Kanbanchi Task & Project Management

Nr	Atividades Pendentes	Atividades Realizadas
1	➤ Criar formulário de contato.	➤ Pesquisa bibliográfica.
2	➤ Implementar tradução dinâmica de conteúdo português->inglês.	➤ Estudos das tecnologias semânticas.
3	➤ Otimizar as consultas.	➤ Identificação dos repositórios fontes de dados.
4	➤ Melhorar a tradução do site.	➤ Definição dos indicadores.
5		➤ Buscar repositórios de ontologias de conteúdos gerais.
6		➤ Buscar repositórios de ontologias específicas da tuberculose.
7		➤ Instalar e configurar o CMS Drupal na versão 8.X
8		➤ Procurar e templates gratuitos para o layout do portal. Instalar e configurar template.
9		➤ Testar conexão com base de dados MySQL do banco do SISTB e testar a extração dados das tabelas necessárias para geração de indicadores.
10		➤ Criar estrutura de menu para organização do conteúdo do portal: 1 - SOBRE; 2 - TUBERCULOSE (Sobre, StopTB); 3 - PARCEIROS (Rede-TB, Geotb, OMS, OMS Brasil); 4 - PRODUÇÃO (Artigo, Portal LIS, Indicadores); 5 - CONTATO
11		➤ Adicionar chamadas (Links) nos itens de menu para chamadas às páginas externas e internas.
12		➤ Criar página básica para que usuários cadastrados no portal possam enviar artigos e trabalhos relacionados ao campo da tuberculose.
13		➤ Criar página para listagem de artigos e trabalhos cadastrados por usuários autorizados no portal.
14		➤ Criar página do menu TUBERCULOSE -> SOBRE
15		➤ Criar páginas de indicadores.
16		➤ Customizar layout do <i>template</i> .
17		➤ Analisar páginas e conteúdos em geral do portal que receberão a marcação semântica. Nos repositórios selecionados determinar as ontologias a serem utilizadas.
18		➤ Selecionar as ontologias nos repositórios e montar as consultas SPARQL.
19		➤ Marcar dados semanticamente com as ontologias selecionadas utilizando o recurso de microdata.
20		➤ Migrar estrutura de desenvolvimento do portal para o servidor cloud USP do LIS disponibilizado pelo laboratório de informática da Faculdade de Medicina de Ribeirão Preto.
21		➤ Alimentar o portal com conteúdo relacionado a tuberculose.

Fonte – Própria do autor