

**Nelson Júlio de Oliveira Miranda**

**Metodologia utilizando as ferramentas da Web  
Semântica para a criação de aplicações semânticas  
no domínio da saúde**

A handwritten signature in blue ink, appearing to be 'Nelson', is centered on the page.

**São Carlos  
Março de 2017**

**Nelson Júlio de Oliveira Miranda**

## **Metodologia utilizando as ferramentas da Web Semântica para a criação de aplicações semânticas no domínio da saúde**

Projeto de pesquisa apresentado ao Programa de Pós-Graduação em Ciências da Computação e Matemática Computacional do Instituto de Ciências Matemáticas e da Computação da Universidade de São Paulo - ICMC/USP, Campus de São Carlos, como requisito para participação no processo seletivo para o Doutorado em Ciências da Computação.

**Linha de Pesquisa:** Sistemas Web e Multimídia Interativos



**São Carlos  
Março de 2017**

## LISTA DE ABREVIATURAS E SIGLAS

Ajax	<i>Asynchronous Javascript and XML</i>
ASCII	<i>American Standard Code for Information Interchange</i>
CERN	<i>European Organization for Nuclear Research</i>
Cetic.br	Centro de Estudos sobre Tecnologia da Informação e Comunicação
CFM	Conselho Federal de Medicina
CMS	<i>Content Management System</i>
DAML	<i>DARPA Agent Markup Language</i>
DC	<i>Dublin Core</i>
eRDF	<i>Embeddable Resource Description Framework</i>
FOAF	<i>Friends of Friends</i>
GRDDL	<i>Gleaning Resource Description from Dialects of Languages</i>
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
ISO	<i>International Organization for Standardization</i>
MFC	<i>Meta Content Framework</i>
NCSA	<i>National Center for Supercomputing Applications</i>
NISO	<i>National Information Standards Organization</i>
OIL	<i>Ontology Inference Layer</i>
OWL	<i>Ontology Web Language</i>
PDF	<i>Portable Document Format</i>
PEP	Prontuário Eletrônico do Paciente
RDF	<i>Resource Description Framework</i>
RDFa	<i>Resource Description Framework in Attributes</i>
RDFS	<i>Resource Description Framework Schema</i>
RES	Registro Eletrônico em Saúde
RUP	<i>Rational Unified Process</i>
SPARQL	<i>SPARQL Protocol and RDF Query Language</i>
TIC	Tecnologias da Informação e da Comunicação
UML	<i>Unified Modeling Language</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>

W3C	<i>World Wide Web Consortium</i>
XHTML	<i>Extensible Hypertext Markup Language</i>
XML	<i>Extensible Markup Language</i>
XP	<i>Extreme Programming</i>
XSL	<i>Extensible Style sheet Language</i>
XSLT	<i>Extensible Style sheet Language for Transformation</i>

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>6</b>
1.1 Problema de pesquisa .....	6
1.2 Hipóteses .....	6
1.3 Objetivos .....	6
1.4 Resultados esperados .....	7
<b>2 REFERENCIAL TEÓRICO.....</b>	<b>7</b>
2.1 Representação do Conhecimento.....	10
2.1.1 Metadados .....	10
2.1.2 Vocabulários .....	12
2.1.3 Semântica dos Dados .....	12
2.1.4 Padrões Semânticos .....	13
2.2 Ontologias .....	14
2.2.1 Ontologias.....	15
2.2.2 OWL.....	16
2.2.3 Engenharia de Ontologias .....	17
2.3 Informação em Saúde .....	17
<b>3 ATIVIDADES, MATERIAIS E MÉTODOS.....</b>	<b>19</b>
<b>REFERÊNCIAS.....</b>	<b>21</b>

## 1. INTRODUÇÃO

Não por acaso que o exemplo utilizado por Tim Berners-Lee no artigo ***The Semantic Web*** aborda uma situação relacionada a área da saúde. As ferramentas nele propostas para a Web Semântica encontram um campo fértil neste domínio, onde os dados dos eventos relativos à assistência prestada aos paciente, durante toda a sua vida, aumentam de volume exponencialmente, da mesma forma que se espalham pelos diversos locais onde ele é atendido. Assemelhando-se em muito a uma das principais características da web: a descentralização.

### 1.1 Problema de Pesquisa

É exatamente neste contexto, que se pretende inserir a presente pesquisa, buscando respostas para a seguinte questão: Quais conjuntos de conjuntos de ferramentas da Web Semântica podem ser utilizados para construir aplicações Web que auxiliem os profissionais de saúde nas tomadas de decisões?

### 1.2 Hipóteses

Têm-se como hipóteses que a utilização das ferramentas da Web Semântica, tais como: XML, RDF/RDFS, OWL, SPARQL, RIF, entre outras; pode contribuir de forma decisiva para a melhoria da qualidade dos serviços de saúde. E que, a definição de uma metodologia que oriente sua aplicação dentro da área de saúde preencherá a lacuna entre a teoria e a prática, tanto na Web Semântica quanto na integração de dados dos prontuários eletrônicos do paciente.

### 1.3 Objetivos

O objetivo geral desta pesquisa é investigar e propor uma metodologia orientadora, que permita a criação de aplicações semânticas voltadas para a solução de problemas dentro do domínio da saúde.

Desse objetivo geral, desdobraram-se os seguintes objetivos específicos:

- a) Mapear as principais terminologias e ontologias em biologia e medicina (CID, SNOMED CT, MeSH, UMLS, OBO, entre outras);
- b) Identificar, importar e triplificar as bases de dados em saúde;
- c) Elaborar a modelagem ontológica;

- d) Construir um *Triple Store* para os dados;
- e) Criar aplicações semânticas utilizando os dados modelados;

De maneira que, esse estudo segue no campo das Ciências da Computação, especificamente, na linha de pesquisa que trata de Sistemas Web voltando nosso olhar para a sociedade, principalmente no que tange aos dados e informações em saúde.

#### **1.4 Resultados Esperados**

Espera-se, ao final deste projeto de doutorado, criar uma metodologia que permita explorar as ferramentas da Web Semântica voltadas para a solução de problemas no domínio da saúde. Para tanto, pretende-se utilizar bases de dados genéricas, a exemplo do RHC Fosp e Siscan. Além de utilizar como lócus de pesquisa a Faculdade de Medicina de Marília - Famema e o Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto - HCFMRP/USP.

## **2 REFERENCIAL TEÓRICO**

Inventada em 1990 por Sir Timothy John Berners-Lee (Tim Berners-Lee) no *European Organization for Nuclear Research* (CERN), baseada no projeto ENQUIRE, para compartilhar documentos (páginas) e, por este motivo, também chamada de Web dos documentos, a Web 1.0 conectou pessoas e possibilitou novas formas de aprendizagem a um clique de mouse.

Seus pilares estão sustentados nos identificadores *Universal Resource Identifiers* (URIs), no protocolo *Hypertext Transfer Protocol* (HTTP) e na linguagem de marcação *Hypertext Markup Language* (HTML), que permitem o acesso e a apresentação de documentos independente de sua localização física, utilizando um navegador (browser).

Marc Andreessen, ainda na *National Center for Supercomputing Applications* (NCSA), lançou o Mosaic 1.0 - primeiro navegador web. Para em seguida, fundar a Netscape e distribuir o Netscape Navigator - versão melhorada do Mosaic (ISOTANI; BITTENCOURT, 2015).

Durante a 1ª Conferência Internacional sobre a Web, ocorrida em maio de 1995, em Genebra, foi criado o *World Wide Web Consortium* (W3C) e Tim Berners-Lee proferiu uma palestra sobre semântica na web.

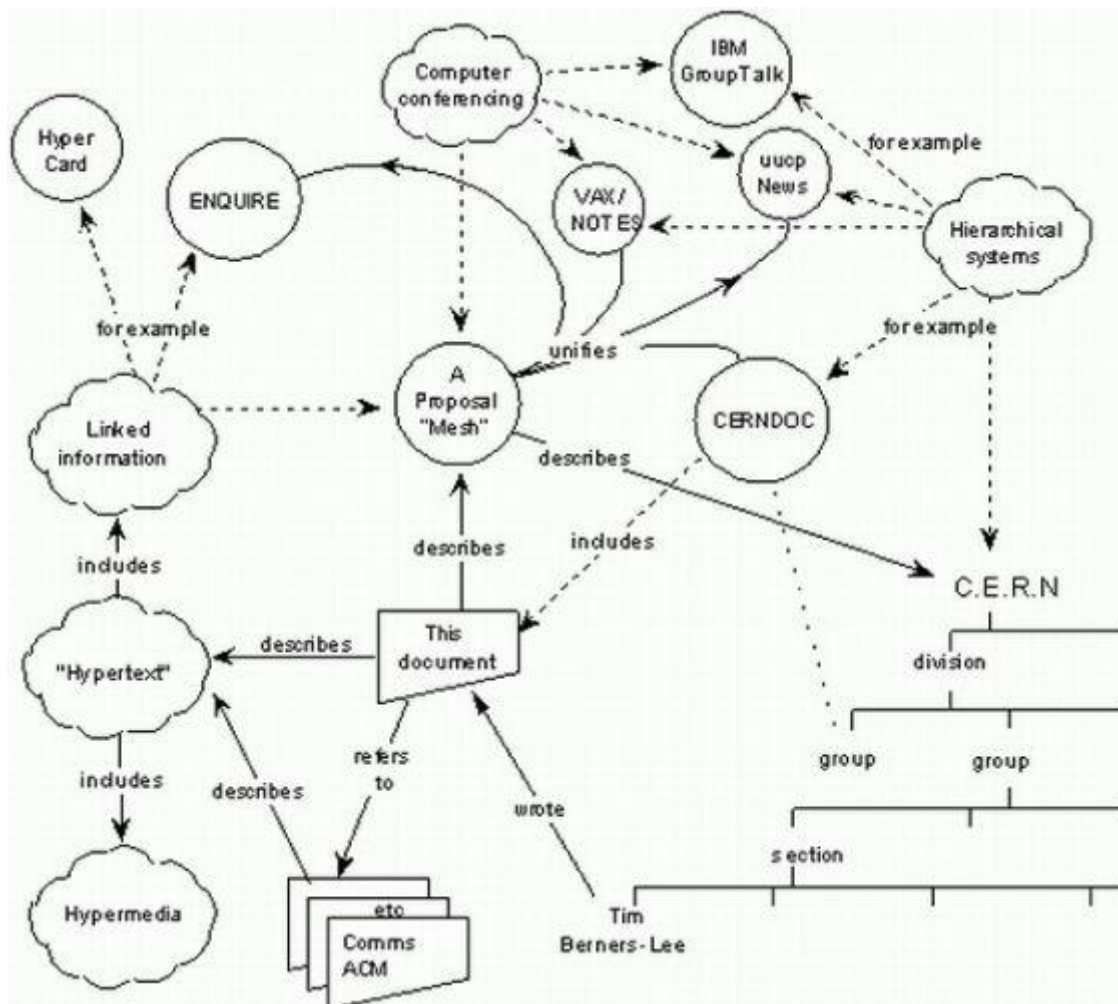


Figura 1 - Information Management: A Proposal

Já a Web 2.0 ou Web social, surge a partir dos anos 2000 com a criação dos *Content Management Systems* (CMS) que constituem os alicerces essenciais de plataformas de prestação de serviços de publicação e colaboração apoiadas em *blogs* e *wikis* que, por sua vez, só foram possíveis graças ao advento da tecnologia Ajax - acrônimo em inglês para *Asynchronous Javascript and XML* - em português JavaScript e XML Assíncrono.

Esta Web "programável" empoderou seus usuários, transformando-os de apenas consumidores para potenciais produtores e etiquetadores (*folksonomia*) de conteúdos. As linguagens de marcação *eXtensible Markup Language* (XML) e



*Extensible Stylesheet Language* (XSL) separaram a estrutura da apresentação; já os microformatos procuraram enriquecer semanticamente as páginas na Web.

O termo **Web Semântica** foi usado pela primeira vez por Tim Berners-Lee no livro *WEAVING THE WEB -The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*.

[...] o primeiro passo é colocar os dados na web de uma forma que as máquinas possam naturalmente entender, ou convertê-los para essa forma. Isso cria o que eu chamo de "Web Semântica" - uma Web de dados que podem ser processados direta ou indiretamente por máquinas (BERNERS-LEE; FISCHETTI, 2000, p.177, **tradução nossa**).

Esta Web Semântica, Web dos dados ou Web 3.0 estende a colaboração também para as máquinas, que se tornam capazes de analisar todos os dados na web - o conteúdo, links e transações entre pessoas e computadores.

Outro marco importante é o artigo *THE SEMANTIC WEB A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, publicado por Tim Berners-Lee, James Hendler e Ora Lassila em 2002. Nele, os autores demonstram o potencial da Web Semântica por meio de um exemplo da área da saúde, onde os irmãos Pete e Lucy precisam agendar sessões de fisioterapia para a sua mãe e utilizam agentes da Web Semântica para vasculhar os dados na Web, analisá-los e compartilhar os resultados com outros agentes (BERNERS-LEE, 2001, p. 25).

Na sequência, apresentam as tecnologias necessárias para a Web Semântica funcionar, quais sejam: Representação do Conhecimento (RDF, XML, etc.), Ontologias, Agentes e Serviços.

Muitas destas tecnologias já estavam postas. A Web Semântica não é uma Web separada, mas uma extensão da atual. Para tanto propõem-se uma pilha de tecnologias chamada "bolo de noiva".

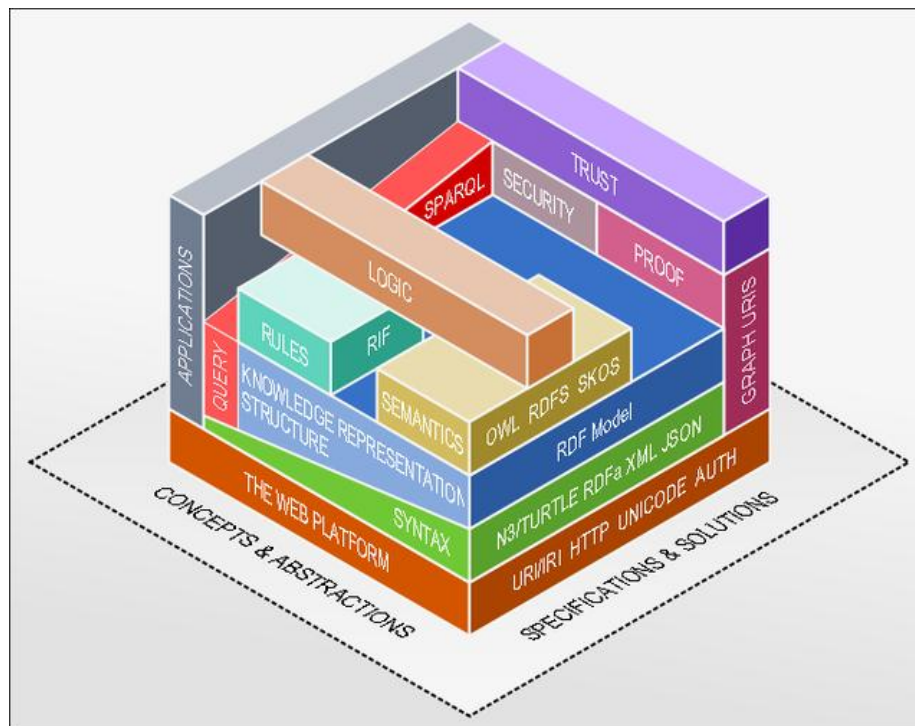


Figura 2 - The Semantic Web Technology Stack

## 2.1 Representação do Conhecimento

Coleções de informações e conjuntos de regras de inferência formam a base da Representação de Conhecimento. Para a Web Semântica funcionar, os computadores devem ter acesso a elas para conduzir raciocínios de forma autônoma.

O elemento central da Representação do Conhecimento é a utilização das mesmas definições para os conceitos por todos os envolvidos: pessoas e máquinas; sendo que um termo serve para referenciar um conceito específico (FEITOSA, 2006, p.73).

O desafio da Web Semântica é fornecer uma linguagem para expressar dados e regras para raciocinar sobre estes dados.

### 2.1.1 Metadados

A definição mais usual para metadados é que "são dados sobre outros dados". Contudo, diante da importância dos mesmos para a Web Semântica, a definição da *National Information Standards Organization* (NISO) expande esta

definição para uma estrutura que descreve, explica e localiza recursos digitais, facilitando sua recuperação, uso e gerenciamento .

Metadados, segundo Alves (2010), são descritos como:

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação (ALVES, 2010, p.47).

A primeira iniciativa para embutir metadados no código fonte de uma página Web, tais como: título, autor, palavras-chave, copyright, idioma, etc.; foi através do elemento <meta> da linguagem HTML. Muito utilizada para facilitar a indexação de páginas web pelos mecanismos de busca.

Pode-se também embutir metadados em página Web por meio da utilização de Microformatos e Microdados. Ambos, visam enriquecer semanticamente os documentos publicados na Web, permitindo que máquinas possam, minimamente, entender as informações da página dentro do contexto semântico.

Microformatos são coleções de tags para embutir metadados em HTML, pequenos e livres, dão mais significado as páginas web, porém os metadados são mais rígidos e sensíveis. O microformato RDFa permite a inclusão de metadados baseados no RDF tanto em documentos XHTML quanto em XML. O *Embeddable RDF* (eRDF), apesar de não ser uma recomendação do W3C, extrai informação utilizando um analisador RDF, mas somente em documentos HTML. O *Gleaning Resource Description from Dialects of Languages* (GRDDL), assim como o RDFa, também é uma recomendação do W3C, e utiliza scripts de transformação XSLT para extrair RDF de XHTML e XML. As especificações hCard (contato) e hCalendar (evento) são exemplos de microformatos.

Microdados permitem marcações semânticas, são atributos colocados nas tags HTML utilizando elementos de um vocabulário *Schema.org*. O Google reconhece microdados.

### 2.1.2 Vocabulários

No famoso artigo *THE SEMANTIC WEB* de Tim, James e Ora (2001), alguns termos são destacados e os autores afirmam que a semântica e/ou significado, foram definidos previamente no contexto da Web Semântica. Esta definição é feita pela especificação de um **Vocabulário**.

Os autores retomam a discussão sobre a real importância do entendimento do significado dos termos afirmando que duas bases de dados podem utilizar diferentes identificadores para o mesmo conceito, tal como **cep**. A utilização de vocabulários comuns resolve parte deste problema.

Os vocabulários são usados para descrever dados, representar conceitos e identificar um conjunto de dados. Um vocabulário pode ser entendido como um conjunto de termos que descrevem dados (predicados, atributos ou propriedades), aumentando a expressividade e reduzindo a ambigüidade (FREITAS;SCHULZ, 2009).

Alguns dos vocabulários mais utilizados:

- *Friends of Friends* (FOAF)
- *Dublin Core* (DC)
- *GeoNames* (GN)
- *Good Relations* (GR)
- *WordNet*
- *Food*
- *Schema.org*

### 2.1.3 Semântica dos Dados

A maneira mais tradicional de se estruturar os dados é organizá-los em tabelas, a exemplo das planilhas em Excel e a tag <table> em HTML. Tabelas facilitam mostrar, ordenar, imprimir e editar os dados. A semântica dos dados é garantida pelos rótulos colocados no topo de suas colunas e início de suas linhas, para uma pessoa é relativamente fácil entender o que representa determinado dado analisando estes identificadores. Contudo, tabelas apresentam dificuldades para representar valores múltiplos e geram duplicidades quando da junção de várias

delas (SEGARAN, 2009).

Segaran (2009) aponta que nos Bancos de Dados Relacionais, como o Oracle e o MySQL é muito rápido e fácil unir tabelas, utilizando atributos comuns. O que permite criar sofisticadas consultas. A semântica fica a cargo dos esquemas de banco de dados, que definem entidades, relacionamentos e rótulos para os dados. Entender o significado dos dados é uma questão de conhecer o esquema. Da mesma forma que os esquemas são o ponto forte dos bancos relacionais, representam seu "calcanhar de Aquiles" no que tange a integração de dados, migração e complexidades dos esquemas.

Segundo Tim Berners-Lee (2001), a propriedade essencial da Web é a universalidade. Estruturar dados na forma de tabelas ou bancos relacionais, trás inúmeras dificuldades para garantir tal propriedade. Para tanto, faz-se necessário uma nova modelagem dos dados. Aliás, não tão nova assim. Um modelo mais flexível, onde as relações sejam descritas pelos próprios dados, onde eles mesmos se auto-descrevam. Um modelo baseado em chave/valor, que seja a prova de futuro (SEGARAN, 2010, p.12-15).

Este modelo de dados semântico é baseado em triplas, blocos de construção fundamentais para a semântica, onde os dados são estruturados em declarações do tipo: <sujeito> <predicado> <objeto>. Onde <sujeito> pode representar qualquer coisa do universo, <predicado> é uma propriedade deste sujeito como nome ou e-mail, e <objeto> pode ser um sujeito noutra tripla ou armazenar valores literais: caracteres e números (ISOTANI;BITTENCOURT, 2015).

#### **2.1.4 Padrões Semânticos**

Segundo Davies (2006), os padrões semânticos estão baseados nos seguintes conceitos:

- *Resource Description Framework* - RDF
- *Uniform Resource Indentifiers* - URIs
- *SPARQL Protocol and RDF Query Language* - SPARQL

O RDF é um padrão para expressar grafos para permitir o compartilhamento de dados entre pessoas e, principalmente, entre máquinas. Recomendado pelo

W3C, baseou-se no *Meta Content Framework* (MFC), proposto por Ramanathan Guha e Tim Bray em 1990 para descrever metadados e organizar a Web. Sua primeira versão foi lançada em 1990. Por ser uma recomendação W3C, possui uma variedade de ferramentas e serviços disponíveis (POLLOCK, 2010).

As declarações RDF seguem o mesmo modelo de triplas <sujeito> <predicado> <objeto> para descrever recursos, que podem representar qualquer coisa e todas as coisas.

Uma URI é uma string que referencia de forma única um recurso, que pode ser qualquer coisa, eletrônica ou não. Por isso, é chamada de *URI reference* ou *URIref*.

Diferente da URL (*Uniform Resource Locator*), que é um subconjunto da URI e identifica onde a informação digital pode ser recuperada, provendo uma identificação única; a URI apenas fornece uma chave forte unívoca para um recurso que se queira representar, seja ele uma pessoa, uma cidade, um livro, etc. (LAUFER, 2015).

## 2.2 Ontologias

Apontada como o terceiro componente básico da Web Semântica, as Ontologias são coleções de informações. Originárias da Metafísica, área da Filosofia, que estuda a natureza do ser e a existência.

Na Computação, uma ontologia é um conjunto de conceitos e suas relações que busca representar o entendimento de um domínio, através de uma representação formal compreensível tanto por humanos quanto por computadores.

Segundo Gruber (1993), ontologia é uma especificação explícita de um conceitualização. Entendendo por conceitualização o significado dos conceitos e as relações entre eles, e especificação como a representação formal destes conceitos e relações.

Para Guarino (2009), a modelagem de ontologias tem o objetivo de compartilhar e reutilizar o conhecimento em diferentes aplicações.

## 2.2 Tipos de Ontologias

As ontologias são constituídas por um conjunto de conceitos essenciais sobre um domínio, representados pelos vocabulários especializados; um corpo de conhecimento numa hierarquia de classes (C), um conjunto de relações (R) e restrições semânticas ou axiomas (A) (ISOTANI, 2015).

Que somados ao conjunto de instâncias (I), completam o conjunto de elementos de uma ontologia  $O = \{C, R, I, A\}$ .

Pode-se classificar as ontologias quanto ao porte em leves ou pesadas, e quanto a área de atuação de domínio ou de tarefa.

As leves enfatizam as taxonomias para categorizar grandes volumes de dados, já as pesadas primam pelo rigor na definição, organização e formalização dos conceitos e relações - fundamental para o compartilhamento e reuso de bases de conhecimento.

As ontologias de domínio descrevem onde as tarefas são executadas, representando o conhecimento sobre um assunto.

As de tarefa detalham os processos e atividades necessárias para resolver um problema, isto é, a habilidade de aplicar o conhecimento de um domínio; seus conceitos podem ser categorizados em: *tasks roles*, *tasks actions* e estados dos objetos/fluxos de informação (MIZOGUCHI, 2004).

Existem duas maneiras de representar uma ontologia: formal e gráfica.

A formal é consumida por computadores e utiliza linguagens para a representação, tais como: lógica de predicados, lógica descritiva (frames) e descrição de ontologias. Desta última, as mais conhecidas (recomendações do W3C) são RDF/RDF-S e OWL.

A representação gráfica, destinada a pessoas, utiliza formas como: grafos, UML e estrutura de árvore. Tanto a representação formal quanto a gráfica contribuem de forma decisiva para a qualidade da ontologia (LAUFER, 2015).

### 2.2.2 OWL

A *Ontology Web Language* (OWL) nasceu da combinação das linguagens OIL (*Ontology Inference Layer*), desenvolvida pela comunidade de defesa europeia e da DAML (*DARPA Agent Markup Language*), criada pelo departamento de defesa americano. Logo tornou-se uma recomendação do W3C.

Baseada no RDF, de onde herda a estrutura em triplas e a descrição de recursos com URI, e no RDF-S ao herdar sua semântica; é uma linguagem declarativa que descreve o conhecimento e permite o uso de ferramentas de inferência. Porém, não delimita uma estrutura sintática (POLLOCK, 2010).

Armazena informações (instâncias) dentro do conceito de mundo aberto (*Open-World Assumptions*) onde, se um fato não está presente, ele é considerado desconhecido, e não falso como ocorre no mundo fechado (*Closed-World Assumptions*) dos bancos de dados tradicionais (LAUFER, 2015).

Do ponto de vista da sintaxe, é possível serializar documentos OWL nos formatos: RDF/XML (padrão), OWL/XML, Turtle, entre outros. Sempre considerando aspectos importantes da modelagem de conhecimento, tais como: entidades, expressões e axiomas.

Do ponto de vista da semântica, OWL estende o RDF-S e é uma especialização da Lógica de Descrição.

Pode-se especificar semântica em OWL de duas formas: Semântica Direta (OWL 2 DL) e Semântica baseada em RDF (OWL 2 Full). Sendo a primeira mais simplificada e baseada Lógica de Descrição e Semântica Direta; e a segunda, grafos RDF e na própria RDF (ISOTANI;BITTENCOURT, 2015).

A versão 2 da OWL possui diversos perfis, projetados para atender as mais variadas necessidades com diferentes níveis de expressividade:

- OWL 2 DL: usado para especificar classes complexas e axiomas;
- OWL 2 QL: voltado para bancos de dados relacionais (SQL);
- OWL 2 RL: próprio para aplicações com necessidades de raciocínio .



### 2.2.3 Engenharia de Ontologias

As metodologias para a criação de ontologias (engenharia de ontologias) originam-se em três áreas da computação: Inteligência Artificial, Engenharia de Software e Interação Humano-Computador (ISOTANI; BITTENCOURT, 2015):

- **Inteligência Artificial:** TOVE, Enterprise Model Approaches e IDEF5;
- **Engenharia de Software:** RUP, XP, Methondology, Ontology Development 101 e RapidOWL;
- **IHC:** DILIGENT e HCOME.

Dentre as ferramentas utilizadas para a criação de ontologias, podemos citar:

- Appolo;
- OntoStudio;
- Hozo;
- Protégé;
- TopBraid Composer

## 2.3 Informação em Saúde

No entendimento de Bentes Pinto (2014), nas organizações de saúde, independentemente, do tipo de atividades de atenção primária, secundária ou terciária, as ações pertinentes aos pacientes geram como produto um conjunto enorme de registros que vão constituir a chamada **documentação sanitária**. Essa documentação é integrada tanto pelos registros de **dados e informações clínicas** concernentes à atenção primária, especializada ou em outros níveis, de cuidados e atenção dispensados aos pacientes. Bem como outros registros de **dados e informações não clínicos**, porém, necessários ao processo assistencial de um cidadão para o seu melhor atendimento nas organizações de saúde. Esses dois tipos de documentação serão discutidos nos parágrafos seguintes:

- a) Documentação clínica:** é aquela que está relacionada diretamente ao estado de saúde da pessoa. Sendo, portanto, produzida em consequência da atenção primária de cuidado a ela dispensada. São documentos de uso, cujos registros são originários da atenção primária e de documentos de uso da organização de saúde.
- b) Documentação não clínica:** é aquela, que mesmo sendo fundamental no processo assistencial, não está relacionada com

dados sanitários. São exemplos desse tipo de documentação, aquelas referentes às atividades administrativas, nutricionais, da farmácia, de manutenção, protocolos etc. (BENTES PINTO, 2014, p. 3-4)

Em se tratando da Documentação Sanitária do tipo clínica, o Prontuário do Paciente é tido como o seu representante maior. O nome prontuário provém do latim *prontuarium*, lugar em que se guardam as coisas que devem estar à mão, despensa, armário. Daí, por extensão, manual de informações úteis; de *promptus*, preparado, que está à mão; de *promere*, tirar uma coisa de onde está guardado, fazer sair. (Houaiss, 2001).

A Resolução do Conselho Federal de Medicina nº 1638 de 10 de julho de 2002, define prontuário da seguinte maneira

[...] documento único constituído de um conjunto de informações, sinais e imagens registradas, geradas a partir de fatos, acontecimentos e situações sobre a saúde do paciente e a assistência a ele prestada, de caráter legal, sigiloso e científico, que possibilita a comunicação entre membros da equipe multiprofissional e a continuidade da assistência prestada ao indivíduo (CONSELHO FEDERAL DE MEDICINA, 2009, p. 1).

Citando a *International Organization for Standardization* (ISO, 2015), Galvão e Ricarte (2011, p.78), definem Prontuário Eletrônico como

[...] a coleção de informação computadorizada relativa ao estado de saúde de um sujeito, armazenada e transmitida em completa segurança, acessível a qualquer usuário autorizado. O prontuário eletrônico estabelece um padrão lógico para a organização da informação, aceito universalmente e independente do sistema. Seu principal objetivo é assegurar serviços de saúde integrados de forma contínua, eficiente e com qualidade, juntamente com informação retrospectiva, corrente e prospectiva.

Na prática, o PEP, como é chamado o Prontuário Eletrônico do Paciente, é composto por diversos objetos digitais complexos, ou seja, uma combinação de objetos digitais simples, tais como: textos, imagens, sons, seus identificadores e metadados. Reunidos em formulários ou fichas para registrar os atendimentos ambulatoriais e de emergência; evoluções médicas, de enfermagem e dos demais profissionais de saúde; prescrições; exames complementares e seus respectivos laudos (resultados); descrições cirúrgicas; resumos de altas; entre outros.

Contudo, o PEP, de uso interno e exclusivo da organização de saúde, evoluiu para o conceito de Registro Eletrônico em Saúde (RES) – cuja premissa básica é o compartilhamento das informações sobre a saúde dos cidadãos dentro e fora da

organização de saúde onde o atendimento é prestado; seja nas esferas municipal, estadual ou federal.

### 3 ATIVIDADES, MATERIAIS E MÉTODOS

A definição dos procedimentos metodológicos, assim como o planejamento minucioso de todas as fases que irão compor a pesquisa, é de suma importância para que se alcancem os objetivos propostos. O objetivo da pesquisa é gerar novas possibilidades para a Web Semântica no âmbito da saúde, o que a classifica de natureza teórica-aplicada. Assim o tipo de pesquisa que norteará esse trabalho contempla os chamados estudos exploratórios e qualitativos. Conforme Gil (2010), as pesquisas exploratórias buscam o aprimoramento de ideias e a descoberta de intuições, permitindo maior familiaridade com o problema através de levantamentos bibliográficos, entrevistas e análise de exemplos. Contudo, podem contribuir para apontar hipóteses de estudos futuros. No que tange aos procedimentos técnicos, a presente pesquisa baseia-se no levantamento e revisão bibliográfica, experimentos, elaboração de suposições e definição de novas metodologias.

Propõe-se a realização das seguintes atividades, apresentadas no quadro 1.

Quadro 1 – Atividades da pesquisa

Atividade	Descrição
1. Créditos	Obtenção de créditos em disciplinas obrigatórias e opcionais de acordo com as exigências do programa de pós-graduação.
2. Revisão da Literatura	Levantamento bibliográfico abrangente e minucioso da literatura a cerca do estado da arte nos domínios da Web Semântica, Dados Conectados e Informação em Saúde.
3. Fontes de Dados	Obtenção de bases de dados de saúde para triplificação e modelagem ontológica para a criação de <i>triple store</i> que sirva de base para a criação de aplicações semânticas e definição de uma metodologia.
4. Reuniões Grupo de Pesquisa	Realização de reuniões periódicas com integrantes do grupo de pesquisa em Sistemas Web e Multimídia Interativos.
5. Qualificação	Redação do documento e realização do exame de qualificação.
6. Teste da Metodologia	Realização de experimentos, utilizando as bases obtidas, para validar a metodologia proposta.
7. Redação de Artigos	Redação e submissão de artigos científicos relativos ao projeto a eventos e periódicos nacionais e internacionais.
8. Redação da Tese	Elaboração da redação da tese final do doutorado.
9. Defesa da Tese	Apresentação e defesa da tese contendo os resultados do projeto.

Fonte: Elaborado pelo autor.

Para a definição da metodologia, pretende-se utilizar bases de dados com informações sobre câncer, obtidas na Faculdade de Medicina de Marília – Famema, nos ambulatorios de oncohematologia adulto e infantil e no serviço de radioterapia. Além de bases genéricas, tais como: RHC Fosp e Siscan. Pretende-se, também, utilizar bases de traumatologia, obtidas no Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto – HCFMRP/USP.

O cronograma, ilustrado na tabela 1, mostra quais serão as macro atividades, suas respectivas durações e seqüência de execução, iniciando em agosto de 2017 e finalizando até junho 2021.

Tabela 1 – Cronograma de atividades

CRONOGRAMA																
Atividades	2017		2018				2019				2020				2021	
	3º	4º	1º	2º	3º	4º	1º	2º	3º	4º	1º	2º	3º	4º	1º	2º
1. Créditos																
2. Revisão																
3. Dados																
4. Reuniões																
5. Qualificação																
6. Teste																
7. Artigos																
8. Redação																
9. Defesa																

Fonte: Elaborado pelo autor.

## REFERÊNCIAS

ALVES, Rachel C.V.; SANTOS, Plácida L.V.A.C; **Metadados no Domínio Bibliográfico**. Rio de Janeiro: Intertexto, 2013.

AMARAL SERIQUE, Kleberson Junio; **Anotação de imagens radiológicas usando a web Semântica para colaboração científica e clínica**, dissertação de mestrado, 92 p São Carlos 2012.

BENTES PINTO, Virginia. **Prontuário Eletrônico do Paciente: Documento Técnico de Informação e Comunicação do Domínio da Saúde**. Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n.21, 1º sem. 2006.

BENTES PINTO, Virginia. **A contribuição da terminologia para a gestão da documentação sanitária em organizações de saúde**. In: MEDINFOR, 3., 2014, Salvador. **Anais...** Salvador: UFBA, 2014.

BENTES PINTO, Virginia; SILVA NETO. Casemiro. Representação indexal como mediação informacional de prontuário do paciente. In: BENTES PINTO, Virginia; CAMPOS, Henry de Holanda. **Diálogos paradigmáticos sobre informação para a saúde**. Fortaleza: Edições UFC, 2013.

BERNERS-LEE, Tim et. al.; **The Semantic Web**: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American. Maio 2001, disponível em <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>, Acesso em: Março 2017.

BERNERS-LEE, Tim; FISCHETTI, Mark; **Weaving th Web**: The original Design and Ultimate Destiny of the World Wide Web by Its Iventor. New York: HarperBusiness, 2000.

CERVO, A C.; BERVIAN, P. A. **Metodologia científica**. 4.ed.. São Paulo: Makron, 1996.

CONSELHO FEDERAL DE MEDICINA. **Código de Ética Médica 2009: Resolução CFM nº 1.931/2009**. Aprova o Código de Ética Médica. Diário Oficial da União. 24 set. 2009. Disponível em: <<http://www.portalmédico.org.br/novocodigo/integra.asp>>.

CONSELHO FEDERAL DE MEDICINA. **RESOLUÇÃO CFM nº 1.605/2000**. O médico não pode, sem o consentimento do paciente, revelar o conteúdo do prontuário ou ficha médica. Revoga-se a Resolução CFM nº 999/1980. **Diário Oficial da União**, Brasília, 29 set. 2000.

CONSELHO FEDERAL DE MEDICINA. **Resolução CFM nº 1.638/2002**. Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde. **Diário Oficial da União**, Brasília, 9 ago. 2002. Disponível em: <[http://www.portalmédico.org.br/resolucoes/cfm/2002/1638\\_2002.htm](http://www.portalmédico.org.br/resolucoes/cfm/2002/1638_2002.htm)>.

CONSELHO FEDERAL DE MEDICINA. **Resolução CFM nº 1.821/2007**. Aprova as normas técnicas concernentes à digitalização e uso dos sistemas informatizados para a guarda e manuseio dos documentos dos prontuários dos pacientes, autorizando a eliminação do papel e a troca de informação identificada em saúde. **Diário Oficial da União**, Brasília, 23 nov. 2007. Disponível em: <[http://www.portalmédico.org.br/resolucoes/cfm/2007/1821\\_2007.htm](http://www.portalmédico.org.br/resolucoes/cfm/2007/1821_2007.htm)>.

DAVIES, John et. al. **Semantic Web technologies**: trends and research in ontology-based systems. West Sussex, John Wiley & Sons Ltd: 2006.

DODDS, Leigh; NEWMAN, Andrew. **Guia do Modelo de Maturidade de Dados Abertos: Avaliando a publicação e a utilização de dados abertos**. São Paulo: Secretaria de Governo, 2015.

FEITOSA, Ailton; **Organização da informação na Web: das tags à Web Semântica**. Brasília: Thesaurus, 2006.

FERREIRA, Aurélio Buarque de Holanda. **Dicionário Aurélio básico da língua portuguesa**. Rio de Janeiro: Nova Fronteira, 2006.

FREITAS, Fred; SCHULZ, Stefan; **Ontologias, Web semântica e saúde**. RECIIS - R. Eletr. de Com. Inf. Inov. Saúde, Rio de Janeiro, v.3, n.1, p-4-7, mar 2009.

FREITAS, Fred; SCHULZ, Stefan; **Pesquisa de terminologias e ontologias atuais em biologia e medicina**. RECIIS - R. Eletr. de Com. Inf. Inov. Saúde, Rio de Janeiro, v.3, n.1, p-8-20, mar 2009.

GALVÃO, M.C.B.; RICARTE, I.L.M.. **O prontuário eletrônico do paciente no século xxi: contribuições necessárias da ciência da informação**. R. Ci. Inf. e Doc., 2(2):77–100, jul/dez 2011

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

GRUBBER, Thomas R. **A Translation Approach to Portable Ontology Specifications**. Knowledge Systems Laboratory - Technical Report KSL 92-71, April 1993.

GUARINO, Nicola et. al. **What Is an Ontology**. Handbook on Ontologies, International Handbooks on Information Systems, Berlin, 2009.

HOUAISS, Antônio. **Dicionário Houaiss da Língua Portuguesa**. Rio de Janeiro, Ed. Objetiva, 2001.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados abertos conectados**. São Paulo: Novatec Editora, 2015.

LAUFER, Carlos; **Guia de web semântica**. São Paulo: Secretaria de Governo, 2015.

MIZOGUCHI, R. **Tutorial on ontological engineering: part 3: Advanced course of ontological engineering**. New Generation Computing, v. 22, n.2, p. 198-220, 2004.

POLLOCK, Jeffrey T.; **Web Semântica para Leigos**. Rio de Janeiro: Alta Books, 2010.

RAMALHO, Rogério Aparecido Sá. **Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação**. 2006. 120f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2006.

REINHARD, Nicolau; GERMANO, Edson Carlos. **Guia de Incentivo ao (Re)uso de Dados Abertos**. São Paulo: Secretaria de Governo, 2016.

SEGARAN, Toby et. al. **Programming the Semantic Web by: Build Flexible Applications with Graph Data**. Sebastopol: O'Reilly, 2009.

SOUZA, Amanda Damasceno et. al.; **A Informação em Oncologia na Era do Big Data.**

XVI Enancib, UFPB, 2015. Disponível em:

<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015/paper/viewFile/3046/1270>. Acesso em: mar 2017.

TIC SAÚDE, Cetic.br. Disponível em:

<[http://cetic.br/media/docs/publicacoes/2/tic\\_saude\\_2014\\_livro\\_eletronico.pdf](http://cetic.br/media/docs/publicacoes/2/tic_saude_2014_livro_eletronico.pdf)>. Acesso em: 22 jun. 2016.