



Materialização da Web Semântica: um modelo de construção dinâmica de consultas baseados em mapeamento de ontologias

Caio Saraiva Coneglian

Doutorando e mestre no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista. Bacharel em Ciência da Computação pelo Centro Universitário Eurípides de Marília

José Eduardo Santarem Segundo

Professor Doutor da Universidade de São Paulo. Doutor em Ciência da Informação pela Universidade Estadual Paulista

<http://dx.doi.org/10.1590/1981-5344/2728>

Aproximar a linguagem computacional da humana é um desafio no processo de Recuperação da Informação. Nessa perspectiva, a Web Semântica tem permitido novos meios de tratamento dos dados contidos na Web, afim de permitir melhor compreensão do significado destas informações. Contudo, alguns dos conceitos que envolvem a Web Semântica, como as ontologias, apresentam alta complexidade, dificultando o seu uso. A dificuldade de compreensão, a estrutura que estas possuem, e a diversidade como elas são construídas, tornam o manuseio de ontologias para a descoberta do significado dos conceitos, algo não efetivo, impossibilitando oferecer um panorama concreto do sentido dos conceitos. Assim, essa pesquisa tem como objetivo propor um modelo que verifique o contexto e o significado de um conceito dentro de uma ontologia. Para tal, foi realizado pesquisa bibliográfica, construção do modelo e a implementação de um protótipo, como prova de conceito, para a verificação dos resultados. O modelo e a implementação apresentam como resultado o tratamento genérico de ontologias, obtendo o significado e o contexto que um conceito possui. Baseados nos resultados identifica-se que a utilização de ontologias permite que sistemas computacionais possam apresentar

uma perspectiva ampliada do contexto dos dados, atendendo mais eficazmente as necessidades informacionais dos usuários.

Palavras-chave: *Web semântica. Ontologias. SPARQL.*

Semantic Web materialization: a dynamic building model of queries based in ontology mapping

Approach the computer language to human language is a challenge in seeking to improve the process of information retrieval. From this perspective, the Semantic Web has allowed a new mode of data processing contained in the Web, in order to allow better understanding of the meaning that these information possess. However, some of the concepts involving the Semantic Web such as ontologies, have highly complex, making it difficult to use. The difficulty in understanding, the structure that they have, and diversity as they are built, make the ontologies manipulation to discover the meaning of the concepts, something not effective, making it impossible to offer a concrete overview of the meaning of the concepts. Thus, this research aims to propose a model that can verify the context and meaning of a term in an ontology. To this end, it performed bibliographic research, model building and implementation of a prototype as proof of concept for the verification of results. The model and implementation have resulted in the generic treatment of an ontology, to obtain the meaning and the context that has a term. Based on the results it is identified that the use of ontologies allows information systems to provide a broader perspective of the context of data, enabling informational user needs can be met more effectively.

Keywords: *Semantic Web. Ontologies. SPARQL.*

Recebido em 03.05.2017 Aceito em 04.06.2018

1 Introdução

Computadores e humanos se expressam e se comunicam distintamente, e por mais que os computadores sejam capazes de acessar

e processar dados, produzindo informações relevantes aos usuários, estes necessitam de meios eficientes para que a comunicação com os humanos ocorra, na tentativa de tornar as necessidades informacionais dos usuários claras para as máquinas. Neste sentido, há diversos estudos que tem como objetivo uma aproximação entre a linguagem natural e a linguagem computacional, como pesquisas voltadas para a linguística aplicada a computadores, a inteligência artificial e os novos meios para a realização da recuperação da informação, que melhoram a comunicação existente, porém não resolvem a lacuna existente.

Neste âmbito, a Ciência da Informação necessita ter um papel mais ativo nestas questões, pois como relatado por Borko (1968), a Ciência da Informação deve se preocupar, entre outras questões, com o estudo do processamento e de técnicas aplicadas aos computadores e seus sistemas de programação, sem esquecer do componente prático da área, visto que esta disciplina desenvolve serviços e produtos.

Dentro da Web, as áreas citadas possuem desafios ainda maiores em torno da resolução de questões, envolvendo uma aproximação das linguagens naturais e artificiais, principalmente pelo crescimento exponencial das informações que são inseridas na rede, ainda mais pelo fato de que a maioria dessas informações estão desestruturadas, sem haver nenhuma classificação a respeito de seu conteúdo (DIAS; SANTOS, 2013).

Com o intuito de resolver tal problemática, e buscando tornar a Web um ambiente mais intuitivo, com dados estruturados, e com um contexto claro do significado das informações, Berners-Lee, Hendler e Lassila propuseram em 2001 a Web Semântica, tendo como foco permitir com que os computadores auxiliem mais eficientemente o usuário, deixando o significado dos dados mais claro tanto para os programas computacionais quanto para os próprios usuários (BERNERS-LEE; HENDLER; LASSILA, 2001).

Nessa perspectiva, surgem as ontologias como um meio de representar as informações dentro de um contexto, em que as relações que um conceito possui demonstra as suas características. Vale destacar que dentro dessa estrutura representacional, um conceito pode ser entendido como algo existente no mundo real. Nesta abordagem, as ontologias são capazes de representar um domínio, descrevendo as relações semânticas existentes, evitando ambiguidades.

No entanto, observa-se uma certa complexidade no uso de ontologias, pois estas apresentam um grande conjunto de mecanismos para a realização dos relacionamentos e para a definição dos domínios. Ressalta-se que estruturas informacionais como taxonomias, uma classificação hierárquica de entidades, e tesouros, um vocabulário controlado que apresenta relacionamentos básicos entre os conceitos, apresentam elementos e funções distintos das ontologias. No entanto, quando as ontologias são utilizadas como taxonomias ou tesouros, as principais funções que uma ontologia fornece não sejam utilizadas, como

as restrições, os axiomas para execução de inferências sobre os dados, e características das relações que são construídas (BREITMAN, 2005).

Outra questão relacionada ao uso de ontologias em programas computacionais, está no fato de que elas são construídas para um propósito específico, sendo necessário o seu conhecimento prévio para que um agente computacional possa se apropriar e usar o conhecimento nela explicitado. Assim, a construção de sistemas que utilizam mais profundamente tais tecnologias, em sua maioria, se baseiam em uma única, ou em um conjunto de ontologias pré-determinadas, limitando as possibilidades em torno do uso dessas ferramentas dentro de sistemas de Recuperação da Informação.

Dessa forma, esta pesquisa tem como objetivo criar um modelo para determinar o contexto e o significado de um conceito, verificando a sua localização dentro de uma ontologia e todas as suas relações, sendo a base desse modelo um motor de geração automática de consultas SPARQL. Para determinar o contexto e o significado do conceito, o modelo analisa e classifica as relações, obtendo um panorama amplo do sentido daquele conceito, possibilitando que inferências sejam realizadas.

A proposta desse trabalho apresenta-se como a base de projetos de Recuperação da Informação, que desejam utilizar as principais funções que uma ontologia oferece. Tal afirmação é justificada, pelo fato de que o primeiro passo para a inserção de semântica em um projeto de Recuperação da Informação, é a ocorrência de um processo de contextualização do conceito, que consiga extrair características relevantes deste conceito.

A metodologia desta pesquisa tem natureza qualitativa, com caráter descritivo e exploratório. A pesquisa foi dividida em três momentos: primeiramente foi feito um levantamento bibliográfico a respeito do protocolo SPARQL, Web Semântica e ontologias; posteriormente foi criado um modelo que trata das ontologias genericamente, podendo gerar consultas que consigam verificar as relações existentes; e por fim, foi desenvolvido um protótipo, como prova de conceito, que gera estas consultas, verificando se o resultado é satisfatório.

2 Web semântica

Criada em 1989, a Web tinha como princípio oferecer um ambiente com interfaces mais amigáveis e intuitivas para os usuários, além de oferecer espaço que possibilitasse o acesso, transação e recuperação das informações.

Contudo, a evolução das tecnologias utilizadas no contexto da Web, visavam principalmente a apresentação das informações para o usuário. As páginas estruturadas em *HyperText Markup Language* (HTML) apresentam um conteúdo com alta interação, porém estando desestruturadas e com significados incompreensíveis para os computadores e agentes computacionais.

Com o intuito de resolver tal questão, a Web Semântica foi proposta em um artigo de Berners-Lee, Hendler e Lassila, publicado em 2001. Nesse artigo, a proposta é que a Web Semântica seria utilizada como uma extensão da Web atual, onde as tecnologias auxiliariam as pessoas a realizarem tarefas básicas, pois os dados contidos na Web estariam relacionados, o que permitiria, entre outras coisas, com que os computadores entendessem o contexto e o significado das informações (BERNERS-LEE; HENDLER; LASSILA, 2001).

A partir desse artigo, foram surgindo diversas pesquisas expandindo a visão acerca da Web Semântica, e como ela se inter-relaciona com a Ciência da Informação e a Ciência da Computação. Como consequência, a Web Semântica vem em um processo de evolução, na busca de sua materialização (SANTAREM SEGUNDO, 2014).

Compreende-se a Web Semântica como uma estrutura que pretende embutir inteligência e contexto nas páginas Web, melhorando a interação dos programas com os conteúdos e, assim, possibilitar o uso mais intuitivo dos usuários. Essencialmente, a Web Semântica tem o propósito de criar e implantar padrões tecnológicos, para que seja possível que agentes, programas e dispositivos personalizados possam trocar informações, automatizando as tarefas cotidianas dos usuários (SOUZA; ALVARENGA, 2004).

Souza e Alvarenga (2004) apontam alguns pontos que a Web Semântica pode trabalhar em conjunto com a Ciência da Informação, auxiliando na melhora de diversos processos, não somente relacionado a Web. Desses pontos, destacam-se o projeto de novos e melhores motores de buscas, construção automática de tesouros e vocabulários controlados, indexação automática de documentos e gestão de conhecimento organizacional.

Diversas tecnologias foram surgindo com o intuito de aprimorar o processo de implementação dos conceitos propostos pela Web Semântica. Dentre elas, destacam-se *Resource Description Framework (RDF)*, *Uniform Resource Identifier (URI)*, *eXtensible Markup Language (XML)*, *Web Ontology Language (OWL)*, *SPARQL Protocol and RDF Query Language (SPARQL)* e diversos outras ferramentas que são descritos pelo W3C. Na sequência será descrito os principais padrões utilizados nesse projeto.

Uma tecnologia fundamental dentro da Web Semântica é o RDF, considerado por Catarino e Souza (2012) como o alicerce da Web Semântica. Este padrão tem a função de estabelecer os relacionamentos entre os dados, possibilitando a realização de diversos tipos de ligações e inferências, por meio de seu uso (CATARINO; SOUZA, 2012). O RDF apresenta a seguinte estrutura: existe um sujeito, um predicado e um objeto, sendo o sujeito um recurso, o predicado uma propriedade e o objeto, o valor desta propriedade.

Conteúdos apresentados apenas em HTML não permitem que o computador interprete o sentido destes, porém se esses dados forem disponibilizados também em RDF, um programa computacional conseguirá

verificar as relações existentes, deixando o conteúdo mais claro para a máquina.

Um outro conceito de grande relevância dentro da Web Semântica, diz respeito às ontologias e da sua linguagem de implementação computacional recomendada pela W3C, o OWL. A seguir será tratado especificamente sobre este conceito e sobre a linguagem OWL.

2.1 Ontologia

Ontologia é definido como uma especificação explícita e formal de uma conceitualização compartilhada (BORST, 1997). Partindo dessa definição, compreende-se uma ontologia como um meio de conceituar elementos do mundo, em uma estrutura que seja processável por máquinas, e aceito pela comunidade relacionada ao domínio tratado.

Na Ciência da Informação, as ontologias indicam modelos para representar pressupostos ontológicos e epistemológicos que tem relevância para o entendimento de pesquisas e de seu tratamento computacional (CAMPOS; CAMPOS, 2014). Podendo também, ser utilizadas com o objetivo de ser um vocabulário compartilhado, em que dados utilizando essa estrutura possam ser trocados e usados por outros usuários.

De acordo com Santarem Segundo e Coneglian (2015, p. 227): "Para o uso como tecnologia da Web Semântica, entende-se as ontologias como: artefatos computacionais que descrevem um domínio do conhecimento de forma estruturada, através de: classes, propriedades, relações, restrições, axiomas e instâncias."

A afirmação feita pelos autores contextualiza as ontologias na Web Semântica, destacando também, o papel das ontologias, quando se trata do entendimento das relações existentes entre os dados, levando em consideração o domínio a qual pertence as informações. Dessa forma, observa-se que as ontologias trabalham evitando a ambiguação, permitindo a realização de inferências e de diversos tipos de relações, entre outras funções.

Existem outros tipos de vocabulários, como tesouros e taxonomias, que conseguem representar conceitos e alguns tipos de relacionamentos entre tais conceitos, mas o uso de ontologias é capaz de representar um domínio com um nível de semântica formal mais elevado, frente aos outros tipos citados, devido às ontologias possibilitarem funções como, relacionamentos lógicos, inferências, restrições, que as outras opções não permitem. Tais funções são importantes, pois quanto mais possibilidades de explicitar os relacionamentos e as características dos conceitos, amplia-se a capacidade de representar um domínio com mais detalhes e de modo mais compreensível pelos agentes computacionais (BREITMAN, 2005).

Entretanto, o termo ontologia se refere a um conceito, assim, para existir uma ontologia, do ponto de vista prático e computacional, é necessário que seja utilizado alguma linguagem que implementa as

características que envolvem as ontologias. A linguagem OWL foi desenvolvida com essa finalidade, buscando representar as características e propriedades, descrevendo computacionalmente os conceitos pertencentes as ontologias.

Os elementos da linguagem OWL são essencialmente classes, propriedades e indivíduos. As classes são representações concretas de um conceito ou de uma entidade. Os indivíduos são os objetos de um domínio. E as propriedades são relações entre as classes e atributos destas classes (W3C, 2004).

Os três elementos são fundamentais na descrição de um domínio, porém, são as os diversos tipos de propriedades que distingue as ontologias de outros tipos de vocabulários, além de possibilitar inferências e descrever as características das classes e dos indivíduos.

Destaca-se uma função que as propriedades podem possuir, onde a semântica apresenta o seu maior nível de representatividade, pela inserção de características com alta capacidade representacional. Essas propriedades são chamadas de propriedades especiais, e algumas delas são enumeradas a seguir: "*owl:TransitiveProperty*", uma propriedade pode possuir características transitivas, ou seja, se A possui relação X com B, e se B possui relação X com C, logo A possui relação X com C; "*owl:SymmetricProperty*", uma propriedade simétrica indica que se há um relacionamento X de A para B, logo há também um relacionamento X de B para A; "*owl:FunctionalProperty*" define que uma propriedade desse tipo deve ter um único valor, sendo utilizado diversas vezes como um atalho para indicar que a cardinalidade mínima é 0 e a máxima é 1; e "*owl:InverseFunctionalProperty*" define uma propriedade que dois objetos não podem ter o mesmo valor, por exemplo um campo id de identificação, onde cada valor deve ser único (W3C, 2004).

O uso de uma ontologia construída com OWL possibilita uma representação bastante elaborada das relações entre as classes, os indivíduos e as propriedades existentes, permitindo uma estrutura informacional bastante complexa e clara, que trata as relações de uma forma satisfatoriamente consistente.

A evolução apresentada pelo OWL, e o amplo uso dessa linguagem por diversas comunidades, proporcionou com que, atualmente, o uso de ontologias, no contexto computacional, esteja fortemente vinculado ao OWL em si. Verifica-se essa questão quando os pesquisadores para estudar e entender a gama de possibilidade que uma ontologia pode oferecer, busca na literatura do OWL as respostas, inclusive porque tal linguagem apresenta característica que podem ser aplicados na prática, podendo assim, tornar real a implementação de ontologias.

Para a realização de buscas dentro de ontologias em OWL, ou em dados em RDF, foi criado uma linguagem chamada de SPARQL. A seguir será tratado com mais detalhes sobre essa linguagem de consultas.

2.2 SPARQL

O SPARQL se apresenta como recurso importante na aplicação das técnicas da Web Semântica. Este protocolo permite a realização de consultas a dados publicados em RDF e OWL, e possibilita realizar recuperação de informações semanticamente, pois trabalha levando em consideração as relações que existem entre os dados e os axiomas presentes nas ontologias.

Santarem Segundo (2014) diz que o uso do SPARQL possibilita a utilização dos dados do *Linked Data* e da Web Semântica como um todo. Isso ocorre, pois necessita-se de meios para acessar os dados armazenados em *data sets*, base de dados em formato compatível com as tecnologias da Web Semântica, sendo o SPARQL uma forma de tornar isso possível.

A linguagem SPARQL tem o intuito de recuperar dados, por meio de consultas à dados que estão nativamente armazenados em RDF ou que podem ser transformados em RDF utilizando algum serviço externo. Além disso, as consultas SPARQL possibilitam agregação, consulta dentro de uma outra consulta, negação, criação de novos valores, expressões, entre outros, e os resultados de consulta SPARQL são apresentados em conjuntos de dados ou em grafos RDF (W3C, 2013).

Outra característica importante do SPARQL, é que este dá a possibilidade de realizar inferências a respeito dos dados, isto porque, é possível realizar associações relacionando as triplas, além de, prover formas de verificar informações como a quantidade, soma e outros operadores, no momento que são realizadas as consultas (DUCHARME, 2013).

Basicamente, o SPARQL tem sua estrutura baseada no RDF, funcionando com relacionamentos em forma de triplas, com sujeito, predicado e objeto. Uma consulta SPARQL contém uma estrutura chamada de *select*, onde é definido como serão exibidos os dados de resposta, contém também uma cláusula *from* para indicar a origem dos dados selecionados, e uma estrutura chamada de *where*, que será responsável por indicar as lógicas de buscas que os dados deverão atender para serem selecionados.

A partir dos pressupostos teóricos discutidos, a seguir, apresenta-se a proposta de Modelo deste trabalho.

3 Modelo

A construção de meios para verificar as relações de um conceito, analisando ontologias, e assim entendendo o contexto e o significado na qual aquele conceito está inserido, permite que a Ciência da Informação e mais especificamente a área de Recuperação da Informação melhore seus processos, com o foco em atender mais eficientemente as necessidades informacionais do usuário.

As ontologias permitem com que o contexto e o significado de um conceito possam ser encontrados, por meio dos relacionamentos e das propriedades que este conceito possui. As ontologias construídas em OWL,

propiciem que estas características sejam processadas por computadores, e que programas consigam explorar estas funções. Além disso, um agente computacional (programa de computador) para explorar uma ontologia, em RDF ou OWL, e verificar o sentido dos conceitos, necessita do uso da linguagem SPARQL.

O modelo proposto nessa pesquisa tem a função de verificar o sentido e os relacionamentos dos conceitos dentro das ontologias, tendo como base um motor de geração de consultas SPARQL, que explora ontologias construídas em OWL e RDF.

Este modelo propõe que o usuário escolha uma ontologia qualquer, e que a partir disso, possam ser verificadas as relações semânticas que um conceito possui. Sendo possível identificar quais e de que tipo são as relações existentes, além de realizar inferências utilizando os axiomas presentes nas ontologias.

O desafio nesse modelo está em encontrar as relações independente da ontologia ou do conceito escolhido, pois o uso de consultas SPARQL em sistemas que foram construídos em cima de determinadas base de dados ou de determinadas ontologias não é algo tão complexo, devido a esquematização que pode ser realizada, quando se conhece os conceitos que se deseja buscar e a estrutura das ontologias. Porém, quando não se conhece a ontologia, é necessário que o sistema faça uma análise desta, e gere consultas SPARQL automaticamente, com o intuito de verificar dentro da ontologia as relações existentes.

Dessa forma, o funcionamento do modelo ocorre da seguinte maneira: uma ontologia é inserida dentro do sistema/ambiente, sendo feito o processamento e mapeamento computacional da mesma, para que sejam aplicadas consultas SPARQL, encontrando as relações existentes entre qualquer termo que está contida na ontologia.

Outro ponto fundamental que este modelo trata é sobre a utilização das funções que uma ontologia pode oferecer. É recorrente em grande parte do material publicado, o uso de ontologias tratando somente das relações de hierarquia existentes, ou seja, como se fosse uma taxonomia ou tesouro, onde verifica-se somente as classes de nível superior ou inferior ao conceito principal. Nesse modelo que apresentamos, mais importante do que as hierarquias são os tipos de relacionamentos que permitem que a semântica dos conceitos fique mais clara. As ontologias em OWL oferecem um conjunto de propriedades que possibilitam que o contexto de um objeto fique definido para o computador, sendo possível identificar conceitos relacionados.

O modelo proposto executa três tarefas essenciais: mapeamento da ontologia; localização do termo dentro da ontologia e; localização das relações entre os conceitos contidos na ontologia. A seguir, estão descritos os passos que descrevem tais atividades:

O primeiro passo dentro desse modelo é a realização do mapeamento de uma ontologia ainda desconhecida para o sistema. O usuário escolhe uma ontologia que pode estar na Web, ou que ele mesmo desenvolveu, e o sistema deverá realizar o mapeamento computacional,

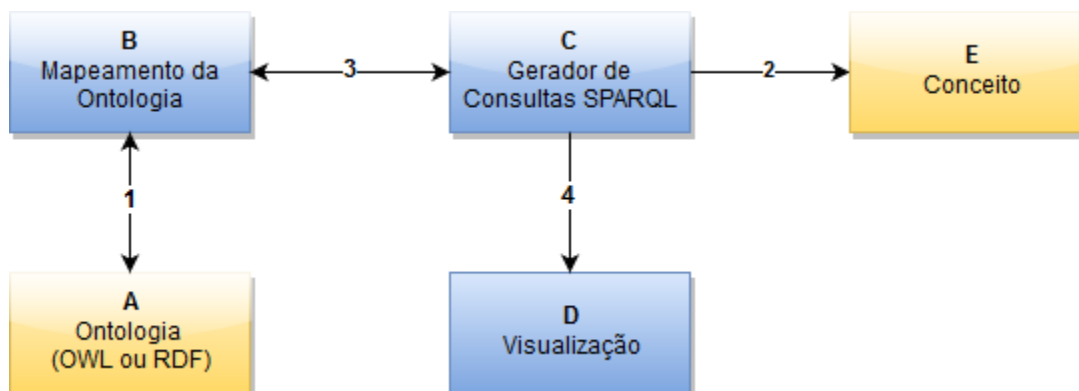
constituindo uma estrutura apta a realizar consultas por meio de agentes computacionais. Para tal, o usuário deverá submeter o link ou o arquivo para ser mapeado.

O segundo passo trata da localização do conceito dentro da ontologia. Este processo se dá por uma consulta SPARQL, gerada automaticamente, que encontrará a posição deste conceito dentro da ontologia. Nessa fase, é necessário o conhecimento exato da representação do termo na ontologia, ou seja, qual o nome deste conceito na ontologia. É fundamental encontrar a posição do conceito, para que seja possível a construção das próximas consultas SPARQL. O papel do usuário neste processo é escolher ou definir o conceito a ser localizado.

Posteriormente, sabendo a localização do conceito, é necessário que seja encontrado todas as relações existentes. Por meio de consultas SPARQL, que são geradas em tempo de execução, deverá ser verificado as ligações que um conceito escolhido possui. É preciso que seja gerada consulta após consulta, analisando a ontologia e verificando todas as relações.

O modelo para a realização da análise do sentido e relações que um conceito possui, pode ser visualizado na Figura 1. Nessa figura, as caixas circulares representam os elementos externos ao programa, que podem variar conforme as necessidades do usuário, e as caixas retangulares, que representam a parte de descobertas e apresentações que o modelo deverá realizar.

Figura 1 - Modelo de Análise das Relações de um Conceito



Fonte: Elaborado pelos autores.

Na Figura 1 o bloco A, demonstra que nesse modelo podem ser utilizados qualquer ontologia, no formato OWL ou RDF, e esta ontologia é desconhecida a princípio do gerador de consultas. No outro extremo do modelo, tem o bloco E, que representa o conceito, que assim como a ontologia, é desconhecida do gerador, podendo ser usado qualquer conceito que esteja contida dentro da ontologia escolhida. O bloco B, do mapeamento da ontologia, é a ontologia em OWL disponível de maneira que um programa construído em uma linguagem de programação consiga acessar, e que quando for realizada uma consulta SPARQL, seja possível extrair informações dessa ontologia. O bloco C, o gerador de consultas

SPARQL, é o centro do modelo, ele irá verificar as relações que um determinado termo possui, acessando ao mapeamento da ontologia, e transmitindo os resultados para o bloco D, de visualização. Na visualização, serão apresentadas as relações que existem entre o termo e os demais componentes da ontologia, organizado pelo tipo de relacionamento.

O modelo da Figura 1 possui setas numéricas que representam a sequência de passos que o modelo propõe fazer. A seguir estão descritos estes passos detalhadamente:

Primeiramente, no processo da seta número um, é feito a inserção de uma ontologia representada em OWL ou em RDF. Para utilizar as informações que estão inseridas dentro dessa ontologia, deverá ser feito o mapeamento desta em uma linguagem de programação, para que seja possível utilizar ferramentas para a extração dos dados da ontologia.

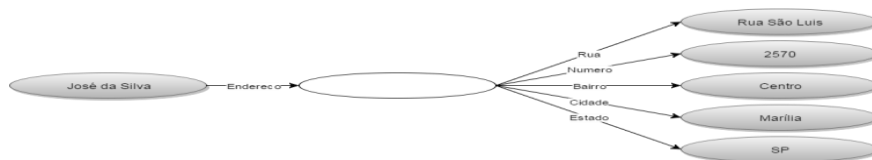
Posteriormente, na seta número dois, o gerador de consultas verifica qual é o termo que deverá ser encontrado na ontologia. Nesse passo será localizado onde o termo escolhido pelo usuário se encontra na ontologia. Para isto, o gerador de consultas, deverá utilizar o SPARQL, obtendo o retorno da existência deste na ontologia, e caso isto seja positivo, será retornado à representação deste termo, verificando o prefixo, e o nome exato (URI).

Estando com a localização do termo, inicia-se uma geração de consultas sucessivas à ontologia, processo informado na ilustração do modelo com a seta número três, que possui como propósito encontrar as relações existentes entre os componentes da ontologia com o termo. As triplas RDF da consulta, tem a seguinte estrutura: o sujeito sendo o elemento que corresponde ao termo principal, e o predicado e o objeto, sendo variáveis, que possibilitam a verificação das relações existentes entre os componentes da ontologia. O gerador para realizar estas verificações se relaciona com o mapeamento da ontologia, que será a fonte de dados da consulta realizada.

Os modelos RDF apresentam os chamados nós em branco, que devem ter um tratamento diferente na execução do procedimento descrito neste modelo. Nó em branco é uma estrutura de dado do RDF, onde uma classe se relaciona com diversas outras classes sem conter informação, como se fosse apenas uma derivação de informações, um exemplo é uma relação entre pessoa e endereço.

A tripla RDF do exemplo apresentado na Figura 2, seria composta por uma pessoa como sujeito, um relacionamento do tipo endereço, e o objeto/valor seria o dado que informa o endereço, porém, como o endereço possui diversos dados, como rua, número, bairro, cidade, estado, o RDF possibilita que o objeto seja um nó em branco, que se relaciona com diversos outros nós, portanto, o nó em branco, se relacionará com rua, número, bairro, cidade e estado.

Figura 2 - Exemplo de Grafo com Nó em Branco.



Fonte: Elaborado pelos autores.

Os nós em branco agregam dificuldades e complexidade extra ao processo de geração automática de consultas, pois é desconhecido quais e quantos são os nós em branco que ocorrem, além de que, um nó deste tipo pode apontar para diversos outros nós, sendo que algum desses também podem ser nós em branco, e assim sucessivamente, sendo necessário, que sejam geradas consultas SPARQL que vão percorrendo todos os nós, para extrair todos os relacionamentos que um determinado objeto possui.

Todos os nós devem ser percorridos pelo gerador de consulta no momento do mapeamento, sendo eles em branco ou com conteúdo. Com o gerador de consultas SPARQL tratando todas estas questões, é possível verificar as relações existentes com o termo principal, extraído da ontologia os objetos relacionados com aquele termo, além do tipo de relação existente. Ao final, é possível ter uma lista com todos os relacionamentos entre o termo principal e os outros objetos da ontologia.

Por fim, após obter a lista com as relações, é necessário apresentar ao usuário essas informações, para isto, é feito um processo de separação pelo tipo de relação, este processo é representado pela flecha número quatro. A exposição dos relacionamentos existentes, mostra o termo e o tipo de relacionamento ao usuário.

Para provar que o modelo cumpre o que é proposto, foi realizado a implementação de um protótipo que segue essa proposta. O próximo capítulo tratará sobre como foi realizado a implementação desse modelo.

4 Prova de conceito

Para realizar a implementação do protótipo foi executado os passos descritos no modelo, conforme a descrição da Figura 1, utilizando-se a ontologia do vinho¹, como exemplo.

O processo iniciou-se pela inserção e mapeamento da ontologia. Nesse procedimento, o protótipo é capaz de receber tanto ontologias que estão na Web, sendo colocado a URL da ontologia e feito o *download* do arquivo, quanto a inserção de um arquivo que está localmente na máquina do usuário. Após esta etapa, inicia-se o mapeamento da ontologia.

Seguindo o modelo, realizou-se a implementação para geração da consulta que localiza o conceito dentro do arquivo OWL ou RDF. Como padrão para encontrar o conceito, é utilizado uma consulta dinâmica, que

¹ Ontologia Vinho é abordada em muitos exemplos sobre Web Semântica, sendo proposta pelo W3C como um modelo que utiliza diversos elementos do OWL. Disponível em: <<http://www.w3.org/TR/owl-guide/wine.rdf>>. Acesso em: 26 maio 2018.

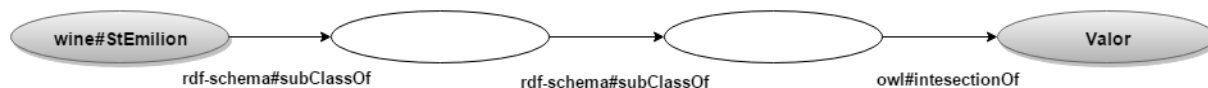
recupera o termo no objeto (no contexto da tripla do RDF: sujeito, predicado e objeto).

Após encontrar o objeto que corresponde ao que o usuário necessita, são localizadas todas as relações existentes do objeto com os outros conceitos da ontologia. Para isso, o gerador de consultas insere como sujeito da tripla RDF o objeto encontrado no passo anterior, e armazena em uma lista todos os predicados e objetos resultantes dessa busca.

Porém, a questão tratada na seção do modelo, da existência dos nós em branco dentro do RDF, traz grandes desafios no processo de geração automática de consultas. Por exemplo, na Figura 2, que demonstra o princípio do nó branco com um exemplo do endereço de uma pessoa, para uma recuperação de todos os elementos relacionados ao sujeito "José da Silva", devem ser recuperados todos os dados ligados ao nó branco. Portanto, se for realizado um algoritmo que gere consultas que considere apenas um nível de relações, existirá uma perda grande de informações. É necessário, assim, serem construídas consultas que vão avançando nível por nível destes nós em branco, até encontrarem todas as informações relacionados aos dados.

Na Figura 3 é ilustrado, em esquema de grafo, como uma consulta SPARQL vai avançando pelos nós em branco até localizar um nó com valor.

Figura 3 - Múltiplos nós brancos dentro de uma ontologia representada em grafo



Fonte: Elaborado pelos autores.

A implementação deste modelo, criou uma consulta para cada dado que está relacionado a algum nó branco. Ou seja, cada vez que aparecer algum nó branco ligado ao objeto principal, ou a algum outro nó, deve ser gerado uma nova consulta. Como não é conhecida a ontologia, isto deve ser tratado dinamicamente, realizando técnicas recursivas para que cada novo nó branco dê origem a nova consulta, até verificar todos os dados que estão relacionados na ontologia. Recursividade é um termo utilizado na Ciência da Computação para falar de uma rotina que pode chamar a ela mesma, executando as mesmas funções diversas vezes. No caso desta implementação, a recursividade é encerrada quando é encontrado algum nó com valor.

Após ser encontradas todas as relações existentes entre o objeto principal e a ontologia, as relações são inseridas em uma lista que armazenará basicamente, a relação e o tipo de relação.

A última etapa da implementação é a apresentação das relações existentes ao usuário. Para isto, primeiramente é realizado uma classificação de todos os resultados que foram encontrados, sendo realizado uma ordenação, agrupada pelas relações. Posteriormente, é

apresentado aos usuários todas as relações existentes. Desta forma, estando classificadas as relações, algum sistema pode fazer uso das relações determinadas, onde o usuário, humano ou não-humano, poderá escolher quais destas serão utilizadas

Com a criação e a implementação do modelo, verificou-se que o protótipo realiza todos os passos propostos, obtendo êxito no processo. Como prova de conceito, foram realizados testes com o protótipo desenvolvido, buscando validar o modelo construído

Após a seleção da ontologia do vinho, foi escolhido um termo para realização da busca dentro da ontologia. Realizou-se, assim, a execução do protótipo, utilizando o termo "StEmilion", sendo encontradas as relações descritas no Quadro 1.

Quadro 1 - Conceitos e relacionamentos encontrados na ontologia

Relação	Conceito
rdf-schema#subClassOf	food#ConsumableThing
owl#equivalentClass	wine#Bordeaux
owl#equivalentClass	food#PotableLiquid
owl#hasValue	wine#StEmilionRegion
food#MealCourse	owl#disjointWith

Fonte: Elaborado pelos autores.

No Quadro 1 é possível verificar as principais relações existente entre o conceito pesquisado e as outras classes da ontologia. Com os dados apresentados, o programa tem uma ideia clara das características e relacionamentos que o conceito pesquisado apresenta. Além do que, se esta informação for retornada a um usuário, que esteja inserido no contexto de um mecanismo de busca, este poderá refinar sua pesquisa, inserindo outros elementos que o programa mostrou a ele. Fundamentalmente, o programa apresenta o contexto em que o conceito se encontra, definindo todas as relações e características que tal termo possui.

Posteriormente foi realizado um segundo procedimento, utilizando uma outra ontologia. Este procedimento foi realizado, para verificar se a implementação corresponde a proposta do modelo de funcionamento em qualquer ontologia. A ontologia utilizada neste procedimento, foi a OpenTox (OPENTOX, 2015).

O termo escolhido para o procedimento foi o termo "NumericModel". Nesse segundo teste, o protótipo funcionou corretamente novamente, conseguindo verificar a localização e os relacionamentos existentes entre o termo e as outras classes da ontologia.

No quadro 2, são apresentadas algumas das ligações que foram encontradas.

Quadro 2 - Conceitos e relacionamentos encontrados na ontologia

Relação	Conceito
A generic OpenTox resource	rdf-schema#comment

1.1#Parameter	owl#disjointWith
1.1#Algorithm	owl#disjointWith
1.1#OpentoxResource	owl#equivalentClass
1.1#Model	rdf-schema#subClassOf

Fonte: Elaborado pelos autores.

No quadro apresentado, é possível verificar que assim como no primeiro caso, o segundo teste trouxe as relações existentes, e deixa claro para o mecanismo, quais são as ligações que o conceito possui, e assim, possibilita uma definição mais clara do significado daquele conceito.

5 Considerações finais

A criação do modelo e sua implementação permitiu que esta pesquisa aprofundasse no estudo da utilização de ontologias genéricas construídas em OWL. Basicamente, percebe-se que o uso concreto de ontologias, trabalhando com as diversas funções que o OWL oferece, expandiu o modo como um sistema consegue “compreender” o contexto e o significado que um termo possui.

Com os testes realizados foi possível verificar como esse modelo trata das especificidades dos relacionamentos, inserindo um alto nível de semântica, e tendo como consequência, uma aproximação da linguagem do humano com a da máquina. Tais aspectos foram verificadas, analisando o processamento realizado nesse protótipo, em que as características da ontologia foram absorvidas e assim, os algoritmos executados puderam ter uma maior clareza do domínio.

Outra contribuição que esta pesquisa traz, trata da possibilidade de uso de uma ontologia desconhecida do sistema, construída com as tecnologias recomendadas pela W3C. Isto é possível, pois o modelo foi construído para ser genérico, em que a utilização do SPARQL, gerando consultas e realizando associações e inferências, auxiliado por algoritmos computacionais, que expande a capacidade de atuação das consultas, torna possível a descoberta de informações em quaisquer ontologias. A implementação foi capaz de provar também, que o processamento da ontologia, pode se dar em tempo de execução. Com isso, o usuário ao utilizar este sistema consegue usar a ontologia que desejar, estando ela disponível na Web, ou construída pelo próprio usuário.

Portanto, o modo genérico, que permite o uso de qualquer ontologia, unido ao uso das principais características e propriedades do OWL, possibilita que a utilização de ontologias em ambientes de Recuperação da Informação se torne mais efetivo e claro. Destarte, verifica-se que o modelo proposto tem como principal função ser a base de projetos que desejam inserir o uso efetivo de ontologias de Recuperação da Informação. Destaca-se o termo uso efetivo, pois grande parte dos projetos construídos utilizando ontologias, não explora as principais funções que essa estrutura pode fornecer, e essa pesquisa

explora essa questão, ao conseguir identificar todas as relações que um termo possui semanticamente. Uma possível aplicação, seria trabalhar na construção de expressões de buscas, onde as relações encontradas seriam a base para composição da expressão, na tentativa de fornecer resultados mais precisos, que atendam com mais eficiência as necessidades informacionais dos usuários.

Referências

BERNERS-LEE, T; HENDLER, J.; LASSILA, O. The semantic web. *The Semantic Web*, v. 284, n. 5, p. 28-37, 2001.

BORKO, H. Information science: what is it? *American documentation*, v. 19, n. 1, p. 3-5, 1968.

BORST, W. N. *Construction of engineering ontologies for knowledge sharing and reuse*. Netherlands: University of Twente, 1997.

BREITMAN, K. K. *Web semântica: a internet do futuro*. Rio de Janeiro: LTC, 2005.

CAMPOS, L. M.; CAMPOS, M. L. A. Aplicação de dados interligados abertos apoiada por ontologia. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15.*, Belo Horizonte, 2014. *Anais... Além das nuvens, expandindo as fronteiras da ciência da informação*. Belo Horizonte, MG: ANCIB, 2014. p. 3822-3841. Disponível em: <<http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>>. Acesso em: 22 maio 2018.

CATARINO, M. E.; SOUZA, T. B. A representação descritiva no contexto da web semântica. *Transinformação*, Campinas, v. 24, n. 2, p. 77-90, 2012.

DIAS, T. D.; SANTOS, N. Web semântica: conceitos básicos e tecnologias associadas. *Cadernos do IME-Série Informática*, v. 14, p. 80-92, 2013.

DUCHARME, B. *Learning Sparql*. Sebastopol: O'Reilly Media, 2013.

OPENTOX. *OpenTox*. 2015. Disponível em: <<http://opentox.org>>. Acesso em: 22 maio 2018.

SANTAREM SEGUNDO, J. E. Web semântica: introdução a recuperação de dados usando Sparql. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15.*, Belo Horizonte, 2014. *Anais... Além das nuvens, expandindo as fronteiras da ciência da informação*. Belo Horizonte, MG: ANCIB, 2014. p. 3863-3882. Disponível em: <<http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>>. Acesso em: 22 maio 2018.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Tecnologias da web semântica aplicadas a organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados. *In: GUIMARÃES, J. A. C.; DODEBEI, V. (Org.). Organização do conhecimento e diversidade cultural*. Marília: Fundepe, 2015. v. 3, p. 224-233.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, v. 33, n. 1, p. 132-141, 2004.

W3C. *OWL Web Ontology Language*. 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em: 22 maio 2018.

W3C. *SPARQL 1.1 Query Language*. 2013. Disponível em: <<http://www.w3.org/TR/sparql11-query/>>. Acesso em: 22 maio 2018.