Aplicação de Tecnologias da Web Semântica em Motores de Busca na Internet

Vitor Rozsa*
Angel Freddy Godoy Viera*
Moisés Dutra*

Artículo recibido:
19 de junio de 2018
Artículo aceptado:
5 de noviembre de 2018

Artículo de revisión

Resumen

La Internet se caracteriza como un escenario dinámico que tiene un vasto volumen de informaciones, y representa un gran desafío para la recuperación de la información. La cuestión principal radica en comprender la intención de los usuarios al hacer sus búsquedas para posibilitar la recuperación semántica de la información. En este contexto, la Web Semántica (WS) proporciona tecnologías empleadas principalmente para la comprensión de las informaciones disponibles en la Web, y que pueden ser utilizadas en la contextualización de las búsquedas. Por tanto este trabajo se

* Programa de Pós-Gradação em Ciência da Informação (PGCIN), Universidade Federal de Santa Catarina (UFSC), Brasil vitor.rozsa@hotmail.com
a.godoy@ufsc.br
moises.dutra@ufsc.br

propone identificar cómo están siendo utilizadas las tecnologías de la WS para suministrar componentes semánticos en los buscadores de la Web, para lo cual se realiza una pesquisa bibliográfica. Se identificaron las aplicaciones de las tecnologías WS en diferentes componentes y etapas del proceso de búsqueda, así como en la recolección, el almacenamiento y la representación de la información; en los procesos de inferencia y en la recuperación en bases de datos; y en la elaboración de consultas y la comprensión de la necesidad de los usuarios en relación con sus búsquedas. También se identificaron diferentes buscadores semánticos en general volcados hacia propósitos específicos en lugar de búsquedas genéricas. Se verifica así que las tecnologías de la WS son empleadas por investigadores orientados hacia contextos específicos, y que aquellos investigadores volcados hacia propósitos generales utilizan abordajes semánticos pero que no están basados en la Web Semántica.

Palabras clave: Investigador Semántico; Tecnologías de la Web Semántica; Ontología; Triplestore (Tripleta)

Semantic Web Technologies Applied to Internet Search Engines

Vitor Rozsa, Angel Freddy Godoy Viera and Moisés Lima Dutra

Abstract

The Web is characterized as a dynamic scenario containing a vast amount of information, representing a great challenge for information retrieval. The main question in this scenario is related to the users' understanding for doing searches, in order to enable the semantic retrieval of information. In this sense, Semantic Web (SW) provides technologies mainly focused on understanding the information available on the Web that can be used Therefore, in this work the aim is to identify how SW technologies are being used to provide semantic components to Web search engines, for which purpose we conducted a bibliographical search. We identified application of SW technologies in different components and stages of the search process, such as in the collection, storage and representation of information; in the process of inference and retrieval in

databases, and in the elaboration of queries and the comprehension of the users' needs. We also identified different semantic search engines, usually aimed at specific purposes, rather than generic searches. It was thus verified that SW technologies are used for researchers oriented towards specific contexts, and that researchers with a general purpose use semantic approaches, but not based in the Semantic Web.

Keywords: Semantic Web Engine; Semantic Web Technologies; Ontology; Triplestore

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

Introdução

Devido ao grande volume de dados disponíveis na Web, a busca por informações por meio de motores de busca tornou-se uma tarefa pouco trivial. Um estudo da International Data Corporation (IDC) identificou que os dados gerados no mundo estão dobrando a cada dois anos, no entanto, apenas 22% das informações são marcadas com metadados que as tornem passíveis de análise por máquina (IDC, 2014). Assim sendo, a busca por palavra-chave, que utiliza principalmente o texto para a indexação, continua sendo a forma mais popular para encontrar conteúdo da Web.

Na busca baseada em palavra-chave, em geral, o buscador gera uma lista de palavras-chave (os índices) e a cada palavra-chave está relacionado um conjunto de documentos. Quando o termo buscado pelo usuário coincidir com alguma palavra-chave, os documentos relacionados serão retornados na lista de resultados. Mas, ou nem sempre os resultados refletem a intenção do usuário ou eles apresentam uma precisão muito baixa devido à grande quantidade de documentos irrelevantes recuperados (Singh e Singh, 2010). Este grande conjunto de resultados é uma tentativa de prover ao usuário soluções para todas as possíveis interpretações de sua consulta, porém, o usuário acessará apenas uma pequena parte destes resultados.

A busca por palavra-chave é útil para recuperar documentos já vistos, mas geralmente não é apropriada para encontrar documentos relevantes sobre um tópico ainda não visto (Nagpál, 2005). Por exemplo, este tipo de busca teria dificuldades em distinguir entre as consultas "futebol na escola" e "escola de futebol". O primeiro caso se refere à prática de futebol em escolas, enquanto o segundo caso se refere às instituições de treino da modalidade esportiva futebol. Em geral, um buscador tradicional não consegue resolver a ambiguidade

em uma consulta porque não conhece o contexto da pesquisa. Ele está focado nas palavras da consulta (sintaxe) e não em seu significado (semântica), o que afeta negativamente a precisão dos documentos retornados (Qu *et al*., 2011). Nesse contexto, torna-se cada vez mais necessário que os buscadores incorporem elementos semânticos em seus mecanismos com o intuito de melhor compreender as intenções dos usuários por trás de suas buscas.

A busca sensível à intenção do usuário e seu contexto é conhecida como busca semântica e os buscadores que fornecem este tipo de busca são denominados "motores de busca semântica". Os motores de busca semântica, ou simplesmente buscadores semânticos, têm por objetivo melhorar a precisão dos resultados da pesquisa por meio da compreensão dos termos e da intenção do usuário em relação à busca (Kalaivani e Duraiswamy, 2012). Eles também são capazes de fornecer informações sobre tópicos relacionados aos termos pesquisados, permitindo ao usuário expandir a abrangência de sua busca.

Os motores de busca semântica tomam forma principalmente por meio das tecnologias da Web Semântica (WS). Um dos principais objetivos da WS é estruturar o conteúdo relevante na Web para que possa ser entendido tanto pelas pessoas quanto por agentes de software, permitindo em especial o processamento automático do conteúdo por estes últimos. Dessa forma, por meio das tecnologias da WS os buscadores habilitam-se a compreender o contexto de busca dos usuários, bem como da coleção de documentos na Web, e assim recuperar resultados mais relevantes.

A WS não é uma proposta recente e tem sido desenvolvida principalmente sob os esforços do World Wide Web Consortium (W3C). Ao longo dos anos foram concebidas diferentes abordagens e tecnologias tendo em vista sua concretização, fornecendo cada vez mais subsídios para os buscadores semânticos.

Diante desse contexto, o presente trabalho busca identificar técnicas e métodos utilizados nos buscadores semânticos, especialmente aqueles baseados nas tecnologias da WS. Além disso, também é objetivo deste trabalho identificar quais são as abordagens utilizadas pelos buscadores semânticos disponíveis na atualidade. Para a identificação das técnicas e métodos, realizou-se uma pesquisa bibliográfica na base de dados Scopus, IEEE e ACM utilizando-se os termos "semantic search", "semantic search engine" e "semantic web". Para a identificação dos buscadores semânticos disponíveis atualmente, realizou-se uma busca por meio da ferramenta de busca Google utilizando-se o termo "semantic search engine". Esta é uma pesquisa exploratória, pois busca identificar e descrever os métodos e técnicas modernos relacionados aos buscadores semânticos. Também se configura como uma pesquisa qualitativa.

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

Nesta seção, introduzimos e justificamos o interesse na investigação sobre buscadores semânticos. Na seção "Web semântica", buscamos fornecer uma visão geral da WS e suas principais tecnologias. Na seção "Aplicação das tecnologias da web semântica nos buscadores web", descrevemos diferentes componentes utilizados em buscadores semânticos e baseados nas tecnologias da WS. Na seção "Buscadores semânticos atuais", apresentamos buscadores semânticos identificados na atualidade. E, por fim, realizamos as considerações finais deste trabalho.

Web semântica

O conceito de Web Semântica (WS) existe há mais de uma década e meia e progrediu consideravelmente ao longo desse tempo. A WS foi inicialmente proposta por Berners-Lee, Lassila e Hendler (2001) como uma extensão da Web atual e hoje é vista como a "Web de Dados". Nesta extensão, as informações disponíveis estão estruturadas e possuem significado bem definido, o que permite que humanos e máquinas estejam melhores aptos a cooperar para utilizar estas informações. Por meio da estruturação do conteúdo na Web, ou seja, da anotação semântica dos dados, os agentes de *software* podem obter maior benefício destes dados e tornam-se capazes de realizar tarefas mais complexas.

Corroborando com a definição anterior, na visão da W3C, a WS provê um conjunto de tecnologias que permite aos usuários criarem bases de dados na Web e ontologias para representar e processar as informações disponíveis (W3C, 2015). O resultado final é uma Web de Dados na qual os computadores estão habilitados a realizar tarefas significativas sobre o conteúdo disponível na Web.

Como esforço para realizar essa proposta, formaram-se grupos de trabalho por meio do W3C, principal organização internacional para a padronização da Web, que definiram padrões fundamentais e a estrutura básica da WS. Entre os padrões definidos estão o Resource Description Framework (RDF), Resource Description Framework Schema (RDFS) e a Web Ontology Language (OWL). Estes padrões são utilizados para a anotação de recursos na Web e possuem diferentes capacidade descritivas, sendo o OWL, entre estes, a linguagem que fornece maior expressividade.

Além dos padrões utilizados na representação dos recursos, também são relevantes as tecnologias utilizadas para armazenar, inferir e recuperar dados. Para estes propósitos são utilizados, respectivamente, *triplestores* (permitem o armazenamento e recuperação de informações representadas no

formato RDF), raciocinadores e linguagens de consulta, como o SPARQL Protocol And Query Language (SPARQL). Convém mencionar que outro padrão relevante empregado na identificação dos recursos na WS é o Uniform Resource Identifier (URI), mas que não será trabalhado nesta pesquisa.

Ontologias

As ontologias são utilizadas para o propósito de representação do conhecimento de um determinado domínio. Uma vez representado o conhecimento, pode-se realizar o processamento semântico de informações e compartilhamento de conhecimento entre entidades de *software*. A definição de ontologia tradicionalmente utilizada na CI e na Ciência da Computação (CC), originalmente proposta por Gruber (1993) e posteriormente expandida por Studer, Benjamins e Fensel (1998) com base em Borst (1997), se refere a ontologia como uma "especificação formal e explícita de uma conceitualização compartilhada" (Studer, Benjamins e Fensel, 1998: 25). Em ambas as áreas o termo ontologia denota uma ferramenta ou artefato utilizado para modelar conhecimento sobre um domínio, real ou imaginado (Gruber, 2009).

As ontologias tomam forma por meio dos padrões definidos para a Web Semântica e possuem diferentes níveis de expressividade. Enquanto o RDF permite a descrição de recursos através de triplas no formato "sujeito", "predicado" e "objeto", o RDFS trata-se de uma expansão semântica do RDF e fornece mecanismos para a descrição de grupos de recursos e o relacionamento entre estes recursos. Já o OWL adiciona vocabulário para descrever propriedades e classes, relações entre classes, cardinalidade, equivalência, características de propriedades e classes enumeradas.

A linguagem OWL possui ainda algumas sublinguagens, como o OWL Lite, OWL DL e OWL Full. Além disso, uma versão mais atual é o OWL 2 (W3C, 2012), que fornece as sublinguagens OWL 2 EL, OWL 2 QL e OWL 2 RL. Estas variações estão representadas e descritas no *Quadro 1*.

Cada variação da linguagem OWL possui diferentes níveis de expressividade e, consequentemente, diferentes níveis de complexidade. A expressividade da linguagem está principalmente relacionada ao que se pode representar na ontologia (tipos de relações, propriedades e regras de inferência), enquanto a complexidade impacta na manutenção da base de dados e durante o processo de inferência sobre as informações anotadas. Entretanto cada variação adapta-se melhor a determinada situação. Por exemplo, em alguns casos, como em uma base médica, teremos uma grande quantidade de entidades, relações e propriedades, exigindo maior capacidade em lidar melhor com ontologias muito grandes (ex.: OWL 2 EL).

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

Variação	Descrição
OWL Lite	Voltada para criação de uma hierarquia de classificação com restrições simples.
OWL 2 EL	Adequada a aplicações com grandes ontologias nas quais a expressividade é trocada por performance computacional.
OWL 2 QL	Possui facilidades de acesso e consulta às bases de dados especialmente por meio de consultas relacionais (e.x.: SQL). Adequada a situações nas quais a ontologia modela poucos conceitos e relações, mas possui grandes quantidades de indivíduos.
OWL 2 RL	Possui facilidades de acesso e consulta às bases de dados para operar diretamente sobre triplas RDF. Adequada a situações nas quais a ontologia modela poucos conceitos e relações, mas possui grandes quantidades de indivíduos.
OWL DL	Provê alto nível de expressividade mantendo a completude (todas as conclusões/inferências serão computadas) e decidibilidade (todas as computações finalizarão em um tempo finito) da ontologia.
OWL Full	Provê máximo nível de expressividade sem as garantias computacionais (completude e decidibilidade).

Quadro 1. Variações da família de linguagens OWL
Fonte: Elaborado pelos autores com base em W3C (2004, 2012)

As ontologias permitem que sejam realizados processos de inferência sobre as informações contidas em uma base de dados. Por meio da descrição formal de conceitos, suas propriedades e relações, possibilita-se a realização de processos de inferência para descobrir novos conhecimentos a partir das informações previamente armazenadas na base. Alguns tipos comuns de inferência por meio de ontologias são herança, transitividade, simetria e equivalência. Estes tipos de inferência estão descritos de forma sucinta no *Quadro 2*. Para o uso de regras personalizadas e mais complexas é possível utilizar a linguagem Semantic Web Rule Language (SWRL, https://www.w3.org/Submission/SWRL/) que, apesar de submetida em 2004 e comumente utilizada em conjunto com ontologias OWL, ainda não se tornou um padrão da W3C.

Tipo de inferência	Descrição
Herança	Permite a criação de hierarquia de conceitos. Um subconceito também é considerado do mesmo tipo de um conceito mais geral. Por exemplo, Pessoa pode ser um subconceito de Mamífero. Assim, um raciocinador entende que Pessoa também é um tipo de mamífero.

Transitividade	Permite a criação de relações transitivas. Quando um indivíduo A está relacionado a um indivíduo B por meio de uma relação transitiva; e o indivíduo B está relacionado a um indivíduo C por meio desta relação, então infere-se que A está relacionado a C por meio desta relação. Por exemplo, a relação "ancestralDe" pode ser definida como transitiva.
Simetria	Permite a criação de relações simétricas. Quando um indivíduo A está relacionado a um indivíduo B por meio de uma relação simétrica, infere-se que o indivíduo B apresenta a mesma relação. Por exemplo, a relação "amigoDe" pode ser definida como simétrica.
Equivalência	Permite a criação de classes sinônimos. Assim, a instância de uma classe também será considerada instância da classe sinônimo. Por exemplo, o conceito Carro pode ser definido como equivalente de Automóvel.

Quadro 2. Exemplos de inferência fornecidos pelas ontologias OWL
Fonte: Elaborado pelos autores com base em W3C (2004)

O SWRL pode ser entendido como uma extensão do OWL que permite a definição de axiomas do tipo Horn. Em SWRL, estes axiomas são compostos por antecedente e consequente, ou seja, quando determinada expressão lógica for verdadeira (antecedente), infere-se que o consequente também seja verdadeiro (consequente). Assim sendo, o SWRL permite a definição de axiomas específicos que potencializam o uso das ontologias.

Triplestores

As *triplestores*, ou RDF *stores*, são bases de dados em forma de grafo utilizadas no armazenamento e recuperação de triplas RDF. Em comparação com outras bases de dados baseadas em grafo, uma característica de destaque das *triplestores* é a capacidade de utilizar ontologias na modelagem dos dados (Ontotext, 2017), obtendo-se assim as vantagens da representação semântica dos dados e a possibilidade de inferência sobre estes dados.

Convém mencionar que a serialização e recuperação de dados utilizando armazenamento em formatos de texto como, por exemplo, o XML (W3C, 2008) ou Turtle (W3C, 2011) atendem a soluções que envolvem um volume reduzido de dados, uma vez que o acesso ao disco rígido representa um alto impacto (ou custo) computacional no processo de busca (Fayer, Curé e Blin, 2012). Alternativamente, *triplestores* baseadas, por exemplo, no armazenamento em memória RAM ou bancos de dados relacionais são capazes de suportar uma quantidade maior de dados semânticos de forma eficiente.

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

Uma vez que as informações nas *triplestores* estão organizadas em forma de grafo, as linguagens tradicionais de consulta a base de dados relacionais, como o Structured Query Language (SQL), não são a forma mais adequada para recuperar informações nas triplestores. É necessária uma linguagem capaz de expressar consultas que representam relações no grafo RDF (Pellegrini, 2006). Assim sendo, a recuperação de dados na base é realizada utilizando-se SPARQL.

Em uma *triplestore*, além de relações entre entidades, também é possível encontrar relações entre entidades e documentos não-estruturados (como textos) dos quais estas entidades foram extraídas (Ontotext, 2017). Sendo assim, a população de uma base de dados pode ser gerada por meio de mineração de texto combinada com tecnologias semânticas e aprendizagem de máquina, utilizadas para identificar e classificar possíveis entidades em uma coleção de dados desestruturados (Pellegrini, 2006). Esta abordagem para povoar a base de dados é particularmente interessante pois demonstra como buscadores semânticos podem minerar a Web e montar suas bases para consulta.

Alguns exemplos de *triplestores* são: AllegroGraph, Stardog, GraphDB, RDFox, Sesame e o Virtuoso Universal Server. Estas bases, consideradas de alta capacidade, comportam de 70 milhões até 1 trilhão de triplas (W3C, 2016). Uma ampla lista de implementações de triplestores abertos e comerciais pode ser encontrada em Wikipedia (2017).

Linguagem de consulta

Para acessar os dados armazenados em triplas RDF é necessária uma linguagem de consulta. O W3C mantém o padrão SPARQL como ferramenta para recuperar, atualizar ou remover triplas RDF, sendo sua versão atual o SPARQL 1.1 (https://www.w3.org/TR/rdf-sparql-query/). O SPARQL pode ser utilizado por meio de motores de consulta como o ARQ.[1]

Uma tripla RDF é composta por sujeito, predicado e objeto. Enquanto o sujeito e o objeto são recursos, o predicado é um elemento de ligação (geralmente declarado em uma ontologia) que relaciona estes recursos. Um recurso pode ser tanto sujeito quanto objeto e possuir diversas ligações, compondo assim o grafo RDF de um conjunto de dados. Na *Figura 1* está representado um grafo RDF.

1 O ARQ é uma ferramenta de consulta para o Jena. Mais informações em: https://jena.apache.org/documentation/query/index.html

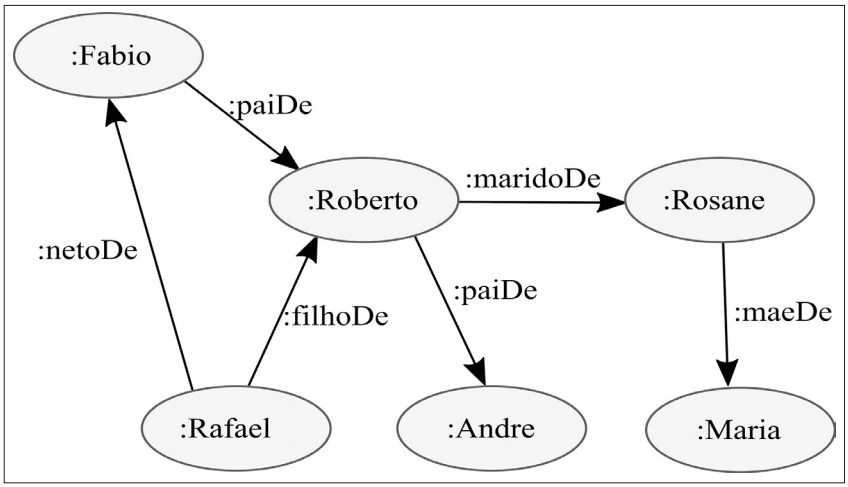

Figura 1. Exemplo de grafo RDF
Fonte: Elaborado pelos autores

O grafo representado na *Figura 1* possui seis triplas e seis recursos que, nesse caso, indicam pessoas. Os recursos estão ligados por relações de parentesco ":paiDe", ":filhoDe", ":maeDe" e ":maridoDe". A direção das setas indica os sujeitos e objetos nas triplas. Para dada relação, o recurso na base da seta é o sujeito e o recurso apontado é o objeto. Convém notar que esta é apenas uma representação visual de um grafo RDF, sendo os formatos mais comuns na serialização de triplas o RDF/XML e o RDF/Turtle, como mencionado anteriormente.

Em uma consulta SPARQL buscamos representar um subconjunto do grafo que pretendemos recuperar. Podemos pensar nessa consulta como um filtro que retorna apenas triplas que atenderem aos requisitos definidos. Para o grafo da *Figura 1*, se buscássemos identificar todos os netos de Fabio, poderíamos gerar uma consulta SPARQL empregando a relação ":netoDe", utilizando ":Fabio" como objeto da relação e o sujeito da relação (os netos de Fabio) como as variáveis a serem identificadas. Esta consulta está ilustrada no *Quadro 3*.

```
PREFIX :    <http://example.org/>
PREFIX rel:  <http://example.org/relacoesDeFamilia>

SELECT ?netos
WHERE {
 ?netos rel:netosDe :Fabio
}
```

Quadro 3. Exemplo de consulta SPARQL
Fonte: Elaborado pelos autores

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

O resultado obtido a partir da consulta no *Quadro 3* submetida no grafo representado na *Figura 1* seria apenas ":Rafael". Visualmente, também podemos identificar, a partir da Figura 1, que Andre e Maria também são netos de Fabio. Entretanto, não há relação direta entre esses recursos como há entre Fabio e Rafael e, sendo assim, ":Andre" e ":Maria" não podem ser recuperados. Esta inferência que realizamos para identificar Andre e Maria como netos de Fabio é efetuada pelos raciocinadores lógicos, explicado em maiores detalhes a seguir.

Raciocinadores

Os raciocinadores são ferramentas capazes de realizar inferências lógicas a partir das ontologias e dos conjuntos de dados nas *triplestores*. Eles podem ser utilizados para validar a consistência da ontologia e das informações na base, ou seja, verificar se não há informações conflitantes; e também para explicitar conhecimentos implícitos a partir das informações armazenadas.

No exemplo da seção anterior, no qual busca-se identificar os netos de Fabio, apenas os recursos que estão ligados a ":Fabio" por meio da relação ":netoDe" são retornados. Por meio de axiomas e inferência também podemos recuperar os outros dois netos de Fabio. Poderíamos criar o seguinte axioma "todos os filhos do filho de alguém são netos desse alguém". Como em nosso exemplo ":Maria" não está declarada diretamente como filha de ":Roberto", o seguinte axioma também seria necessário "todos os filhos(as) de uma esposa também são filhos(as) de um marido". Assim, conseguimos identificar todos os netos de Fabio.

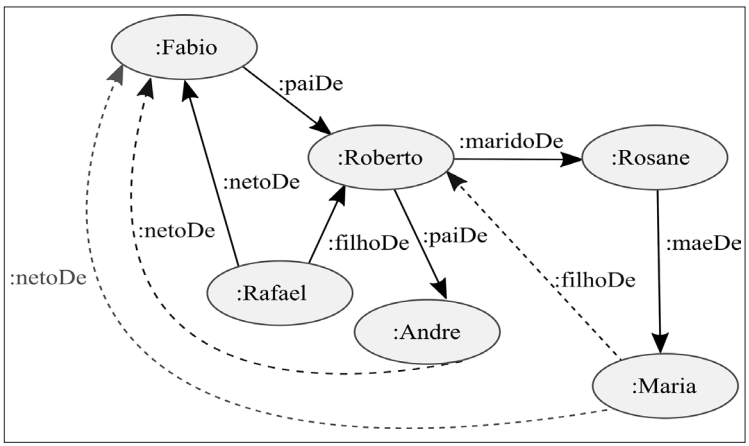

Figura 2. Exemplo de grafo RDF contendo relações inferidas
Fonte: Elaborado pelos autores

Apesar da simplicidade do exemplo anterior, o uso de axiomas e inferência pode revelar informações menos evidentes e de maior complexidade em uma base de dados. Na *Figura 2*, as setas pontilhadas indicam as relações inferidas no grafo RDF.

Existem diferentes raciocinadores semânticos com diferentes capacidades. Por exemplo, FactC++, Pellet, Hermit e ELK são raciocinadores comumente mencionados na literatura. Entre os principais fatores que motivam a escolha de um ou outro raciocinador está o tipo de linguagem de ontologia que é capaz de tratar (ex.: OWL Lite, OLW DL, OWL DL, etc) e sua performance (capacidade de tratar grandes quantidades de triplas). No *Quadro 4* estão listados exemplos de raciocinadores e a linguagem por eles suportada.

Raciocinador	Linguagens suportadas
FactC++	OWL DL e OWL 2
Pellet	OWL DL e OWL 2 EL
Hermit	OWL 2
ELK	OWL 2 EL

Quadro 4. Exemplos de raciocinadores semânticos e linguagens OWL suportadas
Fonte: Elaborado pelos autores com base em W3C (2011)

Aplicação das tecnologias da web semântica nos buscadores web

Os elementos que compõem a arquitetura de um buscador semântico são similares àqueles de um buscador baseado em palavras-chave (ex.: *crawler*, indexador e motor de busca). Entretanto a diferença está no emprego das tecnologias da WS nos diferentes elementos e etapas do processo de busca tendo em vista a recuperação semântica de resultados. Nesta seção, buscamos identificar as oportunidades de emprego das tecnologias WS que viabilizam a busca semântica por meio de exemplos identificados na literatura.

As tecnologias da WS podem ser utilizadas em diversos processos de uma ferramenta de busca como, por exemplo, na coleta, representação e armazenamento da informação; no processo de inferência sobre a informação e consulta (recuperação) na base de dados semântica; e na elaboração de uma consulta pelo usuário e na compreensão da intenção do usuário com esta consulta. No *Quadro 5* estão descritas de forma geral algumas das aplicações da WS que foram identificadas na literatura em cada um dos processos mencionados anteriormente de um buscador.

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

Processo	Aplicação
Coleta	Uso de ontologias na compreensão do significado do texto para classificação e agrupamento.
Representação	Representação semântica da informação por meio de ontologias.
Armazenamento	Armazenamento de grandes quantidades de dados em *triplestores*.
Inferência	Emprego de motores de inferência para o processamento de regras, derivação de novos conhecimentos e verificação de consistência da base de dados.
Consulta a Base de Dados Semânticos	Uso de ferramentas de consulta (SPARQL) para a recuperação semântica.
Auxilio na Elaboração da Consulta	Uso de taxonomias e *tags* semânticas para auxiliar o usuário na elaboração da consulta.
Compreensão da Intenção do Usuário	Uso de ontologias para compreender a intenção do usuário e selecionar o modelo adequado a uma consulta, podendo também expandi-la.

Quadro 5. Aplicações de tecnologias da WS em motores de busca
Fonte: Elaborado pelos autores

Nas subseções seguintes, cada processo apresentado no *Quadro 5* será explicado em maiores detalhes. Na próxima seção, buscamos identificar o emprego de tecnologias da WS tendo em vista estes processos nos buscadores semânticos disponíveis na atualidade.

Coleta

Na coleta, as tecnologias da WS podem ser utilizadas na compreensão do conteúdo da coleção. Laura e Me (2015), buscando identificar conteúdos ilegais na Web, utilizaram a combinação de redes semânticas e bases do conhecimento para compreender de forma precisa o conteúdo de documentos. Esta habilidade permite uma melhor classificação do documento que será refletida em uma maior precisão durante o processo de recuperação no sistema.

Outra aplicação das ontologias na etapa de coleta é o agrupamento de documentos com conteúdo similares. Soliman, El-Sayed e Hassan (2015) propuseram uma solução baseada em ontologias para o agrupamento de resultados recuperados em uma busca baseada em palavras-chave. O agrupamento destes resultados reúne documentos similares e facilita o trabalho do usuário durante a análise destes resultados. As ontologias são utilizadas na construção de redes semânticas que representam cada documento na lista de resultados. Esta rede semântica é pré-processada e então utilizada para

computar a similaridade entre os diferentes documentos de modo que documentos semelhantes possam ser agrupados.

Representação

A representação semântica nos buscadores se refere ao uso de ontologias para descrever domínios do conhecimento. As ontologias permitem que as informações sejam modeladas por meio de conceitos, propriedades, relações e axiomas, possibilitando assim que raciocinadores realizem inferências sobre as informações na base de dados (Ma *et al*., 2012). A representação do conhecimento de um domínio é o primeiro passo para que os buscadores Web tenham acesso aos demais benefícios das ontologias.

A representação do conhecimento de um domínio apresenta benefícios por si só. As ontologias podem ser utilizadas, por exemplo, no processo de coleta de informações, na elaboração da consulta pelo usuário e na compreensão da necessidade de informações do usuário. Durante a coleta, como visto anteriormente, as ontologias podem ser utilizadas na classificação e agrupamento de documentos. Durante a especificação de uma consulta, podem ser obtidos termos a partir da ontologia para auxiliar o usuário a formular a consulta ou expandi-la. Por fim, assim como as ontologias são utilizadas para compreender o texto em documentos, elas também podem ser utilizadas em técnicas de processamento de linguagem natural para compreender a consulta do usuário. Estas técnicas são exemplificadas nas seções seguintes.

Armazenamento

Uma vez coletados os dados, eles podem ser armazenados em *triplestores*. Sayed e Muqrishi (2017) utilizaram o Jena TDB (https://jena.apache.org/documentation/tdb/) para armazenar dados acadêmicos na ferramenta que denominaram de "buscador baseado em ontologia". Na ferramenta mencionada, os autores destacam a capacidade das *triplestores* em realizar inferência sobre os dados, viabilizando assim a descoberta de relações entre diferentes informações. Por exemplo, uma busca por "doenças cardíacas" poderia retornar páginas com doenças específicas no coração sem conter os termos originais da busca.

Buscando beneficiar-se das capacidades de armazenamento, consulta e recuperação semântica, Arenas, Haberlot e Cruz (2014) propuseram e testaram um buscador semântico para informações geoespaciais. Em sua proposta, registros de metadados são armazenados em triplestores e associados como instâncias de conceitos em uma ontologia. As consultas são realizadas

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

utilizando uma linguagem específica que é traduzida para SPARQL/GeoSPARQL,[2] antes de ser submetida à base de dados. A partir desta configuração, os autores destacam que se torna possível buscar instâncias a partir de conceitos; recuperar conceitos específicos a partir de conceitos genéricos; e realizar relações entre informações geoespaciais.

Inferência

A derivação de novos conhecimentos por meio do processo de inferência é um dos benefícios da representação de informações por meio de ontologias. Fernández *et al.* (2016) utilizaram ontologias para gerar uma base de metadados semânticos a partir de metadados e registros médicos do repositório Encyclopedia of DNA elements (ENCODE). Com o intuito de complementar as informações, os autores propuseram a aplicação de técnicas de inferência sobre as informações na base semântica.

Uma das principais motivações dos autores (Fernández *et al.*, 2016) é que, apesar de fornecer informações de alta qualidade, o suporte para busca e recuperação de informações no repositório ENCODE - baseado estritamente na sintaxe dos termos de busca - é insuficiente. Nesse contexto, os autores buscaram melhorar a recuperação nesta base por meio do uso de ontologias e da anotação semântica dos metadados da coleção no repositório ENCONDE.

Um potencializador do processo de inferência é o uso de SWRL. Enquanto o OWL permite a definição de restrições básicas, o SWRL permite uma maior flexibilidade por meio da criação de regras mais complexas. Por exemplo, Ma *et al.* (2012) utilizam o SWRL para combinar diferentes propriedades em uma mesma regra. Neste trabalho, quando uma consulta de usuário assume uma forma similar a uma regra com restrições, é possível que o sistema identifique e carregue automaticamente a regra correspondente em SWRL e a aplique no processo de inferência.

Consulta a base de dados semânticos

Em um buscador semântico, a recuperação de dados nas *triplestores* é realizada por meio da linguagem SPARQL. Sendo assim, os termos de busca definidos pelo usuário precisam ser traduzidos em uma consulta correspondente em SPARQL para então serem aplicados na base de dados. Por outro lado, é comum que bases de dados RDF disponibilizem um SPARQL Endpoint, que se trata de um recurso que pode ser utilizado por pessoas ou algoritmos

2 GeoSPARQL é uma linguagem que permite a representação e consulta de dados geoespaciais na WS.

automatizados para a submissão de consultas na base utilizando-se diretamente a linguagem SPARQL. Um exemplo de SPARQL Endpoint popular é o utilizado para consultas no DBpedia, disponível por meio de navegadores Web no endereço: https://dbpedia.org/sparql.

Um exemplo do uso do SPARQL é o trabalho de Ayaz *et al.* (2016), no qual é proposto um motor de busca semântico para romances da cultura Urdu. Neste motor de busca, as informações são representadas por meio de ontologias e consultadas utilizando-se SPARQL. A ferramenta permite ainda a inclusão ou atualização de informações na base, mas utilizando a OWL API.[3]

Auxílio na elaboração da consulta

Uma questão recorrente em um processo de busca é a possibilidade do usuário não lograr traduzir exatamente sua necessidade de informação em uma expressão de busca. Nesse caso, o buscador, por sua vez, pode não ser capaz de retornar resultados relevantes para o usuário, exigindo que este reformule a busca de forma a melhor representar sua necessidade. Assim, são necessárias funcionalidades que auxiliem este usuário na definição de termos de busca como, por exemplo, uma funcionalidade de autocompletar que busca termos relacionados (ex.: termos mais gerais, específicos ou sinônimos) a um termo inicial; ou disponibilizar uma lista de termos gerais para serem pesquisados correspondendo aos conceitos mais genéricos de uma ontologia.

Fatima, Luca e Wilson (2014) propuseram um *framework* visando a criação de motores de busca semântica completos. Neste *framework*, os autores sugerem dois componentes semânticos complementares que são o otimizador de consultas e o processador de ontologias, que podem ser utilizados para auxiliar na formulação da consulta pelo usuário e na compreensão da mesma pelo motor de busca. Nesse caso, o otimizador de consultas busca entender os termos inseridos pelo usuário e identificar ontologias relacionadas. Este procedimento permite que o buscador defina consultas específicas a serem processadas pelo motor de busca. Já o papel do processador de ontologias é auxiliar o otimizador de consultas a identificar ontologias relacionadas aos termos de busca e identificar e obter diferentes ontologias na Internet.

Compreensão da intenção do usuário

A compreensão da intenção do usuário em uma busca pode ser considerada como um dos principais desafios dos buscadores. Este quesito está diretamente

3 A OWL API é uma ferramenta em Java utilizada na manipulação de ontologias.

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

relacionado ao fosso ou lacuna semântica (do inglês, *semantic gap*), que se refere à diferença entre o que o usuário compreende de sua necessidade de informação e o que ele consegue representar em sua consulta. Alguns benefícios da compreensão da intenção do usuário em uma consulta são, por exemplo, selecionar ou gerar uma melhor representação de sua necessidade do ponto de vista do motor de busca (ex.: traduzir os termos de busca em uma consulta SPARQL); e selecionar os recursos (ex.: o modelo de consulta ou a base de dados) mais adequados para recuperar informações.

Por fim, convém mencionar que podemos encontrar buscadores que são completamente baseados em tecnologias da WS ou buscadores que empregam estas tecnologias de forma pontual. Isto é justificável, dado que a quantidade de documentos no formato RDF na Web ainda é muito pequena (Fatima, Luca e Wilson, 2014), exigindo que os buscadores sejam capazes de tratar tanto de documentos RDF quanto não-RDF.

Um exemplo de um buscador semântico híbrido é o IBRI-CASONTO (Sayed e Muqrishi, 2017), capaz de processar tanto buscas semânticas quanto buscas baseadas em palavras-chave. Para que isso seja possível, é necessário que o buscador possua alguns elementos de infraestrutura separada, como uma base de dados relacional e uma *triplestore* para o armazenamento de dados, bem como um indexador específico para cada base. Diante disso, podemos perceber que um buscador habilitado a tratar tanto de documentos RDF quanto não-RDF, consequentemente, será mais complexo.

Buscadores semânticos atuais

Podemos identificar dois tipos de buscadores, aqueles que realizam buscas a partir de determinado conteúdo e, portanto, conhecem explicitamente o contexto de busca, e aqueles voltados para buscas gerais (generalistas), que precisam inferir o contexto de busca. Os primeiros são criados tendo em vista aplicações específicas, como buscar nas bases de dados de uma organização e, assim sendo, estão geralmente limitados aos contextos para os quais foram criados. Já os buscadores generalistas, por meio de diferentes tecnologias, devem ser capazes de compreender a intenção do usuário e o contexto do conteúdo indexado para retornar resultados adequado às consultas.

Atualmente, encontramos poucos exemplos de buscadores semânticos generalistas. Na primeira década dos anos 2000 surgiram diferentes propostas destes tipos de buscadores, mas que ao longo do tempo tornaram-se restritas (apenas soluções corporativas), foram encerradas, abandonadas ou, no melhor dos casos, compradas por empresas maiores e incorporadas em suas próprias soluções (ex.: Hakia, Yebol, Lexxe, Factbites, Powerset, entre outros).

Os buscadores semânticos generalistas que persistem até os dias atuais, como o SenseBot, Kngine, Google e DuckDuckGO, acabaram incorporando componentes semânticos para evoluir suas soluções. Dessa forma, estes buscadores passaram a compreender cada vez mais o contexto dos usuários e tornaram-se mais cômodos ao fornecer informações de acordo com este contexto.

SenseBot

Um buscador semântico disponível atualmente é o SenseBot (Semantic Engines LLC, 2017). Este buscador encaixa-se no perfil de buscador generalista, porém, é especializado em construir sumários a partir do conteúdo de diferentes documentos para determinada consulta. O sumário construído identifica a origem de cada extrato de texto, permitindo assim que o usuário possa aprofundar-se no tema, além de fornecer pistas sobre a confiabilidade do conteúdo. Na *Figura 3* é apresentada uma parte do sumário criado para a consulta "Universidade Federal de Santa Catarina".

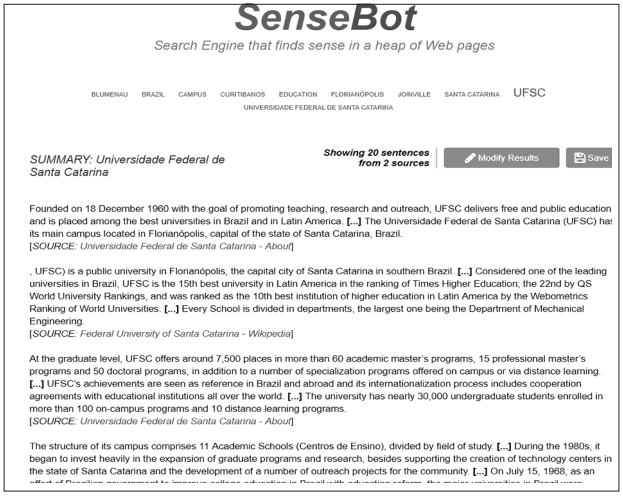

Figura 3. Sumário gerado para a consulta "universidade federal de santa catarina"
Fonte: Captura de tela gerada e adaptada pelos autores

Para gerar este sumário, o SenseBot utiliza mineração de texto e processamento de linguagem natural para identificar conceitos semânticos chave nos documentos. Também são atribuídos pesos aos conceitos, o que permite identificar quais documentos estão mais ou menos relacionados com a consulta do usuário. Além disso, podemos notar na *Figura 3* que diferentes *tags*

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

semânticas (ex.: CAMPUS, EDUCAÇÃO, FLORIANÓPOLIS, etc) são disponibilizadas, permitindo ao usuário acessar conteúdos relacionados à busca inicial. Já a extração de excertos de texto dos documentos que virão a compor o sumário é realizada por um algoritmo proprietário (SEJ, 2007).

Um aspecto relevante nesta ferramenta é que, de acordo com entrevista fornecida ao SEJ (2007), o buscador utiliza motores de busca intermediários para recuperar o grupo de documentos que serão sumarizados. Neste sentido, também podemos atribuir um caráter de meta-buscador ao SenseBot.

Kngine

Kngine é uma ferramenta de busca semântica voltada para serviço de perguntas e respostas. Esta ferramenta é um dos casos que passou de solução para busca na Web para solução apenas corporativa. Sendo assim, atualmente a ferramenta é utilizada em soluções de suporte automático ao cliente, em assistentes pessoais móveis e em soluções de busca corporativa.

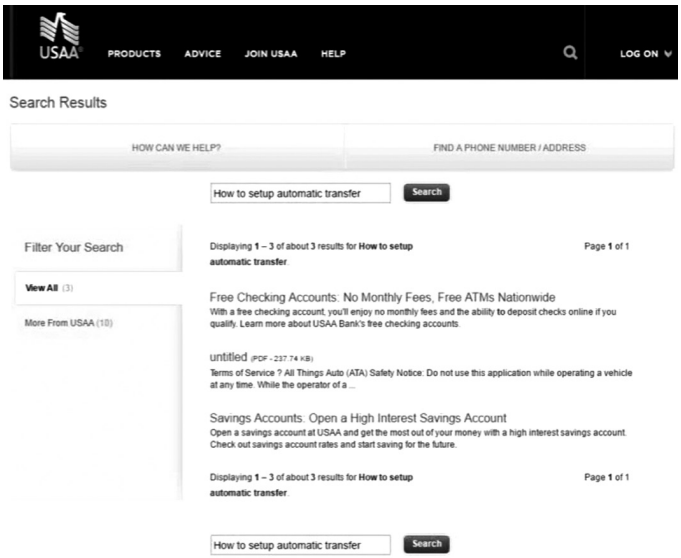

Figura 4. Serviço de consulta sem a ferramenta Kngine
Fonte: Captura de tela gerada pelos autores

Para prover sua funcionalidade, a Kngine baseia-se no uso de grafos de conhecimento e análises estatísticas. O grafo de conhecimento relaciona diferentes conceitos e suas propriedades e é gerado a partir da coleta e análise de dados não-estruturados. Estes dados são processados por meio de análise

de linguagem natural, aprendizagem de máquina e algoritmos de mineração de dados (Kngine, 2017). Nas *Figura 4* e *Figura 5* está representada a comparação entre um serviço de suporte automatizado utilizando-se o Kngine (*Figura 5*) e sem a ferramenta Kngine (*Figura 4*).

A busca por meio do Kngine fornece informações já sumarizadas e voltadas para responder à questão contida na consulta. No exemplo ilustrado na *Figura 4* e *Figura 5*, a consulta é "como configurar transferência automática" no contexto de uma instituição financeira. No serviço de busca ilustrado na *Figura 4*, é retornado uma lista de links que podem estar relacionados com a consulta. Já no serviço do Kngine na *Figura 5*, é retornado uma lista de passos tendo em vista explicar as ações necessárias para resolver o problema ou questão declarado na consulta.

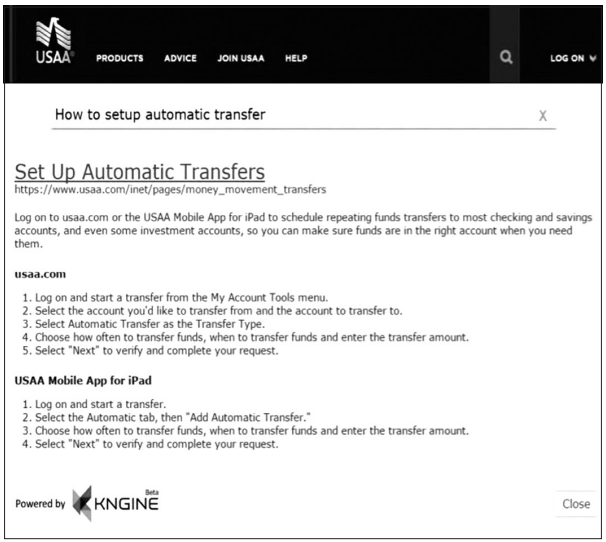

Figura 5. Serviço de pergunta e resposta fornecido pelo Kngine
Fonte: Captura de tela gerada pelos autores

O sistema do Kngine também utiliza análise de linguagem natural e aprendizagem de máquina para compreender a consulta do usuário e recuperar as informações correspondentes no grafo de conhecimento. Além disso, a empresa fornecedora afirma que o Kngine é a primeira ferramenta de pergunta e resposta a suportar múltiplas línguas, como o Inglês, Árabe, Alemão e Espanhol (Kngine, 2017).

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

Google

O motor de busca Google ofereceu seus serviços pela primeira vez em 1998, com buscas baseadas em palavras-chave e utilizando o algoritmo PageRank, desenvolvido por seus fundadores. Entretanto apenas recentemente a ferramenta passou a demonstrar sinais mais claros de componentes semânticos, principalmente por meio de seu algoritmo Hummingbird. Este algoritmo é uma mudança que afasta a ferramenta do uso de apenas termos singulares para a compreensão geral de toda uma frase de consulta (Henshaw, 2013).

O Hummingbird está especialmente focado em semântica, buscando compreender melhor a intenção do usuário por trás da busca. O algoritmo utiliza processamento de linguagem natural e consultas complexas para este propósito (Henshaw, 2013). Uma vez que o Hummingbird é um dos principais componentes na ferramenta de busca, a compreensão total de seu funcionamento é mantida em segredo, assim como acontece com a maior parte do que se desenvolve na empresa Google (Ysasi, 2016), deixando-nos apenas pistas sobre como a busca no Google opera.

Voltado para prover semântica nas buscas do Google são incorporados componentes de pergunta e resposta em combinação com o grafo de conhecimento. O componente de pergunta e resposta é uma abordagem que utiliza consultas baseadas em modelos para mapear a intenção do usuário (Starr, 2013). Estes modelos são na forma de perguntas como, por exemplo, "Quem é W?", "O que é X?", "Quando ocorreu Y?" e "Onde ocorreu Z?". Nesse contexto, normalmente, W corresponde a uma pessoa; X corresponde a uma coisa ou lugar; e Y e Z correspondem a eventos. Uma vez que a consulta do usuário corresponda a um destes modelos, é possível compreender melhor que tipo de informação e sobre que entidade se está buscando.

O grafo de conhecimento pode ser comparado à uma ontologia da WS, contento entidades, propriedades e seus relacionamentos. Isto permite que o Google visualize identidades dentro das consultas ao invés de apenas palavras-chave (Starr, 2014). As respostas para as perguntas que se encaixam nos modelos mencionados anteriormente, em geral, já foram selecionadas, validadas, verificadas e armazenadas no grafo de conhecimento, provendo uma base confiável de informações (Starr, 2013). Nesses casos, o Google é capaz de prover respostas mais concretas, extraindo conteúdo diretamente dos resultados e apresentando este conteúdo de forma sumarizada ao usuário.

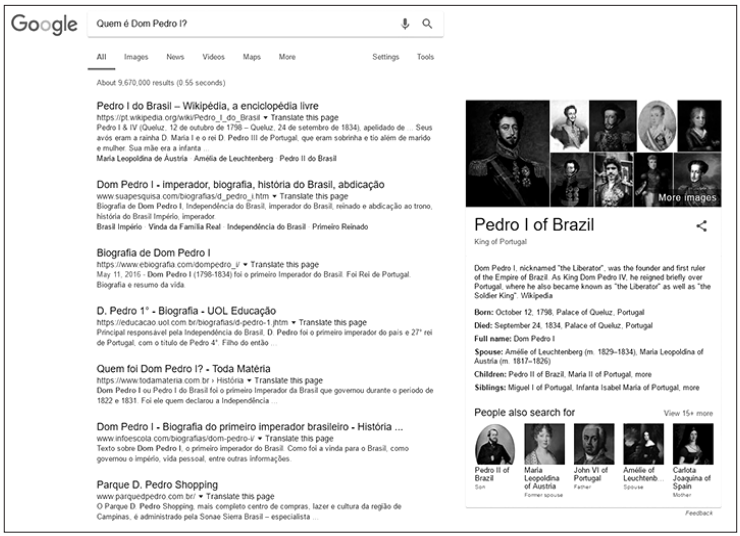

Figura 6. Sumário de informações gerado no buscador Google para a consulta "Quem é Dom Pedro I?"
Fonte: Captura de tela gerada pelos autores

Na *Figura 5* está ilustrado o sumário gerado pelo Google para uma pergunta no modelo "Quem é X?". O conteúdo do sumário é apresentado ao lado direito da lista de resultados e contém diferentes referências para conteúdos relacionados. Além deste sumário, também é possível obter outros tipos de resultados para consultas específicas, como um sumário sobre as condições climáticas de uma região, quando se consulta sobre a temperatura; ou um gráfico ilustrativo quando se consulta sobre estatísticas de um país.

DuckDuckGo

DuckDuckGo é uma ferramenta de busca semântica focada em fornecer respostas diretas para as consultas ao invés da tradicional lista de resultados. Em contraste com o buscador Google, DuckDuckGo é uma ferramenta de busca semântica que visa garantir a privacidade do usuário. A ferramenta não coleta ou compartilha as informações do usuário (para empresas de publicidade, por exemplo), sendo este um dos seus principais diferenciais (Titlow, 2014).

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

Figura 7. Sumário de informações gerado no buscador DuckDuckGo para a consulta "tofu"
Fonte: Captura de tela realizada pelos autores

O mecanismo de busca da ferramenta DuckDuckGo recupera informações de diferentes fontes (Nelson, 2014), além de empregar seu próprio *crawler* nesta tarefa, podendo ser classificado como um motor de busca híbrido. Durante uma busca, os termos inseridos pelo usuário passam por um processo de desambiguação e classificação. Para isso, a ferramenta utiliza uma tecnologia proprietária denominada "detecção semântica de tópico". Depois, a consulta é submetida às fontes com relação mais forte com a consulta do usuário (Weinberg, 2013).

Por exemplo, para uma consulta "tofu com gengibre" os resultados são obtidos a partir de uma base de receitas. Sempre que possível, um sumário de resposta, similar ao fornecido pelo Google, é apresentado e seguido por *links* para diferentes referências. Na *Figura 6* está ilustrada uma busca realizada no DuckDuckGo. Na parte esquerda da *Figura 6*, há a lista de resultados para a consulta "tofu"; enquanto na parte direita há um sumário com informações sobre o tópico da consulta realizada, também denominado "Zero click-info", por não ter sido necessário clicar em nenhum link para obtê-la.

Considerações finais

Neste trabalho buscamos identificar as tecnologias da WS que compõem os buscadores semânticos na atualidade. A partir disso, foi possível perceber que a uma grande parte dos processos de uma ferramenta de busca (coleta, armazenamento, representação, inferência, consulta, elaboração e compreensão) podem beneficiar-se das tecnologias da WS, tendo em vista melhor compreender a intenção do usuário e retornar resultados mais relevantes em suas buscas.

Na segunda parte deste trabalho, descrevemos alguns dos motores de busca semânticos utilizados hoje em dia, visando principalmente identificar seus componentes semânticos. A partir dos resultados obtidos (seção " Buscadores semânticos atuais"), podemos perceber que os motores de busca semântico avaliados confiam principalmente em abordagens baseadas em processamento de linguagem natural, aprendizagem de máquina e mineração de dados para obter melhores interpretações da consulta do usuário e da coleção de documentos disponíveis.

Identificamos que alguns dos motores de busca avaliados (Kngine e Google) utilizam grafos de conhecimento. Estes grafos do conhecimento podem ser entendidos como uma abordagem para estruturação de conteúdo na Web. Nesse contexto, podemos tomar o Google como evidência da necessidade da evolução de buscadores baseados em palavra-chave para buscadores semânticos. Tradicionalmente baseado em palavra-chave, o Google obteve um grande avanço em qualidade após a ativação de seu novo algoritmo Hummingbird, mais voltado para prover semântica nas consultas e baseado em um grafo de conhecimento.

O emprego de tecnologias da WS parece estar no momento voltado para contextos específicos, como foi possível observar por meio dos exemplos na seção " Aplicação das tecnologias da web semântica nos buscadores web". Isto é compreensível, uma vez que a maior parte do conteúdo da Web é desprovido de semântica. Nesse caso, a anotação semântica das informações nestes contextos específicos torna-se parte do processo de construção dos motores de busca semânticos. Por outro lado, outra abordagem é também fornecer a busca baseada em palavras-chave ao lado da busca semântica, de modo a aumentar o alcance do motor de busca permitindo que dados não-estruturados sejam recuperados.

Neste trabalho, fornecemos uma amostra das possíveis aplicações das tecnologias da WS nos buscadores Web. Como trabalhos futuros, sugere-se o aprofundamento da investigação sobre o emprego das tecnologias da WS para cada um dos processos identificados, de forma a identificar técnicas e métodos específicos disponíveis para o aprimoramento dos buscadores Web.

Referências

Arenas, H., B. Harbelot e C. A. Cruz. 2014. "A Semantic Web Approach for Geodata Discovery". *Lecture Notes In Computer Science*, 117-126.

Ayaz, B., W. Altaf, F. Sadiq, H. Ahmed e M. A. Ismail. 2016. "Novel Mania: A semantic search engine for Urdu". *International Conference On Open Source Systems & Technologies*, 42-47.

Berners-Lee, T., O. Lassila e J. Hendler. 2001. "The semantic web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". *Scientific American*.

Borst, W. N. 1997. "Construction of Engineering Ontologies For Knowledge Sharing and Reuse". Tese (Doutorado), Institute for Telematica and Information Technology, Universidade de Twente, Enschede, Holanda.

Fayer, D. C., O. Curé e G. Blin. 2012. "A survey of RDF storage approaches". *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées* 15, 11-35.

Fatima, A., C. Luca e G. Wilson. 2014. "New Framework for Semantic Search Engine". *Uksim-amss 16th International Conference On Computer Modelling And Simulation*, 446-451.

Fernández, J. D., M. Lenzerini, M. Masseroli, F. Venco e S. Ceri. 2016. "Ontology-Based Search of Genomic Metadata". *IEEE/ACM Transactions On Computational Biology And Bioinformatics* 13 (2): 233-247.

Gruber, T. R. 1993. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". *International Journal Of Human-computer Studies* 43 (5-6): 907-928.

Gruber, T. R. 2009. "Ontology". Acesso em 15 jun. 2018. http://tomgruber.org/writing/ontology-definition-2007.htm

Henshaw, J. 2013. "What Google's Hummingbird Update Means for Content Marketers". Acesso em 15 jun. 2018. https://smallbusiness.yahoo.com/advisor/google-hummingbird-means-content-marketers-143119302.html

IDC. 2014 "Executive Summary: Data Growth, Business Opportunities, and the IT Imperatives". *International Data Corporation*. Acesso em 15 jun 2018. https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

Kalaivani, S. e K. Duraiswamy. 2012. "Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment". *Journal Of Computer Science* 8 (9): 1407-1413.

Kngine. 2017. "Technology: Our technology is pure rocket science". Acesso em 15 jun. 2018. http://www.kngine.com/Technology.html

Laura, L. e G. Me. 2015. "Searching the Web for illegal content: the anatomy of a semantic search engine". *Soft Computing* 21 (5): 1245-1252.

Ma, S., W. Zhao, S. Zhang e H. Zhang. 2012. "Material Hub: A Semantic Search Engine with Rule Reasoning". *IEEE 36th Annual Computer Software And Applications Conference Workshops*, 38-44.

Nagpál, G. 2005. "Improving Information Retrieval Effectiveness by Using Domain Knowledge Stored in Ontologies". *OTM Confederated International Conferences On the Move to Meaningful Internet Systems*, 780-789.

Nelson, J. 2014. "What's the source of DuckDuckGo search results?". Acesso em 15 jun. 2018. https://www.quora.com/Whats-the-source-of-DuckDuckGo-search-results

Ontotext. 2017. "What is RDF Triplestore?". Acesso em 15 jun. 2018. http://ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore

Pellegrini, T. 2006. "Jeen Broekstra: The importance of SPARQL can not be overestimated". Acesso em 15 jun. 2018. http://archive.is/YLhci

Qu, J., C. Wei, W. Wang e F. Liu. 2011. "Research on a Retrieval System Based on Semantic Web". *International Conference On Internet Computing And Information Services,* 543-545.

DOI: http://dx.doi.org/10.22201/ibi.24488321xe.2019.78.57977

Sayed, A. e A. A. Muqrishi. 2017. "IBRI-CASONTO: Ontology-based semantic search engine". *Egyptian Informatics Journal* 18 (3), 1-12.

SEJ. 2007. "Summarization, the Answer to Web Search: Interview with Dmitri Soubbotin of SenseBot". Acesso em 15 jun. 2018. https://www.searchenginejournal.com/summarization-the-answer-to-web-search-interview-with-dmitri-soubbotin-of-sensebot/6094/

Semantic Engines LLC. 2017. "SenseBot - semantic search engine that finds sense on the Web". Acesso em 15 jun. 2018. http://www.semanticengines.com/

Singh, B. e H. K. Singh. 2010. "Web Data Mining research: A survey". *IEEE International Conference On Computational Intelligence And Computing Research*, 661-670.

Soliman, S. S., M. F. El-Sayed e Y. F. Hassan. 2015. "Semantic Clustering of Search Engine Results". *The Scientific World Journal* 2015, 1-9.

Starr, B. 2014. "Demystifying The Google Knowledge Graph". Acesso em 15 jun. 2018. http://searchengineland.com/demystifying-knowledge-graph-201976

Starr, B. 2013. "Google Hummingbird: When Evolutionary Becomes Revolutionary". Acesso em 15 jun. 2018. http://searchengineland.com/google-hummingbird-when-evolutionary-becomes-revolutionary-173740

Studer, R., V. R. Benjamins e D. Fensel. 1998. "Knowledge engineering: Principles and methods". *Data & Knowledge Engineering* 25 (1-2): 161-197

Titlow, P. J. 2014. "Inside DuckDuckGo, Google's Tiniest, Fiercest Competitor". Acesso em 15 jun. 2018. https://www.fastcompany.com/3026698/inside-duckduckgo-googles-tiniest-fiercest-competitor

Weinberg, G. 2013. "DuckDuckGo Architecture - 1 Million Deep Searches A Day And Growing. High Scalability". Acesso em 15 jun. 2018. http://highscalability.com/blog/2013/1/28/duckduckgo-architecture-1-million-deep-searches-a-day-and-gr.html

Wikipedia. 2017. "List of subject-predicate-object databases: Implementations". Acesso em 15 jun. 2018. https://en.wikipedia.org/wiki/List_of_subject-predicate-object_databases

W3C. 2008. "Extensible Markup Language (XML)". Acesso em 15 jun. 2018. https://www.w3.org/TR/XML/

W3C. 2016. "LargeTripleStores". Acesso em 15 jun. 2018. https://www.w3.org/wiki/LargeTripleStores

W3C. 2004. "OWL Web Ontology Language: Overview". Acesso em 15 jun 2018. https://www.w3.org/TR/owl-features/

W3C. 2012. "OWL 2 Web Ontology Language: Document Overview (Second Edition)". Acesso em 15 jun. 2018. https://www.w3.org/TR/owl2-overview/

W3C. 2011. "OWL Reasoner". Acesso em 15 jun. 2018. https://www.w3.org/2001/sw/wiki/Category:OWL_Reasoner

W3C. 2014. "RDF 1.1 Turtle: Terse RDF Triple Language". Acesso em 15 jun. 2018. https://www.w3.org/TR/turtle/

W3C. 2015. "SEMANTIC WEB". Acesso em 15 jun. 2018. https://www.w3.org/standards/semanticweb/

Ysasi, E. 2016. "What is Google's Semantic Search?". Acesso em 15 jun. 2018. https://www.theleverageway.com/blog/google-semantic-search/#top

Para citar este texto:

Vitor Rozsa, Angel Freddy Godoy Viera y Moisés Dutra. 2019. "Aplicação de Tecnologias da Web Semântica em Motores de Busca na Internet". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 33 (78): 165-191.
http://dx.doi.org/10.22201/iibi.24488321xe.2019.78.57977