



Ferramentas de Busca Semântica na Web

Bernadete Aparecida de Aquino

**JUIZ DE FORA
DEZEMBRO, 2010**

Ferramentas de Busca Semântica na Web

Bernadete Aparecida de Aquino

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharel em Ciência da Computação

Orientadora: Regina Maria Maciel Braga Villela

**JUIZ DE FORA
DEZEMBRO, 2010**

FERRAMENTAS DE BUSCA SEMÂNTICA NA WEB

Bernadete Aparecida de Aquino

MONOGRAFIA SUBMETIDADA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Regina Maria Maciel Braga Villela
DSc em Engenharia de Sistemas e Computação COPPE/UFRJ

Alessandréia Marta de Oliveira Júlio
Msc em Engenharia de Sistemas e Computação COPPE/UFRJ

Kele Teixeira Belloze
Msc em Sistemas e Computação IME/RJ

JUIZ DE FORA, MG - BRASIL
JULHO, 2010

RESUMO

Em um curto período de tempo, a web tornou-se uma rede rica de informações capaz de levar o conhecimento do mundo para os lugares mais distantes. Contudo, como esse número de informação se expande a cada dia, a obtenção de dados relevantes à pesquisa tornou-se um desafio. A área de pesquisa na web tem sido fonte de crescentes inovações. Neste contexto, a web semântica é uma das vertentes de pesquisa, capaz de melhorar os resultados obtidos pelas ferramentas de busca, usando, não apenas palavras, mas conceitos e relações lógicas. Este trabalho apresenta a web semântica e as ferramentas de busca que a utilizam.

Palavras-chave: Web Semântica. Busca Semântica. Ferramentas de Busca Semântica.

ABSTRACT

In a brief span of time, the web has become a rich network of information capable of bringing the world's knowledge to the farthest places. However, as this information content grows every day, obtaining data relevant to the search become a challenge. The area of web search has been a booming place to find semantic web innovations. In this context, the semantic web is one of the strands of search that can augment and improve traditional search results by using, not just words, but concepts and logical relationships. This work presents the semantic web and search engines that use it.

Keywords: Semantic Web. Semantic Search. Semantic Search Engines.

SUMÁRIO

LISTAS QUADROS	7
LISTA DE FIGURAS	8
LISTA DE ABREVIACÕES	9
1 INTRODUÇÃO.....	10
1.1 MOTIVAÇÃO	10
1.2 OBJETIVOS.....	11
1.2.1 Objetivos Gerais.....	11
1.2.2 Objetivos Específicos.....	11
1.3 ORGANIZAÇÃO DO TRABALHO	11
2 WEB SEMÂNTICA.....	12
2.1 INTRODUÇÃO	12
2.2 ARQUITETURA DA WEB SEMÂNTICA.....	13
2.3 UNIVERSAL RESOURCE IDENTIFIER (URI) E UNICODE.....	14
2.4 EXTENSIBLE MARKUP LANGUAGE (XML)	14
2.5 RESOURCE DESCRIPTION FRAMEWORK (RDF).....	15
2.5.1 RDF Schema	16
2.6 SPARQL.....	17
2.7 ONTOLOGIA.....	19
2.7.1 Linguagem para especificação de Ontologias	20
2.7.2 OWL.....	20
2.8 REGRAS PARA A WEB SEMÂNTICA	22
2.8.1 RIF.....	23
2.8.2 SWRL.....	23
2.9 LÓGICA UNIFICADORA.....	24
2.10 PROVA, CONFIANÇA E CRIPTOGRAFIA	24
2.11 APLICAÇÕES PARA A WEB SEMÂNTICA	25
3 USO DE FERRAMENTAS DE BUSCA SEMÂNTICA NA WEB	26
3.1 INTRODUÇÃO	26
3.2 BUSCA NA WEB SEMÂNTICA	26
3.3 UM MODELO ARQUITETURAL PARA FERRAMENTAS DE BUSCA SEMÂNTICA.....	27
3.3.1 Ontology Crawler.....	28
3.3.2 Ontology Annotator	28
3.3.3 Web Crawler	29
3.3.4 Query Builder e Query Pre-processor	29
3.3.5 Inference Engine	30
3.4 FERRAMENTAS DE BUSCA SEMÂNTICA EXISTENTES	30
3.4.1 TrueKnowledge.....	30
3.4.2 Kngine	32
3.4.3 Wolfram Alpha	33
3.4.4 Duck Duck Go	35
3.4.5 Sig.ma.....	36
3.4.6 Comparação entre as Ferramentas	37
4 ESTUDO DE CASO.....	38

4.1	INTRODUÇÃO	38
4.2	PESQUISA SOBRE AS FERRAMENTAS DE BUSCA SEMÂNTICA	38
4.2.1	Análise dos Dados.....	39
4.3	INTERFACE DE BUSCA PARA AS FERRAMENTAS SEMÂNTICA	40
4.3.1	Conexão Com as API's	41
4.3.2	Arquitetura da Aplicação.....	41
4.3.3	Consultas no Front-End.....	43
5	CONSIDERAÇÕES FINAIS	45
	REFERÊNCIAS.....	46

LISTA DE QUADROS

3.1 Quadro Comparativo entre as Ferramentas	37
---	----

LISTA DE FIGURAS

2.1	Estrutura em camadas da Web Semântica	13
2.2	Representação XML.....	15
2.3	Exemplo <i>statement</i> RDF.....	16
2.4	Camadas RDF e RDFS.....	17
2.5	Exemplo base de conhecimento.....	18
2.6	Ontologia de uma Árvore	22
2.7	Representação SWRL para a regra ‘tio’	24
3.1	Arquitetura para Ferramentas de Busca Semântica	28
3.2	Arquitetura TrueKnowledg.....	31
3.3	Consulta na Ferramenta TrueKnowledge	32
3.4	Consulta na Ferramenta Kngine.....	33
3.5	Consulta na Ferramenta Wolfram Alpha	34
3.6	Consulta na Ferramenta Duck duck go	35
3.7	Dataflow da Ferramenta Sig.ma	36
3.8	Consulta na Ferramenta Sig.ma	37
4.1	Arquitetura da Aplicação.....	42
4.2	Método Responsável por Recuperar as Informações dos XMLs de Resposta	42
4.3	Consulta Xpath no XML de TrueKnowledge.....	43
4.4	Tela Inicial da Aplicação	43
4.5	Pesquisa por <i>Water Cycle</i>	44

LISTA DE ABREVIACÕES

API	<i>Application Programming Interface</i>
DTD	<i>Document Type Definition</i>
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
JSP	<i>Java Server Page</i>
OWL	<i>Web Ontology Language</i>
RDF	<i>Resource Description Framework</i>
RDFS	<i>Resource Description Framework Schema</i>
SQL	<i>Structured Query Language</i>
SPARQL	<i>Simple Protocol and RDF Query Language</i>
SWRL	<i>Semantic Web Rule Language</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>Extensible Markup Language</i>

1 INTRODUÇÃO

As ferramentas de busca surgiram como uma forma de amparar a recuperação de páginas relacionadas a temas específicos disponíveis na web. Sem elas, os usuários teriam que conhecer todos os endereços de sites que os interessassem ou utilizar os hiperlinks para acessar novos endereços.

O grande número de páginas disponíveis na internet, fez com que essas ferramentas de busca não fossem mais capazes de retornar resultados precisos. Isso ocorre porque elas realizam suas buscas baseadas em palavras-chaves, não se preocupando com as palavras que podem ter mais de um significado e com palavras ou frases que podem designar o mesmo conceito.

O principal empecilho para se proporcionar melhores resultados para as buscas é o fato dos significados dos conteúdos disponíveis na web atual não serem acessíveis às máquinas. Com isso, elas não podem interpretar as sentenças e extrair informações úteis aos usuários.

Considerada como uma extensão da web atual, a Web Semântica visa estruturar as informações de forma que as máquinas possam as interpretar e utilizar na realização de inferências. Com isso, tornou-se possível o desenvolvimento de ferramentas de busca capazes de ‘entender’ o conteúdo das páginas web e realizar raciocínio lógico sobre elas para executar consultas complexas e retornar resultados precisos.

1.1 MOTIVAÇÃO

O modelo de busca atual faz uma recuperação de termos descontextualizados, não permitindo uma recuperação precisa, organizada e eficaz das informações. A estruturação dos documentos proposta pela web semântica possibilita uma transformação na forma de se recuperar informação na web, levando a busca para mais perto das necessidades dos usuários.

As ferramentas de busca semântica tornaram-se uma resposta para a superação das limitações apresentadas pelas ferramentas de busca tradicionais. Uma vez que não realizam a busca baseada em palavras chaves, mas sim tentam examinar e ‘entender’ o que

o usuário deseja, através de análises lógicas da consulta, para que assim possam recuperar resultados mais precisos.

1.2 OBJETIVOS

1.2.1 Objetivos Gerais

O desenvolvimento deste trabalho tem por objetivo fazer um estudo sobre as ferramentas de busca por informações na web que utilizam semântica, uma vez que estas ferramentas vêm se apresentando como uma solução para a recuperação de informações relevantes à consulta.

1.2.2 Objetivos Específicos

Realizar um estudo sobre a web semântica e sobre algumas das principais ferramentas de busca semântica atuais, verificando quais dentre estas ferramentas são mais indicadas para determinadas áreas do conhecimento e em seguida, desenvolver um protótipo de busca utilizando ferramentas de busca no contexto da web semântica.

1.3 ORGANIZAÇÃO DO TRABALHO

O trabalho apresenta a web semântica e suas camadas no capítulo 2, onde ainda são citadas algumas das áreas onde ela pode ser aplicada.

Já no capítulo 3, uma arquitetura conceitual para as ferramentas de busca semântica é apresentada, em seguida são mostradas algumas das principais ferramentas de busca que utilizam semântica.

Por fim, o capítulo 4 apresenta um estudo sobre quais das ferramentas citadas no capítulo 3 são mais indicadas para pesquisas em determinadas áreas de interesse. Por ultimo é construído uma interface de busca integrada a tais ferramentas.

2 WEB SEMÂNTICA

2.1 INTRODUÇÃO

A Web atual é composta por diversas páginas que para as máquinas, só oferecem informações sobre como deve ser apresentado o conteúdo. As máquinas não sabem o significado do documento que estão apresentando, na verdade, cada documento na Web parece exatamente o mesmo para elas. Por isso, não são capazes de tomar nenhuma decisão inteligente. Para eliminar esta limitação é necessário adicionar metadados às páginas, a fim de permitir que nesta nova web as máquinas possam “compreender” o significado dos documentos, tornando possível a construção de diversas ferramentas que analisarão os documentos de forma global. Esta nova Web é chamada de Web Semântica (YU, 2007).

Existem diferentes ideias sobre o que é a web semântica. Seu inventor, Tim Berners-Lee, a descreve como sendo uma extensão da web atual, na qual a informação é utilizada com significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação (SEMANTIC, 2010).

Na web semântica, os dados são precisos e interligados de modo que os computadores possam melhor utilizá-los para a descoberta, automatização, integração e reutilização entre as diversas aplicações (ALESSO, 2006).

A web semântica fornece infraestrutura não somente para as páginas web, mas também para bancos de dados, aplicações, sensores e dispositivos pessoais. Com isso, softwares inteligentes podem utilizar essas informações em pesquisas, filtros e diversas outras funcionalidades (ALESSO, 2006).

Uma das aplicações para a web semântica é proporcionar pesquisas mais eficientes, uma vez que os sistemas podem inferir quando as palavras ou frases são similares. Uma busca por “Barack Obama” pode retornar páginas, igualmente válidas, que se referem ao “presidente dos Estados Unidos”. Por outro lado, é possível distinguir onde a mesma palavra é usada com significados diferentes. Ao pesquisar por “Jaguar”, no contexto de indústria automobilística, o sistema poderá desconsiderar referências a “onça-pintada”. Quando pouca informação for encontrada a respeito do tema da pesquisa o sistema, como alternativa, poderá localizar páginas com assunto relacionado (DAVIES, 2006).

Para que as máquinas possam servir aos humanos de maneira mais eficiente é necessário construir instrumentos que forneçam a elas sentido lógico e semântico. Nas seções seguintes, serão descritos estes mecanismos.

2.2 ARQUITETURA DA WEB SEMÂNTICA

O desenvolvimento da web semântica é realizado construindo-se uma camada em cima da outra. Sendo necessário que cada camada seja compatível com a camada inferior e tenha ao menos um entendimento parcial da camada superior. Ou seja, agentes devem interpretar e usar dados de uma camada inferior e ter algum conhecimento dos dados da camada superior (ALESSO, 2006).

A arquitetura da web semântica já passou por diversas alterações, à medida que componentes mais abrangentes ou eficazes são desenvolvidos. Sua configuração atual é apresentada na figura 2.1.

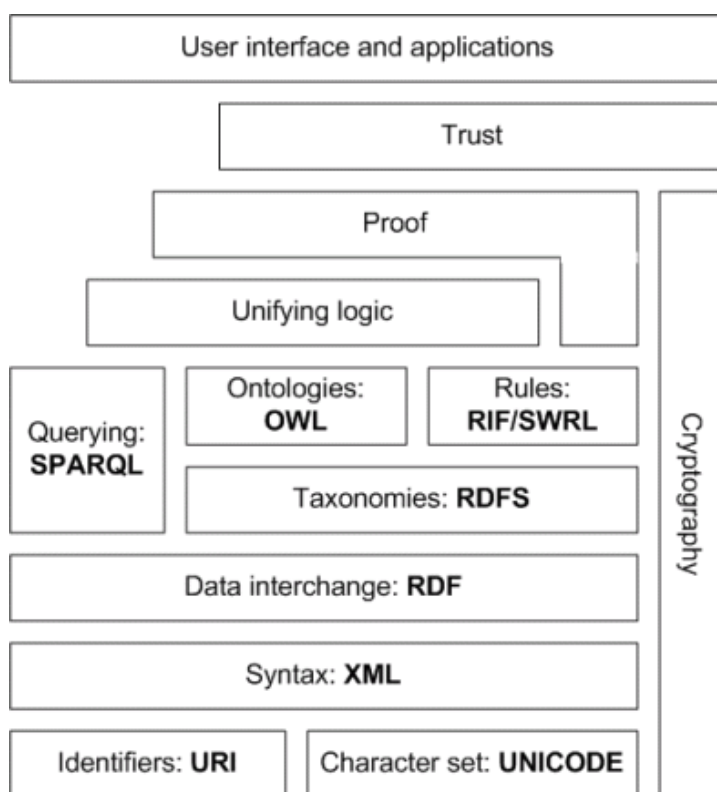


Figura 2.1 - Estrutura em camadas da Web Semântica (SEMANTIC, 2009)

2.3 UNIVERSAL RESOURCE IDENTIFIER (URI) E UNICODE

O URI é um padrão que busca identificar um recurso físico ou abstrato de forma única e global, não necessariamente provê acesso ao recurso, deve ser considerado simplesmente como uma convenção para a nomenclatura de recursos. O tipo mais comum de URI é o Uniform Resource Locator (URL) que é o endereço de um recurso disponível na rede (JARDIM, 2009).

Uma versão simplificada da sintaxe URI seria a seguinte (FENSEL, 2007):

scheme : [//authority][path][?query][#fragid]

O padrão URI deve ser utilizado para evitar a ambiguidade na web. Por isso, é importante a utilização de URIs diferentes para conceitos diferentes ou para a variação do mesmo conceito.

O Unicode é um padrão de codificação para fornecer uma representação numérica e sem ambiguidade para cada caractere de maneira independente da linguagem e da plataforma de software (JARDIM, 2009).

2.4 EXTENSIBLE MARKUP LANGUAGE (XML)

XML é uma meta linguagem, linguagem utilizada para criar outras linguagens. Foi desenvolvida pela W3C (*World Wide Web Consortium*) em 1996, com o intuito de estruturar, armazenar e transportar informações (ALESSO, 2006).

O documento XML permite que as *tags*, elementos que procuram dar significado aos dados, recebam qualquer nome, assim pode-se definir nomes que façam sentido no contexto do documento.

Uma característica importante da linguagem XML é a separação do conteúdo da apresentação. Com isso, é possível que o conteúdo seja utilizado para outros fins além da exibição (ANTONIOU, 2008).

Para descrever a estrutura dos dados, quais são os elementos e atributos utilizáveis e quais elementos podem aparecer como subelemento de outro, é utilizado o DTD (*Document Type Definition*) ou o XML Schema.

A figura 2.2 demonstra um exemplo da utilização da linguagem XML na descrição de um determinado livro.

```

<book>
  <title>
    Nonmonotonic Reasoning: Context-Dependent Reasoning
  </title>
  <author>V. Marek</author>
  <author>M. Truszczynski</author>
  <publisher>Springer</publisher>
  <year>1993</year>
  <ISBN>0387976892</ISBN>
</book>

```

Figura 2.2 – Representação XML (ANTONIOU, 2008)

Devido às peculiaridades do padrão XML, o mesmo é recomendado pela W3C como padrão de troca de dados na web. Porém, o mesmo não oferece capacidade semântica e lógica. O *Resource Description Framework* (RDF) é que começa a estabelecer uma base para a semântica na web, assunto que será tratado a seguir.

2.5 RESOURCE DESCRIPTION FRAMEWORK (RDF)

As *tags* XML podem muitas vezes dar significado aos dados, no entanto, a compreensão dessas *tags* é significativa apenas aos seres humanos. Para solucionar este problema, foi desenvolvido o RDF que adiciona metadados legíveis por máquina nos dados existentes na web (FENSEL, 2007).

Os conceitos fundamentais do RDF são os recursos, as propriedades e as declarações (*statements*). Os recursos representam os objetos, as coisas que se queira modelar, podendo ser livros, autores, lugares, pessoas, dentre vários outros. Já as propriedades são responsáveis por definir as relações entre os recursos, por exemplo, “livro” poderia ter a propriedade “escrito por”. Por fim, as declarações definem uma tripla que consiste de um recurso, uma propriedade e um valor. Uma tripla pode ser representada como: um recurso [sujeito] tem uma propriedade [predicado] com um valor específico [objeto] (ANTONIOU, 2008).

Um exemplo simples de uma *statement* seria: O livro possui o título *Thinking on the Web: Berners-Lee, Godel and Turing*. Que pode ser definido como a tripla (O livro, possui o título, *Thinking on the web*). A figura 2.3 mostra o grafo correspondente desta tripla.

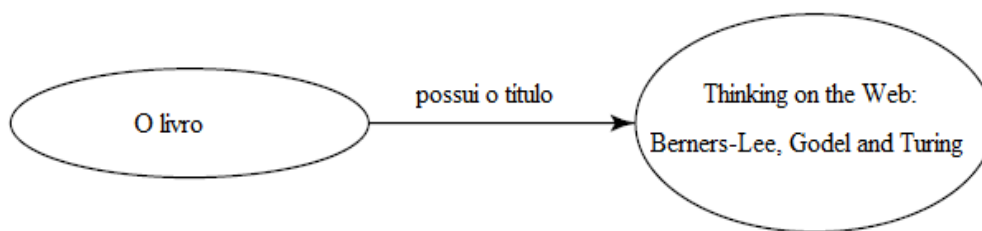


Figura 2.3 – Exemplo *statement* RDF

O sujeito, predicado e objeto no RDF são identificados por referências a URI. Existe algumas situações que seria inconveniente usar um URI para o objeto, por exemplo, quando esse possui um valor inteiro. Portanto, o RDF também permite que valores literais possam ser usados como objeto, mas não como sujeito e predicado (WALTON, 2007).

2.5.1 RDF Schema

O RDF fornece um modelo simples e poderoso para descrever a informação, mas a semântica é descrita usando RDFS. O objetivo do RDFS é fornecer um vocabulário de termos que possa representar as classes e seus relacionamentos (ALESSO, 2006).

Uma classe pode ser considerada como um conjunto de elementos. Os objetos que pertencem a uma determinada classe são instâncias dessa classe. A relação entre as instâncias e a classe é expressa por “rdf:type” (ALESSO, 2006).

As classes são usadas para impor restrições sobre o que pode ser atribuído em um documento RDF. Por exemplo, *statements* como “A sala 210 é ensinada por David” devem ser restringidos, uma vez que somente disciplinas podem ser ensinadas. Em termos matemáticos, foi estabelecida uma restrição ao domínio da propriedade (ANTONIOU, 2008).

Uma vez que as classes já tenham sido definidas, deve-se estabelecer o relacionamento entre elas. Considerando-se a seguinte restrição: Cursos podem ser ministrados somente pela equipe acadêmica. Se David for considerado um professor então, de acordo com a restrição, ele não está apto a ministrar cursos. A razão é que não há declaração especificando que professor é um membro da equipe acadêmica. É necessário que a classe professor herde da classe equipe acadêmica. No RDFS, a relação de herança é proporcionada por “rdfs:subClassOf”, para classes e “rdfs:subPropertyOf” para propriedades (ANTONIOU, 2008).

As propriedades possibilitam expressar relacionamentos entre as classes e suas instâncias ou superclasse, nelas podem existir restrições que são introduzidas através de

“rdfs: domain” que restringe o conjunto de recursos para uma determinada propriedade e “rdfs: range” que restringe o conjunto de valores que pode ter uma propriedade.

O RDF *statement* “Nuclear Physics is thought by Professor Smith” é mostrado na figura 2.4 que faz uma separação das camadas RDF e RDFS.

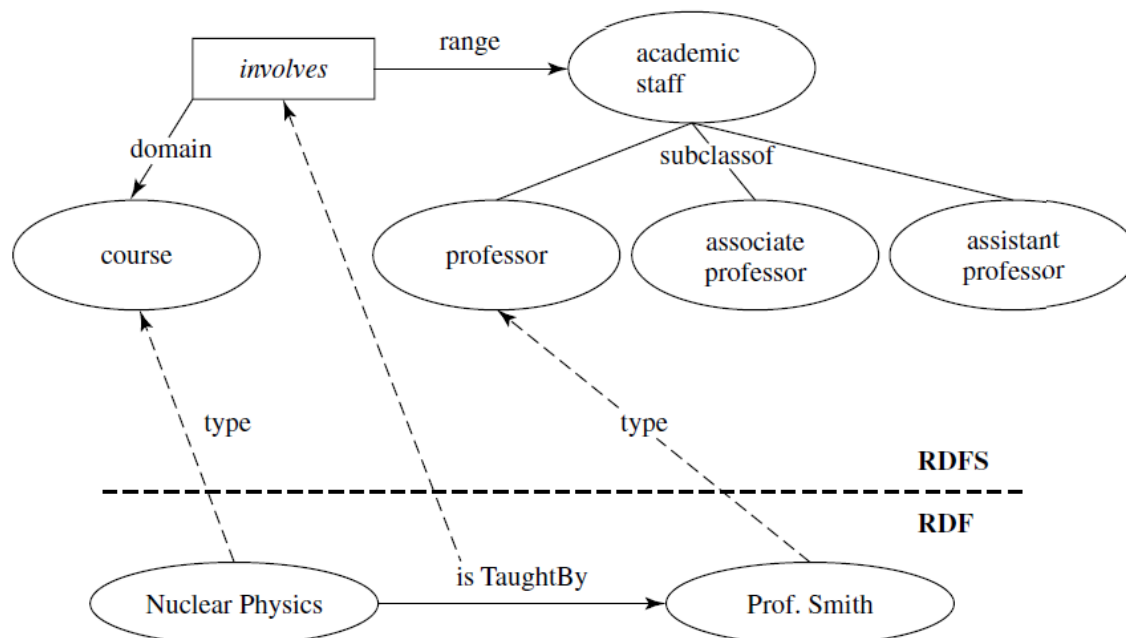


Figura 2.4 – Camadas RDF e RDFS (ALESSO, 2006)

2.6 SPARQL

O padrão Simple Protocol and RDF Query Language (SPARQL) define um protocolo de rede para a troca de consultas e uma linguagem para realizar consultas em RDF (WALTON, 2007).

A maior parte do uso da sigla SPARQL se refere à linguagem de consulta RDF que é como uma SQL mais poderosa, mas para dados RDF (DODDS, 2005).

O protocolo SPARQL é um método de invocação de consultas remotas. Ele especifica uma interface simples que é suportada por HTTP (Hypertext Transfer Protocol) ou SOAP (Simple Object Access Protocol), onde um cliente pode usar esta interface para realizar consultas SPARQL em banco de dados ou em qualquer repositório que possa ser mapeado para o modelo RDF (DODDS, 2005).

Como em SQL, a consulta SPARQL tem a estrutura *select-from-where*. O *select* é usado para especificar quais devem ser os dados recuperados. O *from* é usado para especificar a fonte a ser consultada, não é uma cláusula obrigatória, quando não informada

deve-se assumir que a consulta será realizada em uma base de conhecimento particular. Já o *where* impõe restrições nas possíveis saídas do grafo, para isso utiliza uma tripla entre chaves, onde cada posição da tripla pode ser uma variável, uma referência URI, um literal ou um nó em branco. As referências URI são delimitadas por “< >” para serem distinguidas dos literais e das variáveis que são precedidas por uma interrogação (ANTONIOU, 2008).

Para demonstrar o uso da consulta SPARQL, a figura 2.5 apresenta uma base de conhecimento RDF sobre modelos de câmeras fotográficas. Foi utilizada a notação N3 para definir as triplas RDF, uma vez que este estilo é mais compactado do que a representação XML. Neste modelo, escreve-se uma tripla como sequência de termos e os prefixos, no início da definição, são abreviações que associam nomes a referência URI (WALTON, 2007).

```

@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
@prefix camera: <http://www.mycamera.org/photo#>
@prefix canon: <http://www.mycamera.org/canon#>
@prefix nikon: <http://www.mycamera.org/nikon#>
@prefix pentax: <http://www.mycamera.org/pentax#>
canon:CanonEOS300D camera:manufacturer canon:CanonInc .
canon:CanonEOS300D camera:hasBack camera:DigitalBack .
canon:CanonEOS300D camera:resolution "6.3"^^xsd:decimal .
nikon:NikonD70 camera:manufacturer nikon:NikonCorporation .
nikon:NikonD70 camera:hasBack camera:DigitalBack .
nikon:NikonD70 camera:resolution "6.1"^^xsd:decimal .
pentax:PentaxK1000 camera:manufacturer pentax:PentaxCorporation .
pentax:PentaxK1000 camera:hasBack camera:35mm .

```

Figura 2.5 – Exemplo base de conhecimento (WALTON, 2007)

A consulta SPARQL a seguir seleciona câmeras na base de conhecimento que possuam *back* digital.

```

SELECT ?x
WHERE { ?x <http://www.mycamera.org/photo#hasBack>
        <http://www.mycamera.org/photo#DigitalBack> }

```

O resultado da consulta na base de conhecimentos é apresentado a seguir.

?x
canon:CanonEOS300D
nikon:NikonD70

A forma da resposta à consulta é definida na instrução *select*. No exemplo acima, a resposta deveria consistir apenas do primeiro termo (?x) de qualquer correspondência

com o padrão da consulta. Para obter toda a tripla seria necessário utilizar o asterisco no lugar da variável (WALTON, 2007).

2.7 ONTOLOGIA

O termo ontologia é originário da filosofia, onde é utilizado como o estudo da existência ou do ser. Recentemente, este termo foi introduzido na tecnologia da informação com um sentido bastante diferente do original. Em vez de “ontologia” fala-se agora de “uma ontologia” que segundo R. STUDER citado por ANTONIOU (2008) pode ser definida como uma especificação explícita e formal de uma conceituação.

Uma ontologia fornece um vocabulário preciso com o qual o conhecimento pode ser representado. Este vocabulário permite que sejam especificadas quais as entidades que serão representadas, como elas poderão ser agrupadas e que relações às conectam. Quanto mais precisa a ontologia, maior será o entendimento de como os dados poderão ser usados. Mas, se a ontologia for excessivamente complexa, ela pode se tornar confusa, complicada e difícil de utilizar, manter, e estender (SEGARAN, 2009).

Os benefícios do uso das ontologias podem ser citados como sendo o entendimento comum e compartilhado sobre certos conceitos do domínio; a possibilidade de se reutilizar o conhecimento de um domínio; o fornecimento, juntamente com as linguagens de ontologia, de uma maneira de codificar o conhecimento e a semântica de tal forma que as máquinas possam “compreender” e a possibilidade de se fazer o processamento automático entre as máquinas (YU, 2007).

As ontologias são geralmente expressas em uma linguagem baseada em lógica, de modo que distinções consistentes e significativas podem ser feitas entre as classes, propriedades e relações. Atualmente, as linguagens de ontologia mais importantes para web (ANTONIOU, 2008) são o RDF que fornece recursos para a descrição da informação na web, o RDF Schema que disponibiliza informações sobre a interpretação das expressões existentes em um modelo RDF e a Web Ontology Language (OWL), linguagem mais poderosa que RDF e RDFS para a definição de ontologias.

2.7.1 Linguagem para especificação de Ontologias

As linguagens para especificação de ontologias devem permitir aos usuários escrever de forma explícita e formal os conceitos de um domínio. Tanto RDF quanto RDF Schema permitem a representação do conhecimento. Entretanto, a capacidade de expressão nessas linguagens é muito limitada (WALTON, 2007).

A principal preocupação das linguagens RDF/RDFS é a organização dos vocabulários em hierarquia. Entretanto, algumas características importantes são deixadas de lado. Como exemplos podem ser citados: escopo local de uma propriedade, não se pode declarar uma restrição em um intervalo de valores que se aplica apenas a algumas classes; classes disjuntas; combinações booleanas entre as classes, às vezes é necessário construir uma nova classe através da combinação de outras usando união, intercessão e complemento; restrições na cardinalidade, colocar restrições em quantos valores distintos uma propriedade pode levar; definir características especiais para as propriedades, tais como propriedades únicas, transitivas ou inversas (ANTONIOU, 2008).

Portanto, uma linguagem de ontologia mais rica que RDF/RDFS com relação a esses recursos adicionais é necessária. Na concepção desta linguagem, o equilíbrio entre o poder expressivo e o eficiente apoio ao raciocínio deve ser estabelecido (ALESSO, 2006). Alguns pesquisadores propuseram uma linguagem de ontologia que estende as facilidades da RDF e RDFS. Este trabalho culminou com a definição da OWL que será descrita a seguir (WALTON, 2007).

2.7.2 OWL

A Web Ontology Language (OWL) é uma linguagem desenvolvida pela W3C para a definição de classes e propriedades, bem como para permitir inferências e uma representação mais poderosa de um domínio (SEGARAN, 2009).

A ontologia inclui quatro conceitos que formam a base dos documentos OWL: classes, relacionamento entre as classes, propriedades e restrições sobre as relações entre as classes e propriedades (ALESSO, 2006).

Ao contrário do RDF, o vocabulário OWL é extenso. A OWL faz uso de elementos RDFS, no entanto, possui vários conceitos originais.

Para atender as necessidades da web semântica, a OWL não é uma linguagem única, mas sim uma família de três linguagens: OWL-Lite, OWL-DL e OWL Full. Essas linguagens formam uma sequência crescente de expressividade que está diretamente relacionada com o número e tipo de linguagem OWL permitida. Elas também são compatíveis entre si, onde uma ontologia OWL-Lite é uma ontologia OWL-DL legal e uma ontologia OWL-DL legal é uma ontologia OWL Full legal (WALTON, 2007).

A OWL-Lite suporta restrições simples e classificação hierárquica. Sua vantagem é a de possuir um idioma fácil de entender e aplicar. Já a desvantagem é uma expressividade restrita (ANTONIOU, 2008).

A OWL-DL comparada com a OWL-Lite adiciona suporte completo à negação, disjunção, restrições de cardinalidade, enumerações e restrições de valores. A vantagem desta linguagem é um eficiente apoio ao raciocínio. Enquanto que sua desvantagem é a total perda de compatibilidade com o RDF (ALESSO, 2006).

Considerando que a OWL-Lite e a OWL-DL impõe restrições ao uso do vocabulário e de declarações RDF, a OWL Full não possui tais restrições. Tem a vantagem de ser totalmente compatível com a sintaxe e semântica RDF e sua desvantagem é ser uma linguagem que nem sempre pode chegar a um resultado comprovado (ALESSO, 2006).

A seguir, é mostrada uma ontologia simples que descreve árvores utilizando as linguagens RDF e OWL. A ontologia possui comentários que são escritos usando `rdfs:comment`.

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">

  <owl:Ontology rdf:about="xml:base"/>

  <owl:Class rdf:ID="animal">
    <rdfs:comment>Animals form a class.</rdfs:comment>
  </owl:Class>

  <owl:Class rdf:ID="plant">
    <rdfs:comment>
      Plants form a class disjoint from animals.
    </rdfs:comment>
    <owl:disjointWith rdf:resource="#animal"/>
  </owl:Class>
```

```

<owl:Class rdf:ID="tree">
  <rdfs:comment>Trees are a type of plant.</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#plant"/>
</owl:Class>

<owl:Class rdf:ID="branch">
  <rdfs:comment>Branches are parts of trees.</rdfs:comment>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#is_part_of"/>
      <owl:allValuesFrom rdf:resource="#tree"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:ID="leaf">
  <rdfs:comment>Leaves are parts of branches.</rdfs:comment>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#is_part_of"/>
      <owl:allValuesFrom rdf:resource="#branch"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:TransitiveProperty rdf:ID="is_part_of"/>
</rdf:RDF>

```

Figura 2.6 – Ontologia de uma Árvore. (ANTONIOU, 2008)

2.8 REGRAS PARA A WEB SEMÂNTICA

As ontologias permitem que se possa expressar o conhecimento sobre as classes, suas hierarquias e propriedades. As regras têm uma expressividade complementar: permitem especificar normas de transformação de dados que definam como sintetizar novos fatos a partir de conhecimentos pré-definidos (FENSEL, 2007).

Com as regras, pode-se expressar o conhecimento na forma “se A então B”, onde o lado esquerdo é chamado de *body* ou *antecedents* e o lado direito é chamado de *head* ou *consequents*. Pode-se pensar que o *body* define uma possível situação do sistema, enquanto que o *head* representa uma possível decisão de adaptação (CARDOSO, 2009). Um exemplo que é muitas vezes utilizado para motivar o uso das regras é o “tio” que é representado pela regra: “o irmão dos pais de uma pessoa é o tio da pessoa” (FENSEL, 2007).

$$\text{pessoa}(\text{?x}) \wedge \text{pais}(\text{?x}, \text{?y}) \wedge \text{irmão}(\text{?y}, \text{?z}) \implies \text{tio}(\text{?x}, \text{?z})$$

2.8.1 RIF

O grupo de trabalho (RIF) foi criado pela W3C em 2005, com o intuito de criar um padrão para troca de regras entre os sistemas. RIF focou no desenvolvimento de uma família de linguagens, os dialetos, destinadas a resolver determinados tipos de problemas, pois a complexidade da definição de uma linguagem técnica única para todos os tipos de regras de negócio tornou-se indesejável (POLLOCK, 2009).

A família de dialetos RIF destina-se a ser uniforme e extensível, com sintaxe e semântica rigorosamente especificadas (BOLEY, 2010).

O grupo de trabalho RIF tem centrado seus trabalhos em dois tipos de dialetos: dialetos baseados em lógica e dialetos para regras com ações. Geralmente, os dialetos baseados em lógica incluem alguma linguagem que utiliza algum tipo de lógica. Os dialetos com regras incluem sistemas de produção de regras como Jess, Drools e JRule, bem como regras reativas (evento-condição-ação) tais como RuleML e XChange (BOLEY, 2010).

2.8.2 SWRL

Semantic Web Rule Language (SWRL) foi proposta especialmente para estender os axiomas OWL incluindo as cláusulas de Horn que é uma disjunção de literais com no máximo um literal positivo (ANTONIOU, 2008).

As regras SWRL são compostas de átomos que podem estar nas seguintes formas $C(x)$, $P(x,y)$, $\text{SameAs}(x,y)$ ou $\text{DifferentFrom}(x,y)$, onde C é uma descrição OWL, P é uma propriedade OWL, x e y são variáveis ou indivíduos OWL ou valores de dados OWL (CARDOSO, 2009).

A figura 2.7 ilustra a sintaxe XML para SWRL baseada no RuleML, para isso apresenta a “regra do tio” descrita acima.


```

<ruleml:imp>
  <ruleml:_body>
    <swrlx:individualPropertyAtom swrlx:property="hasParent">
      <ruleml:var>x1</ruleml:var>
      <ruleml:var>x2</ruleml:var>
    </swrlx:individualPropertyAtom>
    <swrlx:individualPropertyAtom swrlx:property="hasBrother">
      <ruleml:var>x2</ruleml:var>
      <ruleml:var>x3</ruleml:var>
    </swrlx:individualPropertyAtom>
  </ruleml:_body>
  <ruleml:_head>
    <swrlx:individualPropertyAtom swrlx:property="hasUncle">
      <ruleml:var>x1</ruleml:var>
      <ruleml:var>x3</ruleml:var>
    </swrlx:individualPropertyAtom>
  </ruleml:_head>
</ruleml:imp>

```

Figura 2.7 – Representação SWRL para a regra ‘tio’ (ALESSO, 2006).

A partir desta regra, se Pedro tem Maria como mãe e Maria tem Paulo como irmão, então Pedro tem Paulo como tio.

2.9 LÓGICA UNIFICADORA

A camada de lógica unificadora da web semântica é ainda vagamente definida. A intenção desta camada é descrever uma lógica matemática formal que represente diferentes modelos (RDF, RDFS, OWL e RIF), fornecendo assim um framework único para possibilitar a combinação dos elementos das camadas inferiores.

Existem softwares, *open source* ou de fornecedores comerciais, que fornecem diversos componentes da web semântica em uma única coleção. Contudo, eles programam sua própria lógica unificadora, cada um faz a unificação de uma forma diferente. Assim, embora a RDF e a OWL permaneçam padrão e portáteis, a aplicação não permanece (POLLOCK, 2009).

2.10 PROVA, CONFIANÇA E CRIPTOGRAFIA

As camadas de prova, confiança e criptografia foram adicionadas à pilha da web semântica para que as pessoas possam acreditar nos resultados apresentados pelas máquinas.

A prova destina-se a fornecer uma maneira matemática de se explicar que as inferências realizadas para gerar uma conclusão ou recomendação são corretas. Contudo,

somente as provas não são suficientes, uma vez que dependem de declarações cuja veracidade não se pode assegurar. Assim, deve ser adicionada a camada de confiança que irá taxar os dados em termos da confiabilidade para que seja possível distinguir os dados bons de dados que provavelmente sejam ruins. Já criptografia irá permitir verificar a autenticidade dos dados (POLLOCK, 2009).

2.11 APLICAÇÕES PARA A WEB SEMÂNTICA

Nas seções anteriores, foram apresentadas tecnologias para a web semântica, agora serão descritas algumas aplicações da web semântica que incluem: serviços web semânticos, busca semântica, ensino à distância, web semântica na bioinformática e comércio eletrônico (ALESSO, 2006).

- Os serviços web semânticos se propõem a automatizar a descoberta, invocação, composição e fiscalização dos serviços web através do uso de agentes, programas que irão coletar o conteúdo de diversas fontes e processar essas informações, podendo trocá-las com outros agentes.
- A busca semântica poderá aumentar e melhorar os resultados das buscas tradicionais por usarem não apenas palavras, mas conceitos e relações lógicas.
- No comércio eletrônico, a web semântica pode ajudar no desenvolvimento de agentes que interpretam as informações dos produtos (frete, comparação de preço), perfis do usuário (para oferecer ao cliente produtos que mais lhes agradam), regras de negócios, dentre outras informações.
- Na bioinformática, a web semântica pode descobrir uma grande quantidade de dados científicos em diferentes formatos e aplicações.
- A web semântica no ensino a distância permite um avanço na criação e disponibilização de cursos e com a adoção de ontologias, a interoperabilidade entre estes cursos e seus usuários poderá ser mais eficaz.
- Existem várias áreas de atuação da web semântica, as que foram apresentadas neste trabalho são apenas algumas delas.

3 USO DE FERRAMENTAS DE BUSCA SEMÂNTICA NA WEB

3.1 INTRODUÇÃO

As páginas web foram construídas principalmente para os olhos humanos, onde as informações inteligíveis para as máquinas são suficientes apenas para dizer como as páginas devem ser exibidas. Com isso, as ferramentas de busca não são capazes de ‘entender’ o significado dos termos e expressões utilizadas, nem as relações entre as páginas, sendo forçadas a realizar pesquisas baseadas em palavras-chave e retornar páginas que possuam tais palavras (YU, 2007).

Atualmente, as ferramentas de busca baseadas em palavras-chave são fundamentais para a recuperação da informação. Contudo, existem diversos problemas associados com seu uso, como exemplos podem ser citados o alto número de páginas recuperadas e a baixa precisão das mesmas, a não obtenção de respostas relevantes à pesquisa ou a não recuperação de páginas importantes e a alta sensibilidade dos resultados ao vocabulário, além desses resultados serem somente páginas web. Mesmo quando a busca é bem sucedida, o usuário é quem deve procurar pelos documentos recuperados a informação que está buscando (ANTONIOU, 2008).

As tecnologias da web semântica oferecem uma maneira para superar estas limitações, uma vez que fornecem a estruturação da informação. Assim, é possível desenvolver ferramentas de busca capazes de examinar, ‘entender’ e conectar as informações, podendo então selecionar somente os conteúdos mais relevantes e apresentá-los em um formato significativo para o usuário (DILLON, 2008).

3.2 BUSCA NA WEB SEMÂNTICA

A busca na web semântica integra as tecnologias da web semântica e ferramentas de busca com o objetivo de melhorar os resultados obtidos em pesquisas baseadas em palavras-chave e evolui para a próxima geração de ferramentas de busca baseadas na web semântica (KASSIM, 2009).

Esse tipo de busca difere da busca convencional por diversas razões. Primeiro, o conteúdo na web semântica está publicado para as máquinas, embora os metadados possam ajudar os usuários a localizar documentos, provavelmente haverá agentes entre os usuários e as ferramentas de busca. Segundo, as ontologias diferem dos textos não estruturados encontrados na maioria das páginas. Terceiro, os documentos na web semântica podem ser uma mistura de fatos concretos, classes, propriedades e restrições lógicas, permitindo que novos fatos sejam produzidos. Por último, a estrutura do grafo dos documentos anotados semanticamente difere da estrutura do grafo de documentos HTML, influenciando tanto na descoberta quanto no *ranking* dos documentos (DING, 2005).

Empresas que tem o foco na internet não esperaram a divulgação dos resultados de pesquisas sobre a busca na web semântica e começaram a implementar algoritmos especializados para atender suas necessidades. A área mais importante, onde se pode ver o desenvolvimento destas novas tecnologias é nas ferramentas de busca semântica que incluem tanto sistemas que tentam aumentar a acurácia da pesquisa, quanto sistemas que tentam mudar profundamente a experiência de busca (HENDLER, 2010).

3.3 UM MODELO ARQUITETURAL PARA FERRAMENTAS DE BUSCA SEMÂNTICA

Diferente das ferramentas de busca tradicionais que se baseiam em índices de palavras, as ferramentas de busca semântica armazenam informações semânticas sobre as páginas web e são capazes de realizar inferências para executar consultas complexas e retornar resultados precisos (YU, 2007).

As ferramentas de busca semântica devem ‘entender’ o que se está perguntando e então apresentar os resultados. Os mesmos resultados devem ser retornados e apresentados na mesma ordem para as consultas que são semanticamente equivalentes (IMIELINSKI, 2009).

Em geral, as ferramentas de busca semântica possuem as seguintes etapas: primeiro, a questão do usuário é interpretada, extraindo-se os conceitos relevantes da sentença. Em seguida, esses conjuntos de conceitos são utilizados para construir uma consulta que é lançada sobre as ontologias. Por último, o resultado é apresentado para o usuário (KASSIM, 2009).

A figura 3.1 apresenta uma arquitetura conceitual para as ferramentas de busca semântica proposta por ILYAS (2004). Cada componente da arquitetura é descrito nas próximas subseções.

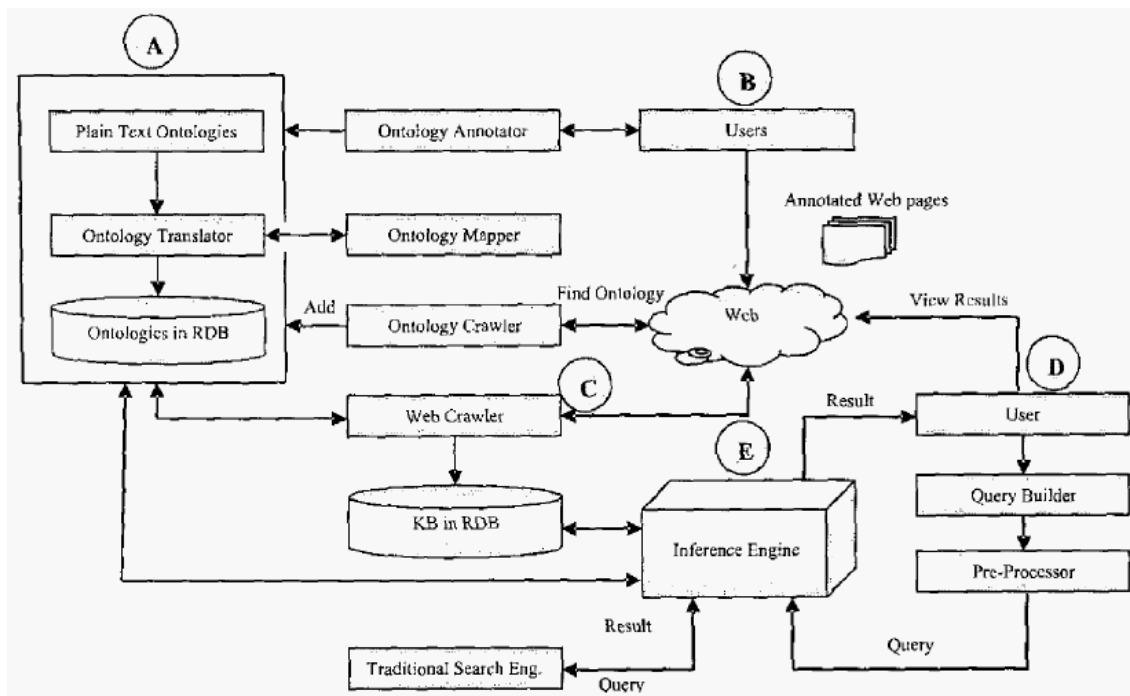


FIGURA 3.1 – Arquitetura para Ferramentas de Busca Semântica (ILYAS, 2004)

3.3.1 Ontology Crawler

O *Ontology Crawler* tem a tarefa de rastrear a web em busca de novas ontologias, devendo ser capaz de lidar com ontologias escritas em diferentes linguagens. As ontologias recuperadas serão traduzidas para as tabelas do banco de dados por meio do *Ontology Translate*, com a ajuda do *Ontology Mapper* que poderá mapear o mesmo conceito em diferentes ontologias. As ontologias armazenadas serão utilizadas para a realização da busca semântica (KASSIM, 2009).

3.3.2 Ontology Annotator

Uma vez que as ontologias foram criadas é necessário anotar as páginas web com esses metadados. Os computadores sabendo como os dados estão relacionados e como essas relações podem ser avaliadas automaticamente, podem realizar operações complexas de filtragem e de pesquisa.

O *Ontology Annotator* deve ser capaz de ler as ontologias do banco de dados ou arquivos de texto e permitir que os usuários anotem suas páginas com esses conhecimentos. O processo é simples quando se está criando uma nova página, mas quando se trata de anotar o conteúdo já existente na web torna-se desafiador, uma vez que não se tem acesso à escrita em páginas criadas por outras pessoas, a solução seria motivar os donos de web sites a anotarem seus conteúdos (ILYAS, 2004).

3.3.3 Web Crawler

O objetivo do *crawler* é pesquisar por documentos anotados semanticamente na web. Assim como na busca tradicional onde os *crawlers* procuram por palavras-chave e constroem a tabela de índices, um *crawler* semântico encontra os conceitos nas páginas anotadas semanticamente e constrói a *knowledge base* (YU, 2007).

Um *crawler* semântico tem que lidar com vários problemas, uma vez que o grafo de documentos anotados semanticamente não é tão denso e bem ligado como o grafo formado por páginas web convencionais. Além disso, muitas das URL's encontradas em documentos anotados semanticamente não apontam para documentos do mesmo tipo. Por isso, é importante o desenvolvimento de heurísticas para limitar o número de links candidatos. Um *crawler* semântico pode também usar ferramentas de busca convencionais para descobrir documentos anotados semanticamente e a partir daí iniciar a navegação (DING, 2005).

3.3.4 Query Builder e Query Pre-processor

Para que o usuário possa realizar pesquisas na web semântica sem a necessidade de conhecer as linguagens de consulta é necessário fornecer ferramentas que construam as consultas necessárias.

O *Query Builder* deve ser capaz de carregar as ontologias do banco de dados e permitir ao usuário construir consultas e selecionar o contexto exato da busca, devendo também fazer análise léxica e dar sugestões para melhorar a busca. Já o *Query Pre-processor* deverá converter as consultas de forma a serem compreensíveis para o *Inference Engine* (ILYAS, 2004).

Se o *Inference Engine* não retornar nenhum resultado para determinada busca, a consulta deve ser enviada para as ferramentas de busca tradicionais. O *Query Pre-processor* para melhorar a consulta poderá adicionar aspas em termos exatos ou “+” para retornar resultados idênticos ao informado, dentre outras melhorias (ILYAS, 2004).

3.3.5 Inference Engine

O *Inference Engine* realiza a pesquisa, através da utilização do banco de dados de ontologias e da *knowledge base*. Os resultados obtidos são enviados ao usuário. Ele é a peça chave do sistema (KASSIM, 2009).

Tendo como exemplo a consulta ‘Encontrar todos os nomes de professores brasileiros que escreveram artigos sobre web semântica no ano de 2010’, existem diferentes conceitos presentes na consulta, nomes, professores, brasileiros, escrever, artigos, web semântica e ano que devem ser considerados pela ferramenta de busca semântica. Assumindo que o banco de dados de ontologias possui todos esses conceitos, o *Inference Engine* irá analisar a consulta para separar os diferentes conceitos. Primeiramente, ele irá saber que nome é uma propriedade de pessoa. Em seguida, irá descobrir que brasileiro é uma pessoa que vive no Brasil que é o nome de um país. Portanto, todas as pessoas que tiverem a propriedade ‘habitante_de’ ou ‘residente_em’ igual a Brasil está qualificada para a pesquisa. ‘Escrever’ é a relação entre ‘professor’ e ‘artigo’, ‘web semântica’ é um tipo de ‘tecnologia_web’ e ‘ano’ é uma medida de tempo. As ontologias descreverão que ‘pessoa’ é ‘escritor_de’ ‘artigos’. Uma vez que o *Inference Engine* conhece os significados e relações entre todos os termos utilizados na consulta, ele é capaz de realizar a busca com maior precisão (ILYAS, 2004).

3.4 FERRAMENTAS DE BUSCA SEMÂNTICA EXISTENTES

3.4.1 TrueKnowledge

Web Site: www.trueknowledge.com

Sede da Empresa: Cambridge, Inglaterra

TrueKnowledge é uma ferramenta de busca onde o usuário pode fazer perguntas e receber respostas diretas para as mesmas, sendo capaz de combinar as informações de sua

knowledge base através de inferências e cruzamento de informações para produzir a resposta correta. Foi lançada em novembro de 2007 e tem como criador William Tunstall-Pedoe (TRUEKNOWLEDGE[1], 2010).

A figura 3.2 apresenta a arquitetura da ferramenta TrueKnowledge, onde a interface com o browser permite que o usuário faça suas perguntas, em seguida o *Natural Language Translation* traduz as questões para a linguagem de consulta TrueKnowledge e a resposta é gerada pelo *Query/Answer System*, com a ajuda da *knowledge base* e do *Knowledge Generator*, responsável por inferir novos fatos (TRUEKNOWLEDGE[2], 2010).

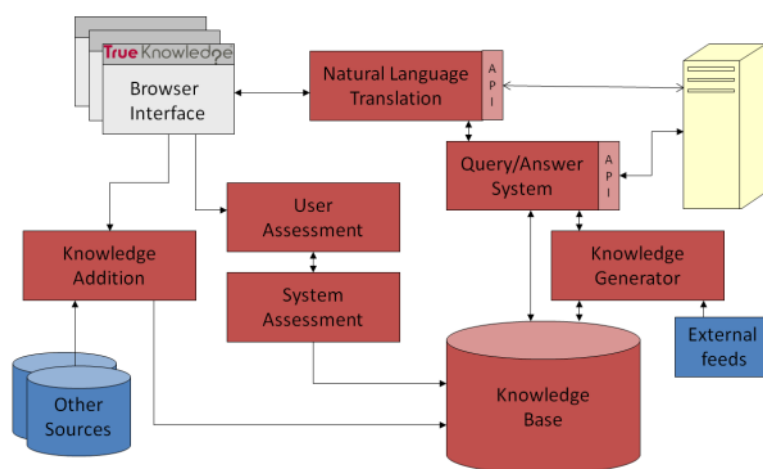


Figura 3.2 - Arquitetura True Knowledge (TRUEKNOWLEDGE [2], 2010)

Sua *knowledge base* possui cerca de 235 milhões de dados estruturados e é construída de duas formas: a partir de dados recuperados da Wikipédia, Free Base, UK Postcodes, Biological taxonomy e pelo web crawler e a partir da submissão de fatos pelos usuários, para garantir a confiabilidade desses fatos os usuários possuem reputação e histórico, além disso, o *System Assessment* verifica se a informação dada não contradiz com outra já existente (OVERELL, 2010).

TrueKnowledge é capaz de responder perguntas complexas, tais como “*Who was US president when Obama was a teenager?*” ou “*In what movies has Joseph Mazzello appeared with Samuel Jackso?*” ou “*What time is it in London now?*”, pesquisa que é mostrada na figura 3.3. Porém, a ferramenta não faz a detecção de palavras ambíguas, com isso pode retornar um significado da palavra que não é a de interesse do usuário.

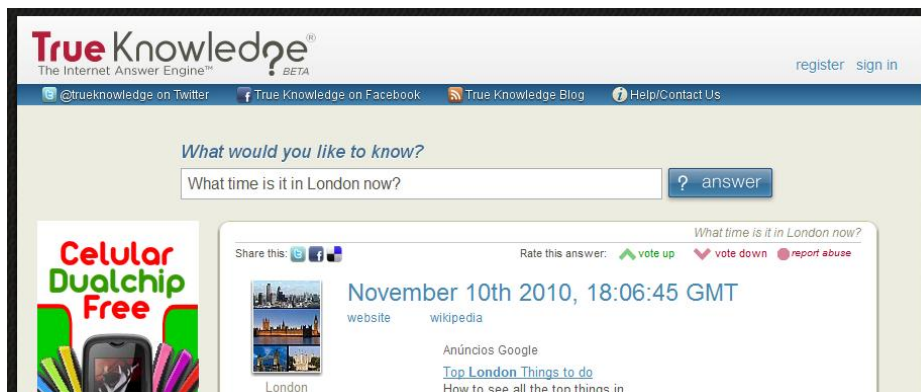


Figura 3.3 – Consulta na Ferramenta TrueKnowledge

3.4.2 Kngine

Web site: www.kngine.com

Sede da Empresa: Kirkland, Estados Unidos

Kngine, abreviação para *knowledge engine*, tem como fundadores Haytham A. Abd EIFadeel e Ashraf A. Abd EIFadeel, foi lançada em fevereiro de 2010. Essa ferramenta é capaz de responder questões dos usuários; realizar comparações entre produtos; pesquisar por informações de tempo real; realizar busca paralela, busca onde é possível navegar por conceitos com características comuns e o usuário pode filtrar e classificar os resultados; além de facilitar a busca fornecendo informações expressivas sobre os diversos conceitos de palavras que possuem significados diferentes. Kngine também está trabalhando em responder questões *fuzzy* através do projeto intitulado *Fuzzy Semantic Search* (LUR, 2010).

A maioria das ferramentas de busca semântica limita sua pesquisa às páginas anotadas semanticamente. Kngine, no entanto, realiza pesquisa em toda a web. Para que isso fosse possível, foi criada uma *knowledge base* com diversas ontologias e seus *web crawlers* percorrem a web tentando compreender os documentos e extrair informações, por meio da utilização da *knowledge base*. Por ultimo, as informações extraídas são adicionas a *knowledge base* que atualmente contém quatro bilhões de pedaços de informações e cerca de dez milhões de conceitos (KNGINE [1], 2010).

A busca na ferramenta Kngine é realizada da seguinte forma: O *Query Layer* analisa a consulta, separando os conceitos. Então, o *Live Objects Identity Indexes* que é um índice invertido, é usado para identificar os conceitos e as palavras que possuam mais de um significado. Por ultimo kngine recupera de sua *knowledge base* as informações sobre a

consulta. Essa recuperação é dependente da análise feita pelo *Query Layer*, por exemplo, uma pesquisa por 'Dan Brown books' irá retornar informações sobre Dan Brown e a lista de livros escritos por ele, enquanto que uma pesquisa por 'Dan Brown' irá retornar apenas informações sobre Dan Brown (KNGINE [2], 2010).

A figura 3.4 apresenta uma pesquisa por 'apple' em Kngine, nela é possível observar a separação dos diferentes significados da palavra pela ferramenta.

The screenshot shows the Kngine search engine interface. At the top, there is a search bar with the word 'Apple' entered. Below the search bar, there are tabs for 'Apple Inc.', 'Apple', and 'Apple (Movie)'. The 'Apple Inc.' tab is selected. On the left side, there is a navigation menu with options like 'Information', 'Share Price', 'Acquisition', 'Softwares', 'Games', and 'Share'. The main content area displays information about Apple Inc., including a logo, a brief description, and various statistics. On the right side, there is a map showing the location of Apple Inc. in Cupertino, California.

Information

Apple Inc.

company, patent assignee, programming language developer, protocol provider, operating system developer, computer manufacturer brand, programming language designer, top level domain sponsor, website category, website owner, software developer, ranked item, venture funded company, cng developer, book subject, employer, author.

Apple Inc., formerly Apple Computer, Inc., is an American multinational corporation with a focus on designing and manufacturing consumer electronics and software products. The company's best-known hardware products include the Macintosh line of personal computers, the iPod line of portable media players, and the iPhone. Apple's software products include the Mac OS X operating system, iTunes media browser, the iLife suite of multimedia and creativity software, and Final Cut Studio, a suite of professional audio- and film-industry software products.

Home Page: apple.com

Industry: Computer software, Computer hardware, Consumer electronics

Slogan: Think Different, The Computer for The Rest of Us

Founders: Ronald Wayne, Steve Wozniak, Steve Jobs

Date Founded: 1 April, 1976

Place Founded: California, Cupertino

Symbol: APC, AAPL, ACP

Market Cap: 170,000,000,000 \$ (2007-12-31)

Revenue: 36,537,000,000 \$ (2009)

Net Income: 5,704,000,000 \$ (2009)

Board Members: Bill Campbell, Steve Jobs, John Sculley, Steve Jobs, Phil Schlein, Al Gore, Gil Amelio, Millard Drexler, Donald T. Valentine, Larry Ellison, Steve Jobs, Jerry York, Arthur D. Levinson, Andrea Jung

Headquarters: 1 Infinite Loop Cupertino, California

Map: Map, Satellite, Híbrido

Figura 3.4 – Consulta na Ferramenta Kngine

Kngine apresenta informações significativas acerca do conceito que se procura, exibindo um texto explicativo e propriedades expressivas do mesmo. Por outro lado, a ferramenta não ajuda o usuário a identificar novas buscas que possam ser mais significativas.

3.4.3 Wolfram|Alpha

Web site: <http://www.wolframalpha.com/>

Sede da Empresa: Champaign, Estados Unidos

Wolfram|Alpha foi lançada em maio de 2009, tendo como criador Stephen Wolfram. Nela, os usuários podem tanto obter respostas às perguntas de determinada área quanto obter resoluções de cálculos matemáticos. Essa ferramenta de busca procura dar informações específicas sobre a consulta e fornecer dados adicionais que o usuário possa estar interessado.

Wolfram|Alpha pode oferecer respostas acerca de tecnologia, culinária, geografia, tempo, viagens, pessoas e músicas. Também é capaz de resolver problemas de sequências numéricas e cálculos, além de responder questões sobre o genoma humano. (ZYGA, 2009).

A ferramenta Wolfram|Alpha contém 10 trilhões de fragmentos de dados, 50 mil tipos de algoritmos e modelos. Foi criada utilizando o Mathematica, software capaz de solucionar diversos cálculos matemáticos e fazendo uso extensivo da web semântica e tecnologias de computação paralela. Especialmente para desambiguação, wolfram|Alpha utiliza informações da Wikipédia (Wolfram|Alpha, 2010).

Para gerar uma resposta, Wolfram|Alpha quebra a consulta em pequenos conceitos, tarefa realizada pelo Mathematica, depois esses conceitos são enviados para os módulos, sistemas que abrangem diferentes temas e, cada um tenta interpretar os dados. A resposta é uma coleção dos dados obtidos por cada módulo (GIBBS, 2010).

A figura 3.5 apresenta uma pesquisa no Wolfram|Alpha pelo próximo eclipse solar total que será visível no Brasil.

The image shows a screenshot of the Wolfram|Alpha search interface. At the top, the Wolfram|Alpha logo is displayed with the tagline 'computational... knowledge engine'. Below the logo is a search bar containing the query 'next total solar eclipse visible in Brazil'. The search results are organized into three sections:

- Input interpretation:** This section shows the query broken down into individual terms: 'next', 'total', 'solar eclipse', 'after today', and 'from Brazil'. Each term is in a separate box, and 'after today' and 'from Brazil' are grouped together.
- Eclipse date:** This section displays the result: '1:41 pm AMT | Saturday, August 12, 2045 (34.78 years from now)'. The date is shown in a light blue box.
- Eclipse path:** This section features a world map with a red line indicating the path of the eclipse. The path starts in the Pacific Ocean, crosses North America, and ends in the Atlantic Ocean. The map is labeled 'Eclipse path:' and includes a 'Show 3D | Orthographic projection' link.

Figura 3.5 – Consulta na Ferramenta Wolfram|Alpha

A ferramenta Wolfram|Alpha apresenta grandes resultados quando solicitada por temas relacionados as ciências, mesmo que a consulta envolva várias ligações complexas entre os termos. Contudo, quando solicitada por temas de outras áreas, mesmo que estes sejam simples, a ferramenta não apresenta o mesmo desempenho.

3.4.4 Duck Duck Go

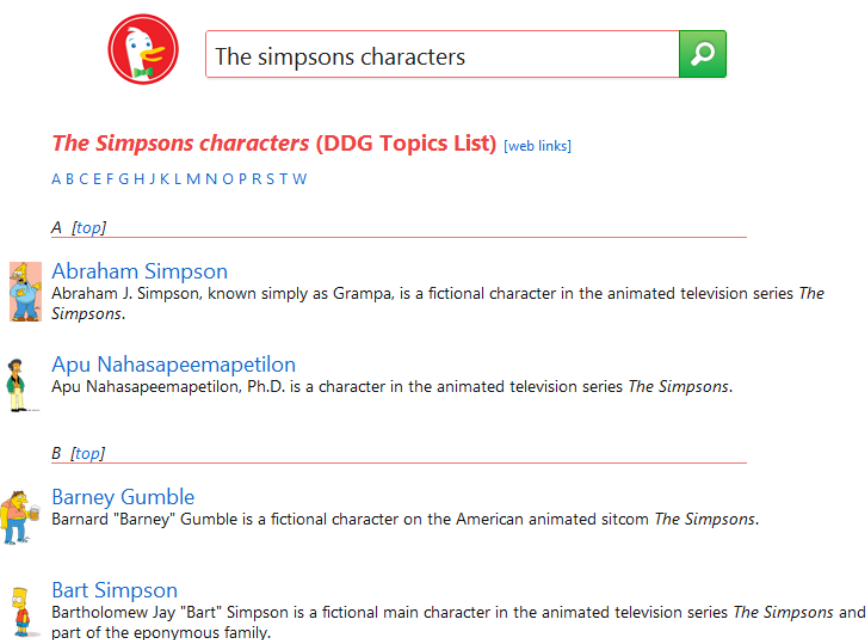
Web site: duckduckgo.com

Sede da Empresa: Valley Forge, Estados Unidos

Duck duck go foi criada por Gabriel Weinberg e lançada em setembro de 2008. Suas principais características incluem: detecção de palavras ambíguas; informação desejada já na primeira página, não sendo necessário navegar pelos links; divisão da página de busca por categorias, agrupa-se os conceitos similares; detecção semântica das palavras, para a recuperação de resultados mais relevantes; política de respeito a privacidade dos usuários, não se armazena históricos nem endereços IPs; e menor recuperação de spans (PANDIA, 2009).

Duck duck go foi construída sobre a API de grandes fornecedores, como o Yahoo! Search Boss que proporciona livre acesso ao Yahoo Search Index e o Worlfram|Alpha. Ela tem a característica de utilizar informações de sites como a Wikipédia, Crunchbase, dentre outros para melhorar a relevância dos resultados, conta também com seu próprio *crawler* e é construída sobre ferramentas open source (WIKIPÉDIA, 2010).

A figura 3.6 apresenta uma pesquisa pelos personagens dos Simpsons, como resultado, a ferramenta apresenta todos os personagens listados em ordem alfabética e com uma breve descrição.



The screenshot shows a search result on the Duck Duck Go website. At the top, there is a search bar with the text "The simpsons characters" and a green search button. Below the search bar, the title "The Simpsons characters (DDG Topics List)" is displayed in red, followed by a link "[web links]". Underneath, there is a list of characters in alphabetical order: A, B, C, E, F, G, H, J, K, L, M, N, O, P, R, S, T, W. The letter "A" is highlighted, and a link "[top]" is provided. The first character listed is "Abraham Simpson", with a small icon of the character and a description: "Abraham J. Simpson, known simply as Grampa, is a fictional character in the animated television series *The Simpsons*." The second character is "Apu Nahasapeemapetilon", with a small icon and a description: "Apu Nahasapeemapetilon, Ph.D. is a character in the animated television series *The Simpsons*." The letter "B" is highlighted, and a link "[top]" is provided. The first character listed under "B" is "Barney Gumble", with a small icon and a description: "Barnard 'Barney' Gumble is a fictional character on the American animated sitcom *The Simpsons*." The second character listed under "B" is "Bart Simpson", with a small icon and a description: "Bartholomew Jay 'Bart' Simpson is a fictional main character in the animated television series *The Simpsons* and part of the eponymous family."

Figura 3.6 - Consulta na Ferramenta Duck duck go

Para auxiliar o usuário encontrar o que deseja, Duck duck go oferece sugestões de buscas. Porém, sua política de zero click não apresenta tantas informações a respeito dos conceitos como se é desejado.

3.4.5 Sig.ma

Web site: <http://sig.ma>

Sede do Instituto: Califórnia, Estados Unidos.

A ferramenta de busca Sig.ma foi desenvolvida pelo DERI, Digital Enterprise Research Institute, com a única finalidade de permitir que as pessoas obtenham a informação de que estão procurando.

O processo de consulta na ferramenta Sig.ma é apresentado na figura 3.7. Primeiro, são identificados os conceitos na consulta do usuário e estes são enviados para diversas ferramentas de busca que retornam uma lista de respostas. As respostas que incorporam metadados são selecionadas e recuperadas da web, os dados que descrevem propriedades diferentes são separados. No processo de *consolidation*, todos os diferentes valores de uma mesma propriedade encontrada são combinados. Por ultimo, os dados são apresentados (TUMMARELLO, 2010).

No *data cache* estão algumas propriedades que descrevem o conceito, elas sempre referenciam outras entidades que podem ser novamente pesquisada na ferramenta. Na figura 3.8, é apresentada uma consulta na ferramenta.

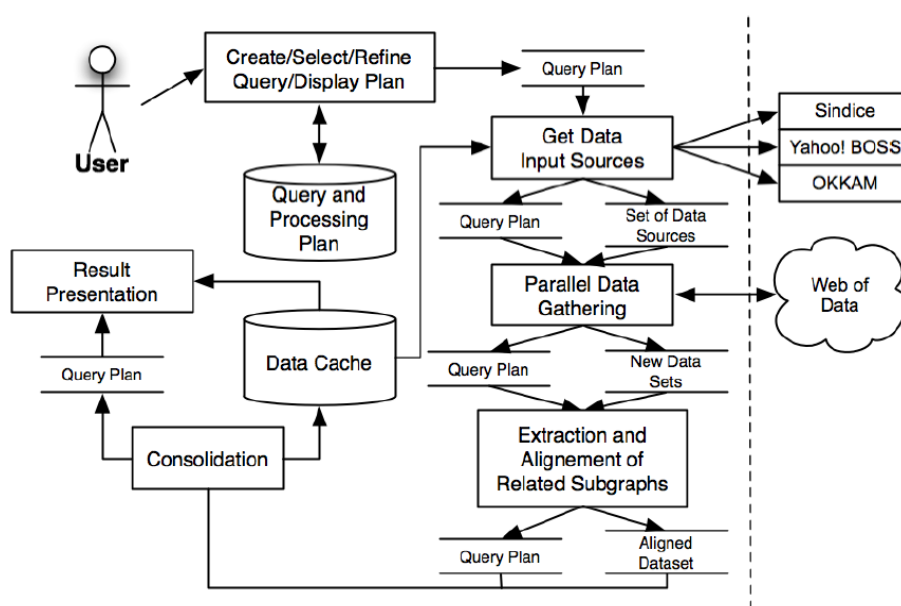


Figura 3.7 – Dataflow da Ferramenta Sig.ma (TUMMARELLO, 2010)

The screenshot shows the Sig.ma search engine interface. The search term is 'Barack Obama'. The main content area displays a profile for Barack Obama with a picture placeholder and a detailed comment. The right sidebar shows a list of 10 sources with their respective URLs and dates. The interface includes navigation buttons like 'Add More Info', 'Start New', and 'Order:: Options Use It'.

Figura 3.8 - Consulta na Ferramenta Sig.ma

Sig.ma apresenta grande quantidade de informações acerca dos conceitos, exibindo textos e algumas propriedades dos mesmos. Contudo, se apresenta mais lenta que as outras ferramentas de busca, uma vez que ainda precisa manipular os dados recuperados em seu processo de pesquisa.

3.4.6 Comparação entre as Ferramentas

O quadro 3.1 apresenta uma comparação entre as ferramentas, revelando algumas das características presentes em cada uma.

Funcionalidade	Duck Duck Go	Engine	Sig.ma	TrueKnowledge	Wolfram Alpha
Perguntas e Respostas	Não	Sim	Não	Sim	Sim
Sugere Novas Buscas	Sim	Não	Não	Não	Não
Resultado já na primeira página	Sim	Sim	Sim	Sim	Sim
Deteção de palavras ambíguas	Sim	Sim	Não	Não	Sim
Busca por palavra-chave quando não conhece o domínio	Sim	Sim	Não	Não	Não

Quadro 3.1 – Quadro Comparativo entre as Ferramentas.

4 ESTUDO DE CASO

4.1 INTRODUÇÃO

As ferramentas de busca semântica devem fornecer informações mais significativas do que as retornadas por ferramentas de busca tradicionais, além de ajudar o usuário a identificar novas buscas que possam ser relacionadas semanticamente com a primeira. Para que isso seja possível é necessário que estas ferramentas tenham conhecimento semântico da área em que a pesquisa está sendo realizada.

Este capítulo propõe um estudo sobre as ferramentas de busca semântica apresentadas no capítulo anterior a fim de verificar quais ferramentas são mais indicadas para determinadas áreas do conhecimento. Em seguida, é apresentado um *front-end* para estas ferramentas que deverá enviar a pesquisa para a ferramenta que obteve melhores resultados na área da consulta.

4.2 PESQUISA SOBRE AS FERRAMENTAS DE BUSCA SEMÂNTICA

Para que as ferramentas de busca semântica possam responder questões sobre determinado domínio, é necessário que possuam ontologias que o represente e que suas bases de conhecimento contenham informações acerca do mesmo. Uma ferramenta de busca semântica só retornará dados significativos para a área do conhecimento que já tenha mapeado.

Com base nisso, este capítulo apresenta uma pesquisa a respeito das áreas de matemática, geografia, história, computação e músicas. A fim de verificar em cada uma dessas áreas qual é a ferramenta, dentre Duck duck go, Kngine, Trueknowledge, Sig.ma e Wolfram|Alpha, mais indicada para se realizar buscas simples.

A fim de inferir quais ferramentas apresentam melhores resultados nas áreas já citadas, foram enviadas duas consultas simples sobre as áreas investigadas para cada uma das ferramentas. A primeira consulta refere-se a um conceito básico e espera-se que as ferramentas apresentem a definição do mesmo. Já a segunda, contém uma pergunta ou algum termo ambíguo e deseja-se que as ferramentas apresentem o significado do termo na área em que se está pesquisando ou ofereçam a resposta, no caso da pergunta.

A ferramenta que melhor definir ou conter um maior número de informações acerca de um termo básico e retornar a resposta da consulta sem a necessidade de se procurar por ela em um texto, será considerada melhor na área em que a pesquisa está sendo realizada. Será também verificado o que faz tal ferramenta mais apropriada para a área investigada.

4.2.1 Análise dos Dados

O apêndice A possui uma tabela mostrando todas as consultas realizadas e uma descrição das respostas apresentadas por cada uma das ferramentas. Com base nele e na forma com que as ferramentas trabalham, será possível deduzir qual é a ferramenta mais indicada para cada área do conhecimento analisada.

Na área de matemática, as ferramentas que apresentaram melhores resultados foram a Wolfram|Alpha e a Duck duck go, já que somente elas apresentaram a resolução do cálculo cinco vezes cinco. Contudo, Wolfram|Alpha será considerada a melhor ferramenta nesta área, uma vez que na busca pela sequência matemática apresentou descrições mais significativas. Wolfram|Alpha é capaz de responder perguntas complexas acerca de matemática por utilizar o software Mathematica que é capaz de solucionar problemas complexos sobre matemática, ciências e engenharias.

Já na área de geografia, as ferramentas Kngine e Duck duck go foram as que ofereceram melhores respostas, visto que as outras ferramentas não apresentaram a definição ou o significado na geografia para ‘*Water Cycle*’ e ‘*Hurricane*’. Como a ferramenta Kngine possui ontologias acerca desta área e seus web crawlers são capazes de recuperar informações em páginas mesmo estas não estando anotadas semanticamente, ela foi capaz de retornar mais informações sobre as consultas do que a ferramenta Duck duck go, sendo então a mais indicada para pesquisas na área de geografia.

Nas pesquisas realizadas por ‘*First world War*’ e ‘*Period of Middle Ages*’ na área de história, as ferramentas Wolfram|alpha, TrueKnowledge e Kngine foram as que exibiram os resultados mais significativos. A ferramenta Duck duck go fugiu do foco da pesquisa ao apresentar a economia na Inglaterra durante a Idade Média para a consulta sobre o período da Idade Média e a ferramenta Sig.ma apresenta somente um texto sobre as duas pesquisas. A ferramenta TrueKnowledge será considerada melhor em pesquisas sobre história, visto que retornou maior quantidade de dados significativos a consulta por possuir

ontologias que descrevem este domínio e por ser capaz de inferir novas informações através do conhecimento armazenado em sua *knowledge base*.

Na pesquisa por ‘*compiler*’ na área da computação, a ferramenta Sig.ma não apresentou resultado para a mesma e TrueKnowledge somente forneceu alguns exemplos de compiladores. Já na pesquisa por ‘*Turing Machine*’ a ferramenta Wolfram|alpha não retornou um conceito para a mesma, somente alguns exemplos de operações. As ferramentas que melhor definiram os conceitos foram Kngine e Duck duck go, contudo Kngine será considerada mais apropriada por ter retornado maior quantidade de dados relacionados a consulta, já que possui ontologias sobre este domínio e é capaz de recuperar informação mesmo de páginas com dados não estruturados.

Nas pesquisas sobre música, somente a ferramenta Trueknowledge retornou resultados significativos para as consultas ‘*She will be loved*’ e ‘*List of songs by Coldplay*’, por possuir ontologias que descrevem este domínio e por fazer buscas no Freebase que possui grande quantidade de dados estruturados sobre músicas e bancas. Trueknowledge então será a mais aconselhada para pesquisas nesta área.

Pelo fato da ferramenta Sigma utilizar ontologias, indexação semântica, inferências lógicas e agregação heurística dos dados e pela observação dos dados retornados nas pesquisas realizadas é possível dizer que para a descrição de entidades esta ferramenta oferece ricas informações.

Todas as ferramentas analisadas cumprem bem o seu papel na busca semântica. A escolha de uma ou de outra se deu principalmente em relação aos diferenciais de cada uma sobre a quantidade e qualidade das informações exibidas.

4.3 INTERFACE DE BUSCA PARA AS FERRAMENTAS SEMÂNTICA

Como a pesquisa realizada apresentou quais são as ferramentas que melhor retornam informações nas áreas analisadas, é possível então construir um *front-end* para tais ferramentas que a partir do conhecimento da área da consulta a envie para a ferramenta mais indicada.

A motivação para o desenvolvimento deste aplicativo foi ajudar na obtenção de resultados mais significativos acerca das buscas realizadas nas áreas investigadas pela pesquisa.

4.3.1 Conexão Com as API's

Para a recuperação das informações a partir das ferramentas de busca, é necessário que estas forneçam uma API que receba as consultas e retorne os resultados a serem exibidos. Contudo, a ferramenta Kngine ainda não possui tal API e a API da ferramenta Wolfram|alpha é paga, mesmo para estudantes e pessoas que a queiram testar. Por isso, para as áreas em que estas ferramentas foram consideradas melhores, a busca será feita em cima da segunda melhor ferramenta.

Observando-se os resultados da pesquisa e levando em consideração a restrição descrita acima é possível dizer que para as áreas de matemática, geografia e computação as buscas serão enviadas para a ferramenta Duck duck go, já nas áreas de história e músicas as pesquisas serão mandadas para a ferramenta TrueKnowledge. E quando for necessário a obtenção da descrição de um conceito, a busca será remetida a ferramenta Sig.ma.

A requisição das consultas é realizada através do protocolo HTTP nas três ferramentas. Já o retorno da consulta nas ferramentas Duck duck go e TrueKnowledge é no formato XML, enquanto que na ferramenta Sig.ma é no formato Json. Os anexos A, B e C apresentam um exemplo de resposta das ferramentas Duck duck go, TrueKnowledge e Sig.ma, respectivamente.

4.3.2 Arquitetura da Aplicação

A aplicação foi desenvolvida utilizando-se a tecnologia Java para Web, o servidor Tomcat e as API's das ferramentas Duck duck go, TrueKnowledge e Sig.ma.

A arquitetura da aplicação é mostrada na figura 4.1. A camada visão possui os JSP's responsáveis pela apresentação dos dados. Alguns desses JSP's possuem a definição de quais dados retornados pelas API's devem ser apresentados e então enviam, para a camada modelo, a solicitação destes dados. Na camada modelo estão as classes que são responsáveis por conectar a API das ferramentas e guardar a resposta recebida, quando solicitadas por alguma informação devem então recuperá-la.

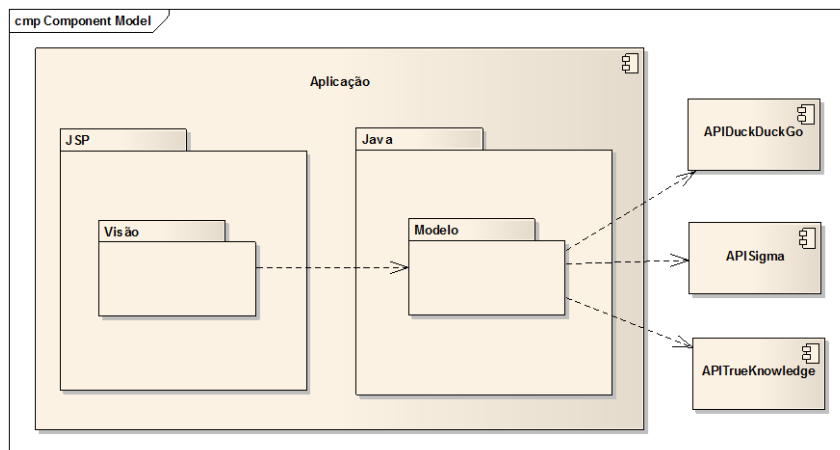


Figura 4.1 – Arquitetura da Aplicação.

Para a recuperação dos dados nos XMLs de resposta, foi utilizado a API JAXP que permite a realização de consultas XPath, além de permitir a utilização de prefixos *namespaces* nestas mesmas consultas, caso que é necessário para os arquivos retornados pela ferramenta TrueKnowledge que os utilizam. Um exemplo do emprego da API pode ser visto na figura 4.2 que apresenta o método consultar da classe TrueKnowledge. Este método recebe qual é o caminho do dado a ser retornado e define qual é o *namespace* que deve ser utilizado nas consultas XPath. Por ultimo, faz a consulta no XML retornado. O método também lança a exceção *MonografiaException* a fim de informar melhor o usuário caso ocorra algum erro na consulta.

```

public String consultar(String caminhoConsulta) throws MonografiaException {
    String resultado = "";
    try {
        TrueKnowledgeNamespaceContext tkNamespace = new TrueKnowledgeNamespaceContext();
        this.getXpath().setNamespaceContext(tkNamespace);

        XPathExpression expression = this.getXpath().compile(caminhoConsulta);
        resultado = expression.evaluate(this.getResultadoPesquisa());
    } catch (XPathExpressionException e) {
        e.printStackTrace();
        throw new MonografiaException(MonografiaException.ERRO_XPATH, e);
    }

    return resultado;
}

```

Figura 4.2 – Método Responsável por Recuperar as Informações dos XMLs de Resposta.

Um exemplo de consulta sobre um caminho XPath pode ser visto na figura 4.3 que apresenta a recuperação do texto entre as tags 'tk:text_result', filha de 'tk:response'. Em seguida, o resultado é apresentado em uma div.

```
String textResult = trueKnowledge.consultar("//tk:response/tk:text_result/text()");
if(textResult != null && !textResult.equals("")){
    resultFound = true; <=>
    <div class="divResultTK">
        <=>textResult <=>
    </div>
<=> }

```

Figura 4.3 – Consulta Xpath no XML de TrueKnowledge.

A recuperação dos dados nos Jsons retornados pela ferramenta Sig.ma foi realizada através de *parsing* de strings, uma vez que o arquivo Json retornado quase sempre não está completo, o que dificultou a manipulação dos arquivos por bibliotecas como Json-lib que poderia, por exemplo, transformar o arquivo em XML para que ele pudesse ser manipulado da mesma forma que as respostas das outras ferramentas.

4.3.3 Consultas no Front-End

O sistema proposto foi nomeado como TDS, iniciais das três ferramentas que ele utiliza, TrueKnowledge, Duck duck go e Sig.ma. Nesta seção, será demonstrado seu funcionamento através da apresentação de algumas telas.

A figura 4.4 apresenta a tela inicial do sistema, onde o usuário poderá digitar sua consulta e selecionar qual é a área da pesquisa. Em seguida, o programa enviará essa consulta para a ferramenta mais indicada, selecionará os dados do arquivo de resposta e por ultimo apresentará o resultado.



Figura 4.4 – Tela Inicial da Aplicação

A figura 4.5 apresenta uma pesquisa por ‘*water cycle*’ que foi enviada para a ferramenta Duck duck go, uma vez que foi a ferramenta mais indicada para pesquisas na área de geografia.

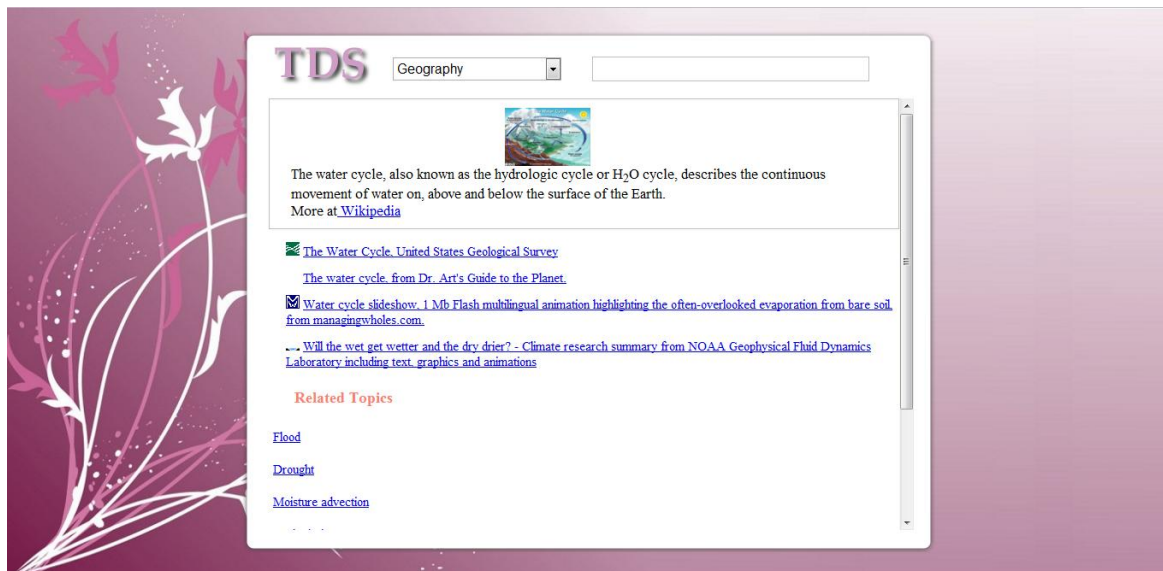


Figura 4.5 – Pesquisa por *Water Cycle*

5 CONCLUSÃO

Neste trabalho, foi realizada uma pesquisa sobre web semântica, que propõe uma estrutura para organizar melhor o conhecimento disperso na internet e sua utilização pelas ferramentas de busca que procuram com isso melhorar os resultados das consultas, apresentando-se como uma alternativa para as ferramentas de busca tradicionais.

Com a chegada da web semântica, as ferramentas de buscas puderam melhorar a recuperação da informação de maneira significativa. Quando as ontologias disponíveis não são aplicáveis ou quando as ferramentas não conseguem agregar valor a busca, algumas destas ferramentas ficam parecidas com as tradicionais, outras não retornam nenhum resultado para a consulta. Mas, quando isso não ocorre os resultados apresentados são muito atraentes, onde o usuário não precisa navegar pelas páginas para obter a informação e a desambiguação de palavras evita a obtenção de resultados desnecessários.

O fato das ferramentas de busca semântica não disponibilizarem descrições mais detalhadas acerca de suas arquiteturas e não informar qual o conteúdo de suas bases de conhecimento, quais ontologias possuem ou em que área do conhecimento possuem mais ontologias, foi uma das dificuldades encontrada para o desenvolvimento das sessões 3.4 e 4.2.

O protótipo criado para as ferramentas de busca semântica procura auxiliar o usuário na obtenção de dados mais relevantes sobre as pesquisas realizadas. Por meio de sua utilização, foram verificadas algumas outras características interessantes que poderiam ser desenvolvidas em trabalhos futuros para a sua melhoria, como a abrangência de mais áreas do conhecimento por parte do protótipo e a detecção automática da área da pesquisa, através da análise da consulta, tornando mais fácil a busca na ferramenta e garantindo que a consulta realmente seja enviada para a ferramenta mais adequada.

REFERÊNCIAS

ALESSO, H. PETER; SMITH, CRAIG F. **Thinking on the Web: Berners-Lee, Godel and Turing**. Canada: Wiley-Interscience, 2006. 261p.

ANTONIOU, GRIGORIS; HARMELEN, FRANK VAN. **A Semantic Web Primer**. 2. ed. Estados Unidos: MIT Press, 2008. 264p.

BOLEY, HAROLD; KIFER MICHAEL. **Rif Basic Logic Dialect**. Disponível em: <http://www.w3.org/TR/2010/REC-rif-bld-20100622/>. Acessado em: setembro de 2010.

CARDOSO, JORGE; LITRAS, MILTIADIS. **Semantic Web Engineering in the Knowledge Society**. Estados Unidos: Information Science Reference, 2009. 425p.

DAVIES, JOHN; STUDER, RUDI; WARREN, PAUL. **Semantic Web Technologies: Trends and Research in Ontology-based Systems**. Inglaterra: Wile, 2006. 312p.

DILLON, THARAM et al. **Advances in Web Semantics I: Ontologies, Web Services and Applied Semantic Web**. 1 ed. Alemanha: Springer, 2008. 397p.

DING, LI et al. **Search on the Semantic Web**. IEEE Computer Society, outubro de 2005.

DODDS, LEIGH. **Introducing SPARQL: Querying the Semantic Web**. Disponível em: <http://www.xml.com/pub/a/2005/11/16/introducing-sparql-querying-semantic-web-tutorial.html>. Acessado em: setembro de 2010.

FENSEL, DIETER. **Enabling Semantic Web Services: The Web Service Modeling Ontology**. Alemanha: Springer, 2007. 188p.

GIBBS, MARK. **Wolfram|Alpha revisited and no longer boring**, maio de 2010. Disponível em: <http://www.networkworld.com/newsletters/web/2010/052410web1.html>. Acessado em: outubro de 2010.

HENDLER, JAMES. **Web 3.0: The Dawn of Semantic Search**. IEEE Computer Society, janeiro de 2010.

ILYAS, QAZI M; KAI, YANG Z; TALIB, MUHAMMAD A. **A Conceptual Architecture for Semantic Search Engine**. IEEE Computer Society, 2004.

IMIELINSKI, TOMASZ; SIGNORINI, ALESSIO. **If you ask nicely, I will answer: Semantic Search and Today's Search Engines**. IEEE Computer Society, 2009.

JARDIM, ANDRÉ DESESSARDS. **Introdução à Web Semântica**. Disponível em: <http://ia.ucpel.tche.br/~lpalazzo/Aulas/IWS/m01/Intro.pdf>. Acessado em: setembro de 2010.

KASSIM, JUNAIDAH M, RAHMANY MAHATHIR. **Introduction to Semantic Search Engine**. IEEE Computer Society, agosto de 2009.

KNGINE[1]. **Insight About Kngine Architecture**, março de 2010. Disponível em: <http://kngine.tumblr.com/post/464652794/insight-about-kngine-architecture>. Acessado em: outubro de 2010.

KNGINE[2]. **Insight About Kngine Architecture – Part 2**, março de 2010. Disponível em: <http://kngine.tumblr.com/post/475424107/insight-about-kngine-architecture-part-2>. Acessado em: outubro de 2010.

LUR, XAVIER. **Kngine: The Smartest Search Engine Ever?**, abril de 2010. Disponível em: <http://www.techxav.com/2010/04/09/kngine-the-smartest-search-engine-ever/>. Acessado em outubro de 2010.

OVERELL, SIMON; KEYNES, MILTON. **True Knowledge Using AI to Get Answers From the Internet**, abril de 2010. Disponível em: http://dces.essex.ac.uk/staff/udo/ecir2010/slides/ECIR_Industry_Day_2010_Overell.pdf. Acessado em: outubro de 2010.

PANDIA. **New Search Engine Duck Duck Go Has Less Spam and More Content**, abril de 2009. Disponível em: <http://www.pandia.com/sew/1596-duck-duck-go.html>. Acessado em: outubro de 2010.

POLLOCK, JEFFREY T. **Semantic Web for Dummies**. Wiley-Interscience. Canada: 2009. 394p.

SEGARAN, TOBY; EVANS, COLIN; TAYLOR, JAMIE. **Programming the Semantic Web**. 1.ed. Estados Unidos: O'Reilly, 2009. 279p.

SEMANTIC Web. 2010. Disponível em: http://semanticweb.org/wiki/Semantic_Web. Acessado em: setembro de 2010.

SEMANTIC Web Standards. 2009. Disponível em: http://semanticweb.org/wiki/Semantic_Web_standards. Acessado em: setembro de 2010.

TRUEKNOWLEDGE[1]. **Technology**, 2010. Disponível em: <http://www.trueknowledge.com/technology/>. Acessado em: outubro de 2010.

TRUEKNOWLEDGE[2]. **Architecture**, 2010. Disponível em: <http://www.trueknowledge.com/architecture/>. Acessado em: outubro de 2010.

TUMMARELLO, GIOVANNI et al. **Sig.ma: live views on the Web of Data**, julho de 2010. Disponível em: <http://fooshed.net/paper/JWS2010.pdf>. Acessado em novembro de 2010.

WALTON, CHRISTOPHER. **Agency and the Semantic Web**. Estados Unidos: Oxford, 2007. 249p.

WIKIPÉDIA. **Duck Duck Go**, outubro de 2010. Disponível em: http://pt.wikipedia.org/wiki/Duck_Duck_Go. Acessado em: outubro de 2010.

Wolfram|Alpha. **About Wolfram|Alpha**, 2010. Disponível em: <http://m.wolframalpha.com/about.html>. Acessado em: outubro de 2010.

YU, LIYANG. **Introduction to the Semantic Web and Semantic Web Services**. 1.ed. Estados Unidos: Chapman and Hall/CRC, 2007. 368p.

ZYGA, LISA. **Wolfram Alpha Could Answer Questions that Google Can't**, março de 2009. Disponível em: <http://www.physorg.com/news155826981.html>. Acessado em: outubro de 2010.

APÊNCICE A – Pesquisas Realizadas nas Ferramentas de Busca Semântica.

Área	Consulta	Duck Duck Go	Knjine	Sig.ma	TrueKnowledge	Worlfram Alpha
Matemática	Sequência Matemática: 1,2,4,8...	Apresenta alguns links que descrevem a série.	Informa que essa é uma série infinita e fornece alguns links que a descreve.	Não apresenta dados significativos para a consulta.	Não apresenta dados sobre a consulta.	Apresenta gráfico, função e uma possível continuação da sequência.
	5*5	Apresenta a resposta à pergunta.	Apresenta informações sobre o cubo mágico e internet explorer 5.	Não apresenta dados significativos para a consulta.	Não apresenta dados sobre a consulta.	Apresenta a resposta à pergunta
Geografia	<i>Water Cycle</i>	Apresenta uma figura sobre o ciclo da água e informa que ele descreve o movimento da água na Terra. E disponibiliza links que o melhor descrevem.	Oferece o mesmo que Duck duck go além de informar que água pode mudar de estados e sua quantidade continua a mesma na superfície da Terra.	Retorna alguns links e figuras sobre o assunto.	Apresenta <i>water cycle</i> como um dos episódios da série de TV Bill Nye the Science Guy.	Não apresenta dados sobre a consulta.
	<i>Hurricane</i>	Apresenta diversos significados da palavra. E seu significado na geografia é descrito como um ciclone tropical que é caracterizado por um sistema de baixa pressão e por trovoadas.	Oferece o mesmo que Duck duck go além de explicar os termos tropical e ciclone.	Apresenta alguns exemplos de <i>hurricanes</i> .	Apresenta <i>hurricane</i> como uma música.	Oferece a possibilidade de se pesquisar por um determinado <i>hurricane</i> .
História	<i>First world war</i>	Informa que a primeira guerra foi um conflito militar centrado na Europa que teve início em 1914. Além de conter links que a descrevem melhor.	Retorna basicamente as mesmas informa que Duck duck go.	Apresenta uma descrição da primeira e da segunda guerra mundial.	Mostra as mesmas informações que Duck duck go, além de informar os aliados e número de mortos em alguns países.	Informa a data, os países e as pessoas envolvidas.
	<i>Period of Middle Ages</i>	Apresenta um texto sobre a economia na Inglaterra durante a Idade Média e alguns links..	Apresenta uma explicação sobre a Idade Média e a resposta para a pergunta.	Apresenta um texto descritivo sobre a Idade Média.	Apresenta a resposta à pergunta.	Apresenta a resposta à pergunta.

Área	Consulta	Duck Duck Go	Engine	Sig.ma	TrueKnowledge	Worlfram Alpha
Computação	<i>Turing Machine</i>	Informa que a máquina é um dispositivo teórico que manipula símbolos de acordo com uma tabela de regras. Além de conter links que a descrevem melhor.	Oferece o mesmo que Duck duck go além de informar que ela foi descrita em 1936 por Alan Turing.	Apresenta um grande texto descritivo sobre o que é a máquina de Turing.	Informa que é uma máquina teórica para a manipulação de símbolos.	Oferece vários exemplos de operações para se fazer na máquina.
	<i>Compiler</i>	Descreve o compilador como sendo um programa que transforma o código escrito em uma linguagem para outra linguagem. E apresenta alguns links relacionados.	Oferece o mesmo que Duck duck go além de apresentar alguns exemplos e outros conceitos para a palavra.	Não apresenta dados sobre a consulta.	Apresenta alguns exemplos de compiladores.	Apresenta o compilador como uma máquina que transforma instruções escritas em uma linguagem de alto nível em instruções assembly.
Músicas	<i>She will be loved</i>	Informa que é uma música da banda Maroon 5 presente no álbum Songs About Jane. Além de apresentar links para a música.	Oferece o mesmo que Duck duck go além de informar que a música foi lançada em 2004 e cita alguns países que ela foi sucesso.	Não apresenta dados relacionados à consulta.	Oferece o mesmo que Duck duck go e informa o autor da música.	Não apresenta dados sobre a consulta.
	<i>List of songs by Coldplay</i>	Apresenta alguns links para páginas que contêm algumas músicas da banda.	Apresenta alguns links para páginas que contêm algumas músicas.	Não apresenta dados sobre a consulta.	Apresenta uma lista de álbuns da banda.	Não apresenta dados sobre a consulta.

Anexo A – Exemplo da Estrutura de arquivos Retornados Pela API da Ferramenta Duck duck go

```
<DuckDuckGoResponse version="0.8">
  <Type>A</Type>
  <Answer>2 + 2 = 4</Answer>
  <Heading>Text</Heading>
  <Image>URL</Image>
  <Abstract>HTML</Abstract>
  <AbstractText>Abstract minus HTML</AbstractText>
  <Results>
    <Result>HTML
      <Icon height="16" width="16">URL</Icon>
      <Text>Result minus HTML</Text>
      <FirstURL>URL</FirstURL>
    </Result>
  </Results>
  <RelatedTopics>
    <RelatedTopic>HTML
      <Icon height="16" width="16">URL</Icon>
      <Text>Result minus HTML</Text>
      <FirstURL>URL</FirstURL>
    </RelatedTopic>
  </RelatedTopics>
</DuckDuckGoResponse>
```

Anexo B – Exemplo da Estrutura dos arquivos Retornados Pela API da Ferramenta TrueKnowledge.

```

<?xml version="1.0" encoding="UTF-8"?>
<tk:response
xmlns:tk=http://www.trueknowledge.com/ns/kengine
xmlns="http://www.w3.org/1999/xhtml" understood="true"
answered="true" type="direct_answer">
  <tk:status>completeness unknown</tk:status>
  <tk:text_result>March 16th 2009, 10:57:30
  CDT</tk:text_result>
  <tk:structured_result>
    <tk:result>
      <tk:object>
        <tk:id>[local timepoint: [timepoint:
        ["2009/3/16/10/57/30"]]; [central daylight
        time]]</tk:id>
      </tk:object>
    </tk:result>
    <tk:question_entities>
      <tk:object>
        <tk:id>[chicago]</tk:id>
        <tk:metadata
        parameter="image128">http://www.truekn
        owledge.com/images/thumbs/128/128/2004
        -07-
        14_2600x1500_chicago_lake_skyline.jpg<
        /tk:metadata>
        <tk:metadata
        parameter="wikipedia">http://en.wikipe
        dia.org/wiki/Chicago</tk:metadata>
      </tk:object>
    </tk:question_entities>
  </tk:structured_result>
  <tk:tk_question_url>http://www.trueknowledge.com/q/what
  's_the_time_in_chicago</tk:tk_question_url>
</tk:response>

```

Anexo C – Exemplo da Estrutura dos arquivos Retornados Pela API da Ferramenta Sig.ma.

```
{
  "sources": [
    {
      "source_uri": "http://strangedarkgypsygirl.com/?tag=obama",
      "source_label": "obama |",
      "domain": "strangedarkgypsygirl.com",
      "date": "2010-08-19",
      "total_values": 3,
      "index": 1
    },
    {
      "source_uri": "http://sterkworks.com/tag/obama/",
      "source_label": "obama",
      "domain": "sterkworks.com",
      "date": "2009-09-05",
      "total_values": 3,
      "index": 2
    },
    {
      "source_uri": "http://en.wikipedia.org/wiki/Barack_Obama",
      "source_label": "Barack Obama - Wikipedia, the free encyclopedia",
      "domain": "en.wikipedia.org",
      "date": "2010-09-28",
      "total_values": 7,
      "index": 3
    },
    {
      "source_uri": "http://ifuwant.cn/tag/obama/",
      "source_label": "Obama",
      "domain": "ifuwant.cn",
      "date": "2010-02-12",
      "total_values": 3,
      "index": 4
    },
    {
      "source_uri": "http://adrianmonck.com/tag/obama/",
      "source_label": "Obama",
      "domain": "adrianmonck.com",
      "date": "2010-03-29",
      "total_values": 3,
      "index": 5
    },
    {
      "source_uri": "http://www.whitehouse.gov/administration/president-obama",
      "source_label": "President Barack Obama | The White House",
      "domain": "www.whitehouse.gov",
      "date": "2010-10-04",
      "total_values": 11,
      "index": 6
    },
    {
      "source_uri": "http://intelligentpestsolutions.com/tag/obama/",
      "source_label": "Obama |",
      "domain": "intelligentpestsolutions.com",
      "date": "2010-08-19",
      "total_values": 3,
      "index": 7
    },
    {
      "source_uri": "http://globpt.com/tag/obama/",
      "source_label": "Obama",
      "domain": "globpt.com",
      "date": "2010-08-19",
      "total_values": 3,
      "index": 8
    }
  ]
}
```