



**EIXO TEMÁTICO:**  
**Organização e Representação da Informação e do Conhecimento**

## **OS MECANISMOS DE BUSCA NA PERSPECTIVA DA WEB SEMÂNTICA.**

### **THE RESEARCH MECHANISMS IN THE PERSPECTIVE SEMANTIC WEB**

Diana Vilas Boas Souto Aleixo - dianavbsouto@gmail.com

Silvana Drumond Monteiro - silvanadrumond@gmail.com

**Resumo:** Com os avanços ocorridos na *Web*, observa-se o aprimoramento de novos mecanismos relacionados à busca. No entanto, o debate a respeito do uso dessas ferramentas, da *Web Semântica* na Ciência da informação ainda é recente, e requer o desenvolvimento de pesquisas que explorem suas características e implicações na organização e recuperação da informação no contexto da *Web*. Este trabalho aborda os mecanismos de busca *Swoogle* e o *Watson*, evidenciando suas particularidades, aproximações e diferenças quanto à estrutura, funcionalidade e normalização. Por meio desta breve discussão, são expostos os pontos que nortearam o desenvolvimento dessas ferramentas, debatendo sobre temática e suas contribuições à Ciência da Informação.

**Palavras-chave:** Mecanismo de busca. Web Semântica. Swoogle. Watson. Semantic Web Search.

**Abstract:** With the advances in the *Web*, there is the improvement of new search-related mechanisms. However, the debate about the use of these tools, the *Semantic Web* in Information Science, is still new and requires the development of research to explore its characteristics and implications for the organization and retrieval of information in the *Web* context. This work addresses *Swoogle* and *Watson* search engines, highlighting their particularities, similarities and differences in the structure, functionality and standardization. Through this brief discussion, the points that guided the development of these tools are shown, debating on the theme and its contributions to the Information Science.

**Keywords:** Research Mechanism. Semantic Web. Swoogle. Watson. Semantic Web Search.

## **1 INTRODUÇÃO**

A *Web Semântica* (WS) é um universo paralelo à *Web* de documentos em *HyperText Markup Language* (HTML), onde os documentos *online* são descritos por meio dos padrões *Resource Description Framework* (RDF) ou *Web Ontology Language* (OWL), sendo caracterizados por uma anotação semântica e referências significativas que os diferenciam dos demais documentos descritos na *web*.

Atualmente, a WS toma corpo, aparecendo de maneira mais visível e robusta

por meio dos mecanismos de busca. Com a adoção de agentes inteligentes, os mecanismos de busca leem e/ou capturam informações que se ligam semanticamente, como é o caso do *Knowledge Graph* do Google.

Desde o momento em que foram desenvolvidos, os mecanismos de busca baseados nos padrões propostos pela WS permitem a compreensão entre os leitores humanos e por agentes de *software* e serviços (as máquinas).

Nesse sentido, o *World Wide Web Consortium* (W3C) tem como meta propiciar uma organização da informação na *web*, por meio de recomendações, tecnologias e padrões. Como exemplo, cita-se a *Semantic Web Education and Outreach* (SWEO), uma comunidade que faz parte do W3C, a qual tem como foco estudar a vinculação de dados abertos na WS, sendo responsável por vários projetos e estudos relacionados à elaboração de mecanismos de busca na WS.

Este trabalho apresenta uma visão geral a respeito dos mecanismos de busca desenvolvidos sob a perspectiva da WS. Para isso, com apoio da literatura e das recomendações do W3C, buscou-se identificar as particularidades, aproximações e diferenças existentes entre os mecanismos de busca *Swoogle* e o *Watson*, ambos desenvolvidos conforme as recomendações do W3C.

O trabalho relaciona inicialmente a Ciência da Informação (CI), sob a perspectiva de Vannevar Bush, para evidenciar um dos paradigmas possíveis para a Ciência da Informação que atendem o objetivo em tela, isto é, a Tecnologia da Informação e Comunicação (TIC) e, por conseguinte, os Mecanismos de busca, que têm sido basilares na constituição dos índices contemporâneos no ciberespaço. Pretende-se, com este artigo, trazer ao debate um tema ainda recente e suas contribuições no universo da Organização da Informação (OI), enquanto linha de pesquisa voltada à CI.

## **2 A CIÊNCIA DA INFORMAÇÃO EM FOCO: VANNEVAR BUSH**

Em uma perspectiva americana, Vannevar Bush é responsável por um dos marcos iniciais da CI (MEDEIROS; VANTI, 2011; SILVA; FREIRE, 2012). De acordo com Barreto (2002), Bush inicia uma revolução na forma de armazenar, recuperar e disseminar a informação, por meio de seus estudos que envolvem a aspiração de um mecanismo capaz de auxiliar o gerenciamento das informações.

Vannevar Bush foi engenheiro elétrico e doutor pelo *Massachusetts Institute*

of *Technology* (MIT) e pela *Harvard*. (MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2013). Conforme Medeiros e Vanti (2011), Bush foi um pesquisador que possibilitou a criação de vários equipamentos, tais como: dispositivo para levantamento topográfico de terras; detector de submarinos por meio de campos magnéticos; criou um aparelho considerado o precursor dos computadores analógicos, e que propiciou sua inserção na CI, o *Memory extensive* (Memex).

As críticas de Bush (1945) aos sistemas de organização da informação se estendiam a pontos como o atraso no acesso à informação; as regras incômodas e complicadas; as classes e subclasses; a incapacidade de acessar um documento; a repetição no processo de busca; e a artificialidade dos sistemas de indexação, crítica aos sistemas de ordem alfabética e ordem numérica, pois a mente humana opera por associações. Diante dessas preocupações, o autor comparou um sistema de organização com a mente humana, Bush relaciona que “Quando um elemento está ao seu alcance, salta instantaneamente para o seguinte, que é sugerido pela associação de pensamentos segundo uma intrincada rede de atalhos contida nas células do cérebro [...]” (BUSH, 1945, p.[9], Tradução nossa).

No artigo "*As We May Think*", no *Atlantic Monthly* em 1945, Bush descreve o *Memex* como um dispositivo no qual livros, registros e comunicações podem ser consultados de forma rápida e flexível. (MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2013).

[...] um memex é um aparelho no qual um indivíduo guarda todos os seus livros, registros e comunicações, e que é mecanizado de forma a poder ser consultado com grande velocidade e flexibilidade. Na verdade seria um suplemento ampliado e íntimo de sua memória. (BUSH, 1945, p. [10] Tradução nossa).

No *Memex*, o armazenamento e a recuperação das informações seriam possíveis por meio da adoção do microfilme, onde em uma mesa seria feita a visualização nas telas, haveria um teclado com botões e alavancas de seleção, e um local para armazenamento do microfilme. A máquina aumentaria a memória humana, permitindo que o leitor fizesse ligações, ou "trilhas associativas" (como o autor denominava), entre os documentos (MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2013).

No que tange aos mecanismos de busca, pode-se inferir sua fala a respeito do que mais tarde receberia a denominação de Hipertexto “[...] consiste em possibilitar que cada um dos elementos selecione ou busque outro elemento

automaticamente e imediatamente.” (BUSH, 1945, p. [11], Tradução nossa). Medeiros e Vanti (2011) consideram o artigo de Bush como parte de uma rede enunciativa de um novo campo em formação: a CI. As bases conceituais sobre a indexação, sistema de organização da informação e intertextos ou ligação entre eles estavam lançadas à época, e tiveram sucessores mais contemporâneos. Suas ideias são tão atuais, que quase 45 anos depois, um físico inglês, Tim Berners-Lee, deu vida e forma à ideia de Bush, e criou a linguagem de programação *Hyper Text Mark up language* (HTML) e os *hyperlinks*.

A associação de ideias organizada por Bush é contemplada na WS, a qual estabelece em sua proposta uma recuperação relativamente mais significativa por meio da associação entre conceitos a compreendidos pelos agentes inteligentes (ALVES *et al.*, 2007).

### 3 MECANISMOS DE BUSCA

Os mecanismos de busca são sistemas especializados para recuperar conteúdo na *Web*. Eles coletam informações nos *sites* e as armazenam em seus bancos de dados, os quais estão disponíveis para consulta, por meio da geração dos índices, mediante consulta, possibilitando a busca e a localização dos conteúdos no ciberespaço.

Conforme Cendón (2001), os mecanismos de busca surgiram à medida que a *Web* atingiu proporções na qual seus recursos não eram capazes de garantir a coleta e a busca de informações. O primeiro mecanismo de busca baseado na *Web* foi criado por Matthew Gray do *Massachusetts Institute of Technology* em meados de 1993, o “*www.wanderer*”. A seguir em 1994, Brian Pinkerton desenvolveu o *Webcrawler*, que “[...] foi a primeira ferramenta de indexar todo o texto dos documentos na *Web*.” (MONTEIRO, 2008, p. 107). Os motores de busca são compostos por componentes listados previamente, mediante critérios de indexação e inclusão nas bases de dados. (CENDÓN, 2001). Monteiro (2009) explica que esses operam através de três processos, sendo:

- o *crawler* que coleta todas as páginas possíveis na *web*;
- o índice, a partir da base de dados gerada pelo *crawler*;
- a interface de busca do usuário. (MONTEIRO, 2008, p. 107).

Para Dziekaniak (2002), os mecanismos de busca são constituídos por três

componentes principais. O primeiro seria o robô<sup>1</sup>, que percorre os *sites* e cria uma cópia do texto, a qual irá compor a base de dados. Essa base de dados (segundo componente) é armazenada no computador. Após, o terceiro componente realiza a pesquisa e percorre os endereços das páginas em busca das palavras ou expressões informadas pelo usuário para consulta, e em seguida os resultados são apresentados.

Em suma, os mecanismos realizam a busca baseados no grau de relevância, em critérios definidos pelos programadores e também por estratégias definidas pelo sujeito no momento da busca. Buscando outras definições expressivas para essas máquinas de busca, destacam-se particularmente as infracitadas:

[...] os índices são povoados com etiquetas, outro tipo de metadados – dados a respeito de dados. (BATTELLE, 2006, p. 18);

Metáfora dos oráculos pós-modernos para os mecanismos. (SOARES, 2008, *online*);

O mecanismo de busca é o ponto focal da experiência de busca online. (NIELSEN *apud* FRAGOSO, 2006, *online*).

Em 2012, Steven Levy em seu livro sobre a biografia do Google, relata o trabalho que os fundadores e os engenheiros estão desenvolvendo, especialmente no desenvolvimento de tecnologias para a recuperação da informação e o aprendizado da semântica que os sujeitos pesquisadores deixam em seus históricos de busca (rastros indiciais), de forma que afirma sobre o Google:

[...] uma máquina; um colecionador onívoro de informações; um cofre hiperenciclopédico do conhecimento humano; um leiloeiro infalível; um estudante de línguas, comportamentos e desejos assustadoramente hábil. (LEVY, 2012, p. 261).

Ou, ainda, o Google é “[...] uma gigantesca máquina de aprendizagem baseada em inteligência artificial.” (LEVY, 2012, p. 15). Em suma, a semântica, apesar de sua estrutura de representação no ciberespaço, torna-se visível e funcional para os sujeitos por meio da busca, por isso os mecanismos de busca com seus índices contemporâneos são tão importantes para o estudo da própria WS.

#### 4 WEB SEMÂNTICA

A ideia da WS nasceu a partir da união de vários campos de estudos

---

<sup>1</sup> Também denominado *robot*, *spider* (aranha), *crawler* (rastejadores), *wanderers* (viajantes), *knowbot*, *worm* (verme), *web-bot*, Dziekaniak (2002).

(inteligência artificial, o hipertexto, os desenvolvedores da *Web*). O ponto central para levantar tal questão é considerar que tudo que pode ser pesquisado encontra-se disponível para ser organizado e recuperado na *Web*. Entretanto, como explicam os pesquisadores Ding *et al.* (2005), retirar significado de um texto não é uma atividade muito fácil para programas de computadores. Assim, a WS é uma abordagem para a codificação e publicação de informações de uma forma que torna mais fácil para computadores a compreensão, através do uso das ontologias<sup>2</sup>.

A origem da WS se relaciona ao artigo da revista *Scientific American*, de 2001, escrito por Tim Berners-Lee, e como coautores James Hendler e Ora Lassila. Nesse artigo, os autores esboçam uma *Web* de estrutura comum, onde os aplicativos compartilham seus dados “linkados” com outros aplicativos. A WS dá às pessoas a capacidade de criarem repositórios de dados na *Web*, de forma a facilitar a troca de informação (WORLD WIDE WEB CONSORTIUM, 2008).

De acordo com Catarino e Baptista (2002), o princípio que norteia a WS é simples, e requer a utilização de linguagens e formatos comuns para a descrição. Nesse sentido, a WS fornece um enquadramento comum que permite o compartilhamento dos dados.

Uma das vantagens da WS é que as pessoas podem criar de forma colaborativa ontologias e construir vocabulário comum, sem controle centralizado. Um bloco de construção de ontologias da WS compõe um termo em WS<sup>3</sup> que desempenha o papel de uma palavra em línguas naturais. Conforme Li Ding *et al* (2005) a WS depende de três ontologias de domínio, descritas a seguir.

Os padrões ou ontologias de domínio que fazem parte da WS são o *Resource Description Framework* (RDF), o protocolo SPARQL e a *Ontology Web Language* (OWL). O RDF é um método geral para troca de dados na *Web*, que permite o compartilhamento e a mistura de dados estruturados e semiestruturados por diversos aplicativos. O SPARQL foi desenvolvido para consultar dados em diferentes sistemas. O OWL permite que o usuário defina conceitos de forma que possam ser misturados e combinados com outros conceitos de variados usos e aplicativos, de

---

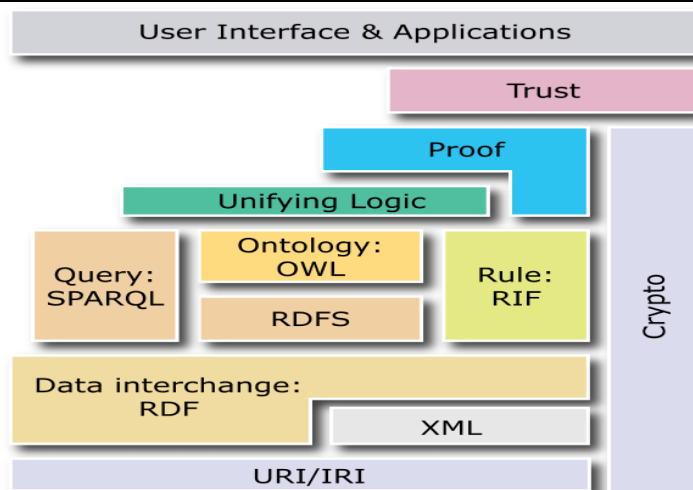
<sup>2</sup> Uma ontologia é uma especificação explícita e formal de uma conceituação compartilhada (STUDER; BENJAMINS; FENSEL; 1998; GRUBER, 1996).

<sup>3</sup> “[...] é um recurso RDF que representa uma instância de RDFS: *Class* ou RDF: *Property* e pode ser referenciado por sua referência URI (URIfref).” (DING, 2005, p. 2, Tradução nossa).

acordo com o tutorial sobre tecnologias de WS da *Cambridge Semantics* (WORLD WIDE WEB CONSORTIUM, 2013).

Na Figura 1 visualizam-se as principais tecnologias presentes na WS, separadas conforme o grau de maturidade (AMARAL SERIQUE, 2012).

**Figura 1 – Camadas da Web Semântica**



**Fonte:** Tennison (2010, p. [6]).

A camada URI é responsável por prover o endereço para localização do recurso na *Web*, sendo útil para a identificação de conceitos (BRATT, 2007). O *Internationalized Resource Identifier* (IRI) é um protocolo que serve como complemento para URIs, composto por uma sequência de caracteres *Unicode* (DÜRST, 2011).

A camada RDF, fornece um framework que representa os metadados sobre os recursos, ou seja, sua padronização estabelece um modelo de dados e sintaxe para representar, codificar e transmitir metadados, de maneira a prover uma interoperabilidade entre os recursos na *Web* (TAYLOR, 2004; BRICKLEY; GUHA, 2000; LASSILA; SWICK, 1999; HEATH; BIZER, 2011).

A camada XML permite a integração de definições da WS com outros padrões baseados em *eXtensible Markup Language* (XML). A *Ontology Web Language* (OWL), possibilita a interpretação dos conteúdos presentes na *Web*, por meio de uma descrição formal eficiente, sendo compatível ao XML, RDF e *Resource Description Framework Schema* (RDFS) (MCGUINNESS; HARMELEN, 2004).

O *SPARQL Protocol and RDF Query Language* (SPARQL) é uma linguagem recomendado pelo W3C, a qual permite realizar consultas e manipular dados a partir

das estruturas em RDF (BREITMAN *et al.*, 2010).

O *Rule Interchange Format* (RIF) é uma camada que permite a definição de regras lógicas junto aos recursos, sendo composta por um conjunto de dialetos, os quais promovem a interoperabilidade entre diferentes sistemas de regras (MORGENSTERN *et al.*; 2013).

Na camada *Unifying logic* são definidas as regras para tratamento das informações descritas nos níveis anteriores. Esta se refere à necessidade de uma lógica, que defina regras para realização de inferências de novos conhecimentos com base nas declarações feitas usando os padrões descritos nas camadas inferiores (LARSEN, 2007).

A camada *Proof* destina-se a providenciar serviços adequados à verificação e comprovação/validação de definições realizadas na WS (BERNERS-LEE, 1998).

Para Palmer (2001), a camada *Trust* garante que as informações estejam representadas de modo correto, dessa forma, apresenta confiabilidade a WS por meio de mecanismos que ofereçam regras de acesso às informações e políticas de segurança.

E por fim, na camada *User Interface & Applications*, encontra-se os sistemas que empregam a WS como plataforma de execução.

Embora englobe várias especificações, modelos e padrões, a WS não é outra *Web* (CATARINO, 2009). A *Web* é legível por humanos e a WS legível por máquinas, ela expande os princípios de uma *Web* de documentos para uma *Web* de dados (BERNERS-LEE, 2009). Hoje, a maior parte do conteúdo presente na *Web* é dedicada à leitura e processamento, por intermédio da interação entre o usuário e o resultado de sua pesquisa.

## **5 RESULTADOS: ANÁLISE DOS MECANISMOS DE BUSCA NA PERSPECTIVA DA WEB SEMÂNTICA.**

Os mecanismos de busca na perspectiva da WS rastreiam os dados RDF na *Web*, fornecendo informações sobre a busca e a respeito dos dados monitorados. Na página do World Wide Web Consortium (W3C) são relacionados treze projetos desenvolvidos pelo SWEOLG, como pode ser visto no Quadro 1.



**Quadro 1 – Mecanismos de busca na Web Semântica**

NOME	CRIAÇÃO	MODIFICAÇÃO	TECNOLOGIA	HOMEPAGE	SITUAÇÃO
<SameAs.org>	Semea4	2012	RDF	<a href="http://sameas.org/">http://sameas.org/</a>	Ativo
Open Link Software	OpenLink Software	2012	RDF, RDFS, SPARQL, OWL, GRDDL, RDFa, RDB2RDF, R2RML, Direct Mapping	<a href="http://lod.openlinksw.com/">http://lod.openlinksw.com/</a>	Ativo
Sigma	DERI	2012	RDF, RDFa	<a href="http://sig.ma">http://sig.ma</a>	inativo
VisiNav	-----	-----	-----	-----	inativo
Falcons	Nanjing University	2010	RDF	<a href="http://w.s.nju.edu.cn/falcons/">http://w.s.nju.edu.cn/falcons/</a>	inativo
Falcons Explorer	Nanjing University	2010	RDF	<a href="http://ws.nju.edu.cn/explorer/">http://ws.nju.edu.cn/explorer/</a>	inativo
Sindice	DERI Gallway	2009	RDF	<a href="http://www.sindice.com">http://www.sindice.com</a>	inativo
Watson	Open University	2009	RDF	<a href="http://watson.kmi.open.ac.uk/">http://watson.kmi.open.ac.uk/</a>	Ativo
Yahoo! Micro Search	-----	-----	-----	-----	inativo
Semantic Web Search Engine (SWSE)	DERI Gallway	2009	RDF, SPARQL	<a href="http://swse.deri.org/">http://swse.deri.org/</a>	inativo
Swoogle	University of Maryland	2009	OWL	<a href="http://swoogle.umbc.edu/">http://swoogle.umbc.edu/</a>	Ativo
Semantic Web Search	-----	-----	-----	-----	inativo
Uriq	-----	-----	-----	-----	inativo

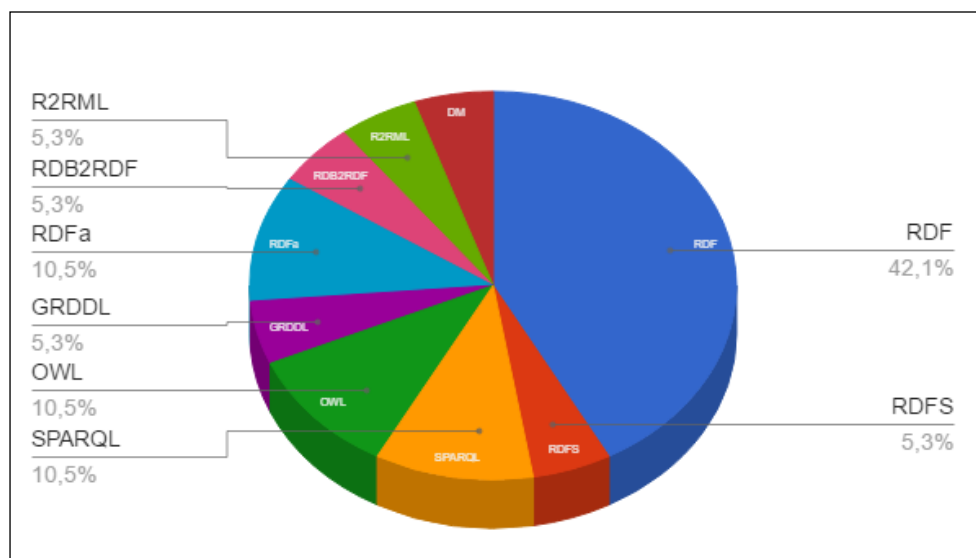
**Fonte:** Dados da Pesquisa.

Conforme apresentado no Quadro 1, dos 13 projetos relacionados na página do W3C, quatro encontram-se ativos. Entretanto, ressalta-se que durante o período de análise dos mecanismos, pôde-se verificar que algumas das páginas onde nesses mecanismos se encontram não estão atualizadas.

Observou-se que alguns mecanismos que estão inativos não apresentam nenhuma informação sobre seu funcionamento na página de busca. Dessa maneira, o usuário só identifica que o mecanismo está inativo ao realizar a busca; em alguns casos, o usuário tem sua busca redirecionada a outros grupos de discussão que debatem sobre a adoção dos mecanismos e a interoperabilidade dos dados.

Em relação à criação e suporte, esses projetos são mantidos por instituições privadas e/ou em universidades. Quanto à tecnologia usada para recuperação dos documentos pôde-se constatar, conforme descrito no Quadro 1, que os mecanismos podem apresentar várias das tecnologias desenvolvidas de acordo com a WS.

Entretanto, conforme visualizado no Gráfico 1, algumas são mais usadas:

**Gráfico 1** – Tecnologias presentes nos mecanismos de busca da *Web Semântica*

**Fonte:** Dados da Pesquisa.

De acordo com o Gráfico 1, um aspecto comum em quase todos os mecanismos de busca citados é a tecnologia da WS utilizada, o RDF. Somente o *Swoogle* apresenta a tecnologia *Web Ontology Language* (OWL).

De maneira a alcançar o propósito elencado neste trabalho, optou-se por analisar dois mecanismos de busca elaborados de acordo com as perspectivas da WS que apresentavam tecnologias distintas e que foram desenvolvidos em universidades. Nesse sentido, a pesquisa restringiu-se aos mecanismos *Watson* e *Swoogle*, considerando que estes até o momento da análise estavam ativos e propiciaram informações que auxiliaram na elaboração do estudo.

### 5.1 MECANISMOS DE BUSCA: WATSON E SWOOGLE

Conforme D'aquin *et al.* (2010), o mecanismo de busca *Watson* realiza buscas por meio das tecnologias semânticas, de modo a recolher, analisar e indexar automaticamente ontologias, fornecendo aos usuários acesso a uma ampla gama de conhecimento disponível na *Web*. Sua interface de apresentação é simples e objetiva, como visualizada na Figura 2.

O *Watson* foi elaborado a partir de três atividades principais (D'AQUIN, *et al.*, 2010): (i) rastreamento e descoberta de ontologias; (ii) validação e análise; e (iii) consulta e navegação.

**Figura 2 – Mecanismo de busca Watson**



**Fonte:** (<http://watson.kmi.open.ac.uk/>).

Já o *Swoogle* (FIGURA 3) é um rastreador, ou seja, um sistema de recuperação para a WS, sendo considerado o avô dos mecanismos de busca no universo da WS. Segundo Ding *et al* (2004), essa ferramenta extrai os metadados de cada documento descoberto e suas relações entre documentos.

**Figura 3 – Interface do Swoogle**

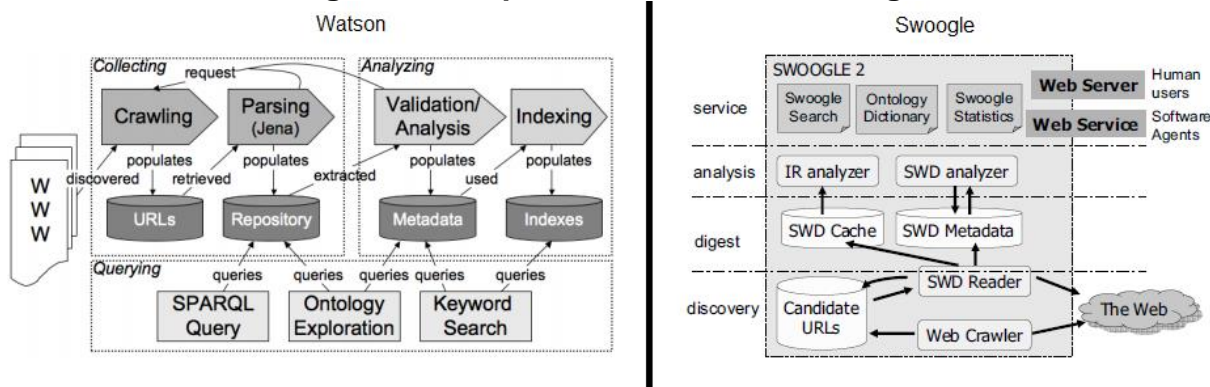


**Fonte:** (SWOOGLE, 2007, s.p).

A arquitetura do *Swoogle* pode ser estruturada em quatro componentes principais: (i) documentos da *Web Semântica* descobertos (SWD), (ii) metadados de criação, (iii) análise de dados, e (iv) interface.

Tanto o mecanismo de busca *Watson* quanto o *Swoogle* apresentam arquiteturas distintas (FIGURA 4):

**Figura 4 – Arquitetura: Watson e Swoogle**



**Fonte:** Elaborado com base em D'Aquin *et al.* (2010, p. [4]) e (LI DING *et al.*, 2004, p. [2]).

O desenvolvimento do *Watson* foi parcialmente financiado pelo *Néon Project* e pelo *OpenKnowledge*. As funções das camadas presentes no *Watson* conforme observado na Figura 4 não diferem dos utilizados no *Swoogle*, e correspondem:

- A camada de descoberta e rastreamento de ontologia é responsável pela obtenção de dados semânticos. Qualquer documento que não pode ser analisado por *Jena* é desconsiderado como uma forma de garantir que são documentos que contenham dados semânticos ou ontologias;
- A camada de validação e análise reúnem metadados sobre a rede semântica, a qual também ser usada para fins de indexação. Além disso, as relações semânticas entre ontologias são consideradas, a fim de detectar outras fontes de ontologias;
- A camada de consulta e de navegação está relacionada com as interfaces de consulta disponíveis que permitem o uso *Watson* funcionalidades. A Web Semântica é caracterizada por uma grande variedade de dados semânticos, enquanto toda a gama de informações é útil, entende-se que é crucial fornecer um acesso relevante aos usuários e suas aplicações. (CARVALHO; CARRAPATOSO, 2012, p. 29).

Dessa forma, o *Watson* opera explorando as relações entre ontologias considerando além de *hyperlinks* clássicos relativos às páginas da *Web* para conteúdo semântico, ou seja, uma ampla e extensível gama de ligações explícitas entre documentos. Uma característica peculiar do *Watson* é que ele é definido pelo W3C como um motor de busca de ontologias, ou seja, ele não indexa todos os documentos da WS, apenas se centra nas ontologias disponíveis na *Web*, diferente do SWSE, o qual é tem como busca exclusiva o padrão RDF.

O *Swoogle* é, portanto, um motor de busca da WS idealizado como uma ferramenta para facilitar o desenvolvimento desta, especialmente no que se refere às especificidades como:

- Encontrar ontologias apropriadas. Incapacidade de encontrar facilmente uma ontologia apropriada para uma tarefa de marcação especial tipicamente leva à criação de uma nova ontologia (ou o abandono do esforço de marcação). Os usuários do *Swoogle* podem consultar ontologias que contêm termos fixados em qualquer lugar do documento (incluindo complementos);
- Encontrar dados de instância. A fim de auxiliar aos usuários a integrar os dados da *Web Semântica* distribuídos na *Web*, o *Swoogle* permite consultar DWS com restrições sobre o que as classes e as propriedades a serem utilizadas/definidas por eles;
- Caracterizando a *Web Semântica*. Através das relações inter-documentos - sobre a *web semântica*, o *Swoogle* revela uma interessante estrutura de propriedades que leva a questões como "Como é que a estrutura de ligação na *Web Semântica*?", "Como são referenciadas as ontologias? ", e" como são apresentadas as ontologias externamente? ". (LI DING *et al.*, 2004, p. [1] Tradução nossa).

O *Swoogle* emprega um sistema de rastreadores para descobrir os documentos escritos em padrão RDF e HTML. Ele fornece serviços aos usuários humanos através de uma interface de navegador e agentes de *software*. Dentre as técnicas utilizadas para classificar os resultados da consulta, cita-se o *PageRank*<sup>4</sup>, algoritmo desenvolvido pelo *Google*, porém adaptado aos padrões da WS.

Durante a análise dos mecanismos, verificou-se que o *Watson* apresenta em sua interface inicial poucas informações que auxiliam o usuário na busca: (i) *What is it*; (ii) *Website*, (iii) *Submit URI*; e (iv) APIs - Interface de Programação de Aplicativos. Para realizar *downloads* e acessar publicações relativas a esse mecanismo deve-se escolher a opção *Website*, onde também é possível explorar a interface de demonstração e contatar os desenvolvedores.

O *Swoogle*, por sua vez, traz na interface inicial várias informações, como: (i) *Manual*; (ii) *News* – que redireciona a um *Blog* com informações referentes ao *Swoogle*; (iii) FAQs; (iv) *Web Service*; (v) *Submit URI*; (vi) *sw-archive - Semantic Web Archive Service*; (vii) *Feedback*; e *Swoogle2005* (esse ícone está inativo).

Considera-se que devido ao tempo que está ativo, o *Swoogle* apresenta mais informações relativas à sua adoção, trazendo exemplos que auxiliam ao usuário, e propiciando canais nos quais as dúvidas podem ser sanadas com mais rapidez.

---

<sup>4</sup> “[...] número que diz o quanto uma página web é relevante na Internet, ou seja, quanto maior esse número, maior é a relevância da página [...]” (SILVA; GALANTE, 2009, p. [4]).

## 6 CONSIDERAÇÕES

O aperfeiçoamento dos sistemas de recuperação de informação é cada vez mais almejado pelos pesquisadores e desenvolvedores no universo da WS. Prova disso, observa-se o anseio em prover aos sujeitos informações de qualidade em meio a avalanche de documentos existentes.

Durante a elaboração deste trabalho, pôde-se observar que a temática em torno dos mecanismos de busca desenvolvidos na perspectiva da WS é pouco explorada, o que demonstra um grande campo para desenvolvimento e aplicação de estudos que relacionem a CI, uma vez que esta pode contribuir nos aspectos relacionados tanto à organização quanto à recuperação da informação na *Web*.

A maioria dos mecanismos de busca citados pelo W3C, conforme já destacado, está inativa. Tal situação remete ao fato, que muitas dessas ferramentas foram desenvolvidas como resultado de trabalhos acadêmicos, dessa forma, não tiveram suas adequações e atualizações após a conclusão dos trabalhos.

O presente estudo realizou uma breve análise dos mecanismos *Swoogle* e *Watson*, pretendem-se posteriormente ampliar esta pesquisa de maneira a abarcar os mecanismos para visualizar possíveis aproximações entre os mecanismos de busca em operação.

Acredita-se que a maior contribuição deste estudo foi relacionar o desenvolvimento dessas ferramentas especializadas em torno da WS com campo de atuação da CI. Se o artigo de Bush (1945), em sua época, colocava em crise a sintaxe da indexação e dos índices, que incidiam também sobre as consequências na recuperação da informação, no que Medeiros e Vanti (2011) consideram como descontinuidade epistemológica da CI, nosso trabalho chama a atenção para a área em tela, no sentido de afirmar o pertencimento dos mecanismos de busca e da WS como objetos de estudo e pesquisa (básica ou aplicada) para a CI.

## REFERÊNCIAS

ALVES, R. C. V. *et al.* Ciência da informação, ciência da computação e recuperação da informação: algumas considerações sobre os métodos e tecnologias da informação utilizados ao longo do tempo. **REIC**, Marília, v. 6, n. 1, p. 28-40, 2007. Disponível em: <<http://www2.marilia.unesp.br/revistas/index.php/reic/article/viewFile/746/648>>. Acesso em: 20 maio 2016.

BARRETO, A. A. A condição da informação. **São Paulo em Perspectiva**, São Paulo, v. 16, n. 3, 2002. Disponível em: <[http://www.scielo.br/scielo.php?pid=S0102-88392002000300010&script=sci\\_arttext0](http://www.scielo.br/scielo.php?pid=S0102-88392002000300010&script=sci_arttext0)>. Acesso em: 20 maio 2016.

BATTELLE, J. **A busca**. Campinas: Campus; Rio de Janeiro: Elsevier, 2006.

BERNERS-LEE, T. Putting government data online. **Design Issues**. Jun., 2009. Disponível em: <<http://www.w3.org/DesignIssues/GovData.html>>. Acesso em: 20 maio 2016.

BERNERS-LEE, T. **Semantic web road map**. (Nota interna, W3C) 1998. Disponível em: <<http://www.w3c.org/DesignIssues/Semantic.html>>. Acesso em: 20 maio 2016.

BREITMAN, K. *et al.* **Publicação de dados governamentais no padrão linked data**. (Cursos) W3C. 2010. Disponível em: <<http://www.w3c.br/cursos/dados-abertos/curso/Parte-6-Modulo-1-AcessoDados.pdf>>. Acesso em: 20 maio 2016.

BRICKLEY, D.; GUHA, R. V. **Resource description framework schema 1.1**. (Recommendation) W3C. 2000. Disponível em: <<http://www.w3.org/TR/rdf-schema/>>. Acesso em: 20 maio 2016.

BUSH, V. As we may think. **The Atlantic Monthly**, Boston, v. 176, n. 1, p. 101-108, 1945. Disponível em: <<http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>>. Acesso em: 20 maio 2016.

CATARINO, M.E.; BAPTISTA, A.A. Web semântica e a qualidade no intercâmbio da Informação. In.: TOMAEL, M.I. **Fontes de informação na internet**. Londrina: EDUEL, 2002, p.31-51.

CENDÓN, B. V. Ferramentas de busca na *web*. **Ci. Inf.**, Brasília, v. 30, n.1, p. 39-49, jan./abr. 2001. Disponível em: <<http://www.scielo.br/pdf/%0D/ci/v30n1/a06v30n1.pdf>>. Acesso em: 20 maio 2016.

CONSÓRCIO WORLD WIDE WEB. **Web semântica**. Disponível em: <<http://www.w3c.br/Padroes/WebSemantica>>. Acesso em: 20 maio 2016.

CRUZ, C. H. B. Vannevar Bush: uma apresentação. **Ver. Latinoam. Psicopat. Fund.**, São Paulo, v. 14, n. 1, p. 11-13, mar. 2011. Disponível em: <<http://www.scielo.br/pdf/rlpf/v14n1/01.pdf>>. Acesso em: 20 maio 2016.

D'AQUIN, Mathieu et al. WATSON: A Gateway for the Semantic Web. Disponível em: <<http://people.kmi.open.ac.uk/marta/papers/eswc07watson.pdf>>. Acesso em 20 maio 2016.

DING, L.; *et al.* Finding and Ranking Knowledge on the semantic web. In.: GIL, H. *et al.* **The Semantic web – ISWC 2005**. Berlim: Springer Berlin/Hidelberg, 2005. P. 156 – 170. Disponível em: <[http://ebiquity.umbc.edu/\\_file\\_directory\\_/papers/197.pdf](http://ebiquity.umbc.edu/_file_directory_/papers/197.pdf)>. Acesso em: 20 maio 2016.

DING, L.; *et al.* Swoogle: A Semantic Web Search and Metadata Engine. 13th ACM **Conference on Information and Knowledge Management**.2004. Disponível em: <[http://ebiquity.umbc.edu/\\_file\\_directory\\_/papers/115.pdf](http://ebiquity.umbc.edu/_file_directory_/papers/115.pdf)>. Acesso em: 20 maio 2016.

DING, Y.; FENSEL, D. Ontology Library Systems: The key for successful Ontology Reuse. **The first Semantic Web Working Symposium (SWWS1)**, Stanford, USA, 2001. Disponível em:<<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.2159&rep=rep1&type=pdf>>. Acesso em: 20 maio 2016.

DÜRST, M. **For a introduction to IRIs vs domain names**. (Recommendation) W3C. 2011. Disponível em: <<http://www.w3.org/International/O-URL-and-ident.html>>. Acesso em: 20 maio 2016.

DZIEKANIAK, G. V. Análise dos sistemas de busca na WEB. **BIBLOS - Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 14, p. 171-184, 2002. Disponível em: <[www.brapci.ufpr.br/download.php?dd0=19626](http://www.brapci.ufpr.br/download.php?dd0=19626)>. Acesso em: 20 maio 2016.

FRAGOSO, S. Quem procura acha? O impacto dos buscadores sobre o modelo distributivo da *Word Wide Web*. **Eptic: Revista de Economia Política de las Tecnologias da La Información y Comunicación**, v.9, n.3, n.p., sept/dic. 2007. Disponível em: <<http://www.eptic.com.br>>. Acesso em: 20 maio 2016.

GRUBER, T. **What is an ontology?** 1996. Disponível em:<<http://www.ksl.stanford.edu/kst/what-is-na-ontology.html>>. Acesso em: 20 maio 2016.

HEATH, T.; BIZER, C. **Linked data: Evolving the Web into a Global Data Space**. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Calypool, 2011.

LARSEN, C. P. **Programming the semantic web** – a microformats compatible GRDL implementation for ActiveRDF. 2007. 78p. (Master's Thesis – Department of Computer Science) – Aalborg Of Computer Science, 2007. Disponível em: <<http://projekter.aau.dk/projekter/files/61071140/1187898631.pdf>>. Acesso em: 20 maio 2016.

LASSILA, O.; SWICK, R. R. **Resource description framework (RDF) model and syntax specification**. (Recommendation) W3C, 1999. Disponível em: <<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>>. Acesso em: 25 jul. 2015.

LEVY, S. **Google: a biografia**. Como o Google pensa, trabalha e molda nossas vidas. São Paulo: Universo dos Livros, 2012.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY (MIT). *Inventor of the week archive*. In: **Lemelson MIT**. Disponível em: <<http://web.mit.edu/invent/iow/bush.html>>. Acesso em: 20 maio 2016.



MCGUINNESS, D. L.; HARMELEN, F. **Web ontology language overview**. (Recommendation) W3C, 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em: 19 mar. 2015.

MEDEIROS, A. L.; VANTI, N. Vanevar Bush e as matrizes discursivas de as we may think: por uma possível história da Ciência da Informação. **Inf. & Soc.: Est.**, João pessoa, v. 21, n. 3, p. 31-39, set./dez. 2011. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/9652>>. Acesso em: 20 maio 2016.

MONTEIRO, S. D. As múltiplas sintaxes dos mecanismos de busca no ciberespaço. **Informação & Informação**, v. 14, n. especial, p. 68-102, 2009. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/2027/3223>>. Acesso em: 20 maio 2016.

MONTEIRO, S. D. Os mecanismos de busca: à guisa de uma tipologia das múltiplas sintaxes. In: TOMAÉL, M. I. (Org.). **Fontes de informação da Internet**. Londrina: EDUEL, 2008. Cap. 5, p. 97-122.

MORGENSTERN, L. *et al.* **RIF primer**. 2 ed. (Recommendation), W3C, 2013. Disponível em: <<http://www.w3.org/TR/2013/NOTE-rif-primer-20130205/>>. Acesso em: 20 maio 2016.

PALMER, S. B. **The semantic web** an introduction. 2001. Disponível em: <<http://infomesh.net/2001/swintro/>>. Acesso em: 20 maio 2016.

SANTOS, F. C.; CARVALHO, C. L. **Aplicações de suporte a web semântica**. (Relatório apresentado ao Instituto de Informática). Universidade Federal de Goiás, 2007. Disponível em: <[http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_004-07.pdf](http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-07.pdf)>. Acesso em: 20 maio 2016.

SILVA, G. R.; GALANTE, R. M. Um mecanismo de detecção de versões de páginas Web para melhoria do desempenho do algoritmo de *PageRank*. In: **IX Escola Regional de Alto Desempenho - Arquiteturas Multicore (ERAD)**, mar 2009, Caxias do Sul, RS, Brasil. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/wtdbd/2008/014.pdf>>. Acesso em: 20 maio 2016.

SILVA, J. L. C.; FREIRE, G. H. de A. Um olhar sobre a origem da Ciência da Informação: indícios embrionários para sua caracterização identitária. **Enc. Bibli. R. Eletr. Bib. Ci. Inf., Florianópolis**, v. 17, n. 33, p. 1-29, jan./abr., 2012. Disponível em: <[www.periodicos.ufsc.br/index.php/eb/.../1518-2924.2012v17n33p1](http://www.periodicos.ufsc.br/index.php/eb/.../1518-2924.2012v17n33p1)>. Acesso em: 20 maio 2016.

STUDER, R.; BENJAMINS, R. R.; FENSEL, D. Knowledge engineering: principles and methods. **Data Knowledge Engineering**, v. 25, n. 1/2, p. 161–197, 1998. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.3940&rep=rep1&type=pdf>>. Acesso em: 20 maio 2016.

TAYLOR, A. G. **The Organization of Information**. London: Libraries Unlimited. 2004.

TENNISON, J. **RDF and semantic web: can we reach escape velocity**. (Cursos). 2010. Disponível em: <<http://www.w3.org/2010/11/TPAC/RDF-SW-velocity.pdf>>. Acesso em: 29 maio 2016.