



UNIVERSIDADE FEDERAL DE SANTA
CATARINA

Centro de Ciências da Educação

**CURSO DE GRADUAÇÃO EM
BIBLIOTECONOMIA**



Giane Cristina Serafim

**WEB SEMÂNTICA: técnicas e ferramentas para tornar mais efetiva a busca
informacional na *Web*.**

Florianópolis, 2014

GIANE CRISTINA SERAFIM

**WEB SEMÂNTICA: técnicas e ferramentas para tornar mais efetiva a busca
informacional na *Web*.**

Trabalho de Conclusão do Curso de Graduação em Biblioteconomia, do Centro de Ciências da Educação da Universidade Federal de Santa Catarina, requisito parcial à obtenção do título de Bacharel em Biblioteconomia. Orientação de: Prof. Moisés Lima Dutra.

Florianópolis, 2014

Ficha Catalográfica elaborada por Giane Cristina Serafim, graduanda em Biblioteconomia da Universidade Federal de Santa Catarina.

S481w Serafim, Giane Cristina

Web Semântica: técnicas e ferramentas para tornar mais efetiva a busca informacional na Web. / Giane Cristina Serafim ; orientador, Moisés Lima Dutra - Florianópolis, SC, 2014. 77 f.

Trabalho de Conclusão de Curso (graduação) –
Universidade Federal de Santa Catarina, Centro de Ciências da Educação. Graduação em Biblioteconomia.

Inclui referências

1. Biblioteconomia. 2. Web Semântica. Metadados. Ontologia. Busca semântica . I. Dutra, Moisés Lima. II. Universidade Federal de Santa Catarina. Graduação em Biblioteconomia. III. Título.

Esta obra é licenciada por uma licença Creative Commons de atribuição, de uso não comercial e de compartilhamento pela mesma licença 2.5



Você pode:

- copiar, distribuir, exibir e executar a obra;
- criar obras derivadas.

Sob as seguintes condições:

- Atribuição. Você deve dar crédito ao autor original.
- Uso não-comercial. Você não pode utilizar esta obra com finalidades comerciais.
- Compartilhamento pela mesma licença. Se você alterar, transformar ou criar outra obra com base nesta, somente poderá distribuir a obra resultante com uma licença idêntica a esta.

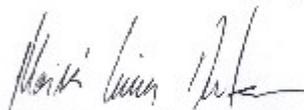
Este trabalho é dedicado a minha pessoa por todo esforço, dedicação e superação.

Acadêmica: Giane Cristina Serafim

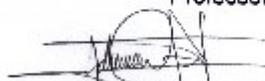
Título: **WEB SEMÂNTICA: técnicas e ferramentas para tornar mais efetiva a busca Informacional na Web.**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Biblioteconomia, do Centro de Ciências da Educação da Universidade Federal de Santa Catarina, como requisito parcial à obtenção do título de Bacharel em Biblioteconomia, aprovada com nota _____.

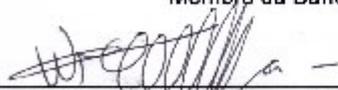
Florianópolis, de 2014.



Moisés Lima Dutra. Dr. UFSC
Universidade Federal de Santa Catarina
Professor Orientador



Adilson Luiz Pinto. Dr. UFSC
Universidade Federal de Santa Catarina
Membro da Banca Examinadora



William Barbosa Vianna. Dr. UFSC
Universidade Federal de Santa Catarina
Membro da Banca Examinadora



Márcio Matias. Dr. UFSC
Universidade Federal de Santa Catarina
Membro Suplente da Banca Examinadora

AGRADECIMENTOS

Agradeço a Deus, por me dar forças e coragem pra seguir em frente e superar os obstáculos.

Aos meus pais, Rosa Maria Serafim e Alberto Serafim, pela presença constante em todos os momentos de minha vida, por todo amor e apoio, por compreenderem a minha ausência em muitos eventos familiares.

A minha irmã Carla Patrícia Serafim que me ajudou a superar um momento muito difícil no decorrer do curso.

As minhas filhas Ingrid de Souza e Giovana Chodren de Souza que entenderam a minha ausência em muitos momentos.

Aos funcionários da **DECTI** – Divisão de Desenvolvimento de Coleções e Tratamento da Informação, em especial a Márcia de Filgueiras Gomes e Ramirez pelo carinho e apoio oferecido durante o período de estágio obrigatório da graduação.

Ao professor Moisés Lima Dutra, pelo apoio e excelente orientação no desenvolvimento deste trabalho.

Aos demais professores do departamento de Ciência da Informação, pelas contribuições ao longo da graduação.

As minhas colegas de turma, pelo companheirismo, em especial a Adélia Gewehr, a Angélica Germanovix, a Eliane Vicente e a Rafaela Paula Schmitz, pelo apoio nos trabalhos acadêmicos e em sala de aula.

A todas as pessoas que, direta ou indiretamente, colaboraram para a realização desta pesquisa.

“O sucesso nasce do querer, da determinação e persistência em se chegar a um objetivo. Mesmo não atingindo o alvo, quem busca e vence obstáculos, no mínimo fará coisas admiráveis.”

(José de Alencar, [18--])

RESUMO

SERAFIM, Giane Cristina. **Web Semântica: técnicas e ferramentas para tornar mais efetiva a busca informacional na Web.** 2014. 77 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Centro de Ciências da Educação, Universidade Federal de Santa Catarina. Florianópolis: UFSC, 2014.

Atualmente as pessoas estão preferindo a busca informacional na *Web*, acessando cada vez menos os tradicionais catálogos das bibliotecas para atender suas necessidades informacionais. Esse fato decorre do uso cada vez maior de tecnologias de informação e comunicação e pelo crescimento exponencial dos recursos informacionais disponibilizados na *Web*. Essa realidade potencializa alguns problemas relacionados à busca, localização, acesso e recuperação dessas informações na *Internet*. Nesse sentido, o problema que originou essa pesquisa está relacionado com a dificuldade na busca e na recuperação efetiva de recursos informacionais na *Web*. Este trabalho examina o desenvolvimento da *Web Semântica* e a aplicação de metadados e ontologias para a representação da informação, pois considera que estas são iniciativas importantes, que proporcionam uma melhor estruturação e representação dos recursos informacionais na *Web*. Com uma metodologia baseada na análise bibliográfica, documental e descritiva do tema, apresenta-se uma análise da *Web Semântica* como uma extensão da *Web* voltada para a organização dos recursos informacionais, além das técnicas e ferramentas tecnológicas que fundamentam a sua construção, com enfoque no uso de metadados e ontologias como elementos fundamentais desta proposta. Constatou-se que as tecnologias da *Web Semântica* relacionam-se com a área de Ciência da Informação, estabelecendo uma correlação na questão da representação descritiva, desta forma o bibliotecário com seus conhecimentos adquiridos podem contribuir para o desenvolvimento da *Web Semântica*, pois as técnicas de representação da informação utilizadas pela biblioteconomia podem colaborar como fundamento para o desenvolvimento de estudos relacionados às tecnologias criadas no seu desenvolvimento. O papel mais importante do bibliotecário na atualidade parece ser o de gerenciador da informação. Porém, o grande desafio é manter-se atualizado em relação às inovações tecnológicas. Como resultado pode-se concluir que a implementação da *Web Semântica* tem um longo caminho a seguir até chegar ao seu ideal no sentido de proporcionar a tão necessária estruturação e representação informacional dos recursos e conseqüentemente uma busca mais efetiva, baseada em recuperação semântica informacional.

Palavras-chave: Web Semântica. Metadados. Ontologias. Recuperação da Informação na *Web*. Busca semântica.

ABSTRACT

SERAFIM, Giane Cristina. **Web Semântica: técnicas e ferramentas para tornar mais efetiva a busca informacional na Web.** 2014. 77 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Centro de Ciências da Educação, Universidade Federal de Santa Catarina. Florianópolis: UFSC, 2014.

Nowadays people are preferring information seeking on the Web, accessing fewer traditional catalogs of libraries to meet their information needs. This fact stems from the increasing use of information and communications technologies and the exponential growth of information resources available on the Web. This reality strengthens some problems related to search, locate, access, and retrieval of such information on the Internet. In this sense, the problem that originated this research is related to the difficulty in seeking and effective recovery of information resources on the Web. This paper examines the development of the Semantic Web and the application of metadata and ontologies for the representation of information, as it believes that these are important initiatives, which provide better representation and structuring of information resources on the Web. With a methodology based on literature review, documentary and descriptive theme, presents an analysis of the Semantic Web as an extension of Web facing the organization information resources, in addition to technical and technological tools that underlie its construction, focusing on the use of metadata and ontologies as key elements of this proposal. It was found that the Semantic Web technologies are related to the area of Information Science, which shall relate to the question of descriptive representation, so the librarian with their acquired knowledge can contribute to the development of the Semantic Web, because the techniques representation of information used by the library can collaborate as a basis for the development of studies related to technology created in its development. The most important role of the librarian at present seems to be the information manager. However, the challenge is to keep up to date with technological innovations. As a result it can be concluded that the implementation of the Semantic Web has a long way to go until you reach your ideal in providing the much needed structure and representation of informational resources and consequently a more effective search, informational-based semantic retrieval.

Keywords: Semantics Web. Metadata. Ontologies. Information Retrieval on the Web. Semantic Search.

LISTA DE FIGURAS

Figura 1 - Arquitetura da <i>Web Semântica</i> proposta em 2005.....	22
Figura 2 - Modelo de dados <i>RDF</i>	28
Figura 3 - Modelo de dados <i>RDF</i> aplicado a um cenário específico	29
Figura 4 - Relacionamentos entre as entidades do Grupo 1	33
Figura 5 - Diretório do <i>DMOZ</i>	47
Figura 6 - Gráfico do <i>Netmarketshare</i> sobre uso mundial dos buscadores em 2014	50
Figura 7 - Página de acesso Buscador <i>Google</i>	50
Figura 8 - Visão geral dos mecanismos de busca.....	54
Figura 9 - Processo de busca utilizando um metabuscador.....	55
Figura 10 - Página de acesso do metabuscador <i>Whos Talkin</i>	57
Figura 11 - Mapa do Conhecimento	59
Figura 12 - Resultado da pesquisa no Mapa do Conhecimento.....	60
Figura 13 - Resultado da busca semântica no <i>Wolfram Alpha</i>	61
Figura 14 - Tela do <i>Sensebot</i> com a nuvem semântica	62

LISTA DE QUADROS

Quadro 1 - Operadores Booleanos.....	16
Quadro 2 - Lista de projetos do buscador Google.....	51
Quadro 3 - Exemplos de Metabuscaadores.....	56

LISTA DE ABREVIATURAS E SIGLAS

AACR2- Anglo American Cataloguing Rules 2ª edição
FRBR - Requisitos Funcionais para Registros Bibliográficos
GILS - Government Information Location Service
HTML - Hypertext Transfer Protocol
HTTP - HyperText Markup Language
IFLA - International Federation of Library Associations and Institutions
ISBD - Descrição Bibliográfica Internacional Normalizada
MARC - Machine Readable Cataloguing
NCSA - National Center for Supercomputer Applications
OCLC - On Line Computer Library
ODP - Open Directory Project
OWL- Web Ontology Language
PDF - Portable Document Format
RDA - Descrição de Recursos e Acesso
RDF- Resource Description Framework
TEI - Text encoding for Information Interchange
TICs - Tecnologias da Informação e Comunicação
URI - Uniform Resource Identifier
XML - eXtensible Markup Language
W3C - World Wide Web Consortium

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Motivação e Justificativa	12
1.2	Objetivos	12
1.2.1	Objetivo Geral.....	13
1.2.2	Objetivos Específicos	13
2	REVISÃO DE LITERATURA	14
2.1	Estratégias de Busca na Web	14
2.2	Web Semântica	17
2.2.1	Dados Vinculados.....	23
2.3	Ontologias	24
2.3.1	Resource Description Framework (RDF).....	27
2.3.2	Web Ontology Language (OWL).....	30
2.3.3	Requisitos Funcionais para Registros Bibliográficos (<i>FRBR</i>).....	31
2.4	Metadados	34
2.4.1	Dublin Core.....	35
2.4.2	Framework de Warwick.....	36
2.4.3	Text encoding for Information Interchange (TEI).....	37
2.4.4	Anglo American Cataloguing Rules 2ª edição (AACR2).....	38
2.4.5	Machine Readable Cataloguing (MARC).....	38
2.4.6	Government Information Location Service (GILS).....	39
2.4.7	Descrição de Recursos e Acesso (RDA).....	40
3	PROCEDIMENTOS METODOLÓGICOS	42
4	BUSCA SEMÂNTICA NA WEB: DESAFIOS E PERSPECTIVAS	44
4.1	Principais Ferramentas de Busca Na Web	46
4.1.1	Buscadores Semânticos.....	57
4.2	Técnicas que facilitam a busca na Web	63
4.3	Papel do Bibliotecário	65
5	CONCLUSÕES	69
	REFERÊNCIAS	72

1 INTRODUÇÃO

A *Web* evoluiu muito nos últimos anos e embora tenha sido projetada para facilitar o acesso, a troca e a recuperação da informação, atualmente se apresenta como um imenso repositório de documentos que deixa muito a desejar quando precisamos fazer uma busca e recuperar a informação necessária. Neste sentido, a explosão informacional e a dificuldade de encontrar a informação desejada tornaram necessária a organização da informação de maneira padronizada.

Neste contexto, a *Web Semântica* ou *Web Inteligente*, vem se apresentando como solução para ordenar o caos informacional existente na *Web*. É uma extensão da *Web* atual, acrescentando semântica ao atual formato de representação de dados para que possam ser interpretados pelas máquinas, a fim de melhorar a pesquisa e o uso dos dados. Uma maneira de permitir esse processamento automatizado é fornecer informações de tal forma que os computadores possam inferir o significado destas. Sob esse aspecto, a função de verificar o assunto do documento ficaria a cargo dos computadores, por meio da utilização de buscadores semânticos, economizando tempo e trabalho ao internauta.

As pessoas são capazes de usar a *Web* para realizar determinadas tarefas, como fazer uma pesquisa num catálogo de uma biblioteca, encontrar o significado de uma palavra, ou procurar por um eletroeletrônico que queira comprar. No entanto, um computador não consegue realizar essas tarefas sem a interação humana, pois as páginas *Web* não foram idealizadas para serem lidas por computadores, apenas por pessoas. Neste sentido, no projeto da *Web Semântica* foi proposta uma série de tecnologias necessárias para que as máquinas sejam capazes de associar e deduzir o conteúdo dos documentos armazenados na *Web*. Algumas dessas tecnologias serão abordadas neste trabalho a fim de alcançar os objetivos propostos.

O problema de pesquisa está relacionado com a dificuldade de se recuperar informações no ambiente *Web*, de maneira eficiente, principalmente quando esta tarefa inclui recuperação da informação com significado (recuperação semântica da informação).

Para tanto, a pesquisa consiste em explorar na literatura científica a evolução da *Web Semântica*, bem como as tecnologias relacionadas ao seu

desenvolvimento, como os Metadados, Ontologias e Dados Vinculados (Ligados ou Linkados).

Embora esteja em desenvolvimento, o objetivo da *Web* semântica é melhorar a busca e a recuperação da informação na *Web*.

Dentro desse contexto, duas questões serão discutidas e analisadas neste trabalho:

- 1) Quais os desafios encontrados pelos usuários na busca da informação na *Web* atual?
- 2) Como a informação vem sendo tratada na *Web* com o intuito de melhorar a busca e facilitar a recuperação da informação?

1.1 Motivação e Justificativa

A motivação para o desenvolvimento deste trabalho surgiu durante o curso de Biblioteconomia. Durante o curso, em todas as disciplinas relacionadas com o processamento técnico cursadas até então, o objetivo final da disciplina era facilitar a recuperação da informação. Ou seja, tratar a descrição e organizar adequadamente a informação conforme os processamentos técnicos para recuperar a informação relevante e pertinente ao assunto pesquisado.

Do ponto de vista pessoal, a motivação surgiu da intensão de saber como eu, na função de bibliotecária, poderei contribuir na aplicação dos conceitos e práticas da organização da informação na *Web*.

Do ponto de vista científico, a motivação se deu pela vontade de pesquisar de que forma a Ciência da Informação vem contribuindo e poderá contribuir para a representação dos recursos informacionais e conseqüentemente para a recuperação da informação na *Web*.

1.2 Objetivos

A seguir são apresentados os objetivos desta pesquisa.

1.2.1 Objetivo Geral

Fazer um levantamento de técnicas e ferramentas no contexto da *Web Semântica* que objetivam melhorar a recuperação da informação e tornar a busca informacional mais efetiva na *Web*.

1.2.2 Objetivos Específicos:

- a) Identificar as barreiras tecnológicas que tornam a recuperação da informação ineficiente em determinados casos na *Web* atual;
- b) Verificar quais são as técnicas e ferramentas mais recentes que seguem as diretrizes da *Web Semântica*, cujos objetivos são facilitar e tornar mais efetiva a busca e recuperação com significado na *Web*;
- c) Destacar o papel do bibliotecário com relação aos processos relacionados com a proposta de organização dos recursos da *Web*, no contexto da *Web Semântica*.

2 REVISÃO DE LITERATURA

Este capítulo apresenta a revisão de literatura para o desenvolvimento da pesquisa, abordando os temas relacionados aos objetivos propostos neste trabalho: Estratégias de Busca na *Web*, *Web* Semântica, Ontologias e Metadados.

2.1 Estratégias de Busca na *Web*

A *Internet* é uma fonte abundante de recursos informacionais, um enorme banco de dados que oferece serviços de informação. E por este motivo, a maioria dos usuários tem dado prioridade às buscas na *Web* para suprir suas necessidades informacionais.

Para realizar uma busca informacional na *Web*, o usuário ou pesquisador precisa utilizar ferramentas de busca. A busca na *Web* é realizada pelo usuário de acordo com um critério específico (utilizando uma dada palavra ou frase) e a resposta aparece como uma lista de referência sobre o assunto pesquisado com vários documentos.

No contexto da *Web*, o termo busca (*Query*) é usado genericamente para se referir às tecnologias e aos protocolos desenvolvidos para a recuperação da informação (W3C, 2014).

A **busca de informações na *Web***¹ pode ser feita de **duas formas**:

a) Por assunto/categorias: a busca é feita por tópicos que estão indexados por **categorias** e **subcategorias** de assunto;

b) Por assuntos específicos: a busca é feita utilizando as ferramentas de busca. Desta forma, o usuário deve informar a palavra chave ou a frase que caracteriza o que quer pesquisar. Esse tipo de pesquisa pode ser feita de duas formas:

1º) Pesquisa simples: pode ser feita na própria caixa inicial das ferramentas e oferece a opção de uso de comando mais gerais.

¹ Fontes pesquisadas:

SILVA, Edna Lucia da; MENEZES, Estera Muszcat. **Metodologia da pesquisa e elaboração da dissertação**. 4. ed. Florianópolis: UFSC, 2005. 138 p. Disponível em:

<https://projetos.inf.ufsc.br/arquivos/Metodologia_de_pesquisa_e_elaboracao_de_teses_e_dissertacoes_4ed.pdf>. Acesso em: 25 out. 2014.

OLIVEIRA, Mônica. **Operadores Booleanos**. 2009. Disponível em:

<<http://www.dbd.puc-rio.br/wordpress/?p=116>>. Acesso em: 06 nov. 2014.

2º) Pesquisa avançada: só pode ser realizada na página das ferramentas de busca, abrindo uma janela especial, na qual é possível usar os comandos mais específicos para aproximar ao máximo o resultado da pesquisa daquilo que se deseja encontrar.

Numa pesquisa simples a lista de documentos recuperados pode ser muito grande. Neste caso, a busca nem sempre atende a necessidade informacional do pesquisador, pois ele provavelmente não terá tempo para analisar o grande volume de documentos recuperados na busca. Nesse caso, o usuário pode utilizar outros recursos ou comandos que geralmente são oferecidos pelas ferramentas de busca para tornar os resultados mais precisos. Estes recursos ou comandos poderão ser identificados pelos usuários no sistema de ajuda de cada ferramenta de busca. Muitos usuários desconhecem estes recursos ou comandos e conseqüentemente não utilizam.

Geralmente os comandos utilizados na busca de informação são:

- a) **Uso de sinais:** o sinal de inclusão + (mais), o sinal de exclusão – (menos), aspas “ ” e o asterisco *;
- b) **Uso de operadores booleanos:** **AND** (e), **OR** (ou) e **NOT** (não) e também o uso de parênteses ().

O emprego dos comandos utilizados na busca simples possibilita:

Uso de aspas “ “

As aspas são utilizadas para que a ferramenta de busca considere as palavras como sendo uma frase. Por exemplo, ao colocar duas palavras entre aspas, “*Web Semântica*” a busca ficará limitada a documentos que contenham exatamente essa frase.

Uso do sinal de mais +

O sinal de inclusão + deve ser utilizado antes de uma palavra ou frase para informar a ferramenta de busca que ela deve selecionar os documentos que tenham obrigatoriamente todas as palavras precedidas do sinal +, em qualquer ordem que seja. Por exemplo: +metadados + “marcação semântica”.

Uso do sinal de menos –

O sinal de exclusão deve ser usado antes de uma palavra ou frase para informar a ferramenta de busca que ela não deve incluir os documentos que contenha aquela palavra (s) ou frase (s). Por exemplo:

+engenharia – “engenharia de produção”

Uso do asterisco *

O asterisco é usado para solicitar ao programa de busca que faça a busca em todos os documentos que contenham a parte inicial da palavra (até o asterisco) com qualquer terminação. Por exemplo:

produ* para recuperar produção, produtivo, produtos, produtividade.

O uso dos sinais pode ser combinado, e estes devem ser utilizados de forma lógica; a primeira palavra ou frase deve ser sempre a de inclusão. Conforme o exemplo a seguir:

+“inteligência artificial” – “redes neurais artificiais”

No caso acima, a ferramenta recuperará uma lista de documentos que tenha a expressão “inteligência artificial”, mas não contenha a expressão “redes neurais artificiais”.

A relação lógica entre os termos a serem pesquisados é estabelecida pelos operadores lógicos também conhecidos como **operadores booleanos**. Os operadores booleanos são usados nas buscas para possibilitar a ampliação ou o refinamento dos resultados. Os operadores válidos numa expressão de pesquisa são os seguintes:

Quadro 1 – Operadores Booleanos

Operador	Significado	Resultado
OR	União	Busca todos os registros onde exista qualquer um dos termos indicados
AND	Intersecção	Busca todos os registros onde ocorram simultaneamente os termos indicados
NOT	Exclusão	Busca todos os registros onde ocorra o primeiro termo exceto o segundo

Fonte: (SILVA; MENEZES, 2005)

O uso dos operadores booleanos pode ser observado nos exemplos abaixo:

AND: o uso do operador AND apresenta como resultado da pesquisa páginas que possuam obrigatoriamente todas as palavras ligadas por esse operador. Por exemplo, na busca:

“engenharia genética” AND ética.

O resultado da pesquisa será uma lista com todos os documentos com a expressão “engenharia genética” que também tenham a expressão “ética”

NOT: o uso dos operadores NOT oferece como resultado da pesquisa páginas que possuam a palavra que precede o operador AND e excluam as palavras que sucedem o operador NOT. Por exemplo, na seguinte busca:

“engenharia genética” NOT ética.

O resultado da pesquisa incluirá todos os documentos que possuam a expressão “engenharia genética”, mas que não contenham a palavra ética.

OR: o uso do operador OR apresenta como resultado da busca documentos que possuam tanto uma palavra como a(s) outra(s) ligada(s) por esse conectivo. Como por exemplo, na solicitação:

“engenharia genética” OR ética.

O resultado da pesquisa incluirá todos os documentos que possuam a expressão engenharia genética e a palavra ética não necessariamente no mesmo documento.

PARÊNTESES: os parênteses são utilizados para agrupar várias palavras ligadas pelos conectivos. Conforme exemplo abaixo:

qualidade AND (empresas OR organizações)

Todos esses recursos apresentados auxiliam a refinar a busca, porém, o pesquisador deve ser igualmente seletivo no uso das informações recuperadas na *Web*. Para isso alguns critérios de seleção devem ser adotados, como por exemplo, verificar as credenciais do autor, a forma como está escrito o documento, e a atualidade do *site*.

2.2 Web Semântica

A *Web* evoluiu muito desde a sua concepção. A *Web* Semântica é uma evolução da *Web* e conforme Feitosa (2006, p.120), “a *Web* Semântica consiste na separação do significado de sua estrutura, ou seja, amplia a noção de “documento” com a possibilidade de ‘semântica legível por máquina’.”

Ao longo dos anos essas transformações ocorreram devido à inserção de novas tecnologias que representam estágios evolutivos classificados em três gerações. De acordo com Feitosa (2006), a primeira geração, baseada na linguagem *HTML*, possibilitou a exibição dos documentos independentemente de sua localização física. Possuía sites com conteúdos estáticos, produzidos em sua

maioria por empresas e instituições, com pouca interatividade entre os internautas; a segunda geração propiciou o uso de diferentes formas de apresentação para a mesma estrutura de um documento, baseada na linguagem XML. Com apresentação de conteúdos produzidos pelos próprios usuários, com maior interatividade online através de *Blogs* e *sites* como o *Youtube*, *Flickr*, etc; já a terceira geração, a da *Web Semântica*, consiste numa *Web* composta por sistemas inteligentes; trata-se de uma *Internet* de significados.

Embora esteja presente na terceira geração da *Web*, a ideia da *Web Semântica* não é nova, conforme lembra Feitosa (2006) em sua obra, pois já havia sido idealizada por Bernes-Lee em 1989 e comentada em 2001, no artigo *The Semantic Web*. A *Web Semântica* não é uma *Web* separada, mas sim uma extensão da *Web* atual, uma *Web* diferenciada onde os conteúdos são melhor organizados de forma estruturada e padronizada com a finalidade de possibilitar que sejam interpretados por máquinas (BERNES-LEE; HENDLER; LASSILA, 2001, tradução nossa).

A *Web* atual é definida como uma “*Web sintática*” porque realiza a busca sem interpretação dos resultados. Quando o usuário faz uma pesquisa, em resposta retornam milhões de resultados, de tal forma que a busca da informação que desejamos em muitas situações se torna inviável e não atende as nossas necessidades. Isso ocorre porque as páginas da *Web* atual são direcionadas para pessoas e não para a interpretação por computadores, pois a sua principal característica é a apresentação da informação. Ficando a cargo do usuário, a função de interpretação dos dados.

Contraopondo a *Web* atual “a proposta da *Web Semântica* é estruturar os dados contidos nos sites de uma forma que o próprio sistema de busca identifique seu assunto e conteúdo e para isso seria preciso embutir semântica na estrutura dos dados” (PICKLER, 2007, p.70).

Semântica² é o estudo do significado. Incide sobre a relação entre significantes, tais como palavras, frases, sinais e símbolos, e o que eles representam, a sua denotação. A semântica linguística estuda o significado usado por seres humanos para se expressar através da linguagem. Outras formas de

² WIKIPEDIA. Semântica. 2014. Disponível em:<<http://pt.wikipedia.org/wiki/Sem%C3%A2ntica>>. Acesso em: 01 abril, 2014.

semântica incluem a semântica nas linguagens de programação, lógica formal, e semiótica.

Uma mesma palavra na linguística pode apresentar significados diferentes, o que dependerá do contexto onde está inserida. Por exemplo:

- a) “Ele ocupa um alto posto na empresa”. Nesta condição, posto - refere-se a um cargo, função que a pessoa ocupa.
- b) “Abasteci meu carro no posto da esquina”. Neste caso, posto - refere-se a um lugar destinado ao atendimento; posto de gasolina.

É o que chamamos de polissemia. Assim como palavras diferentes podem apresentar o mesmo significado e sentido, que vem a ser a sinonímia. Por exemplo:

- a) A **garota** renunciou *veementemente* ao pedido para que comesse.
- b) A **menina** recusou *energeticamente* ao pedido para que comesse.
- c) A **mocinha** rejeitou *impetuosamente* ao pedido para que comesse.

Vemos que os substantivos “garota”, “menina” e “mocinha” têm um mesmo significado, sentido, todos correspondem e nos remete à figura de uma jovem. Assim também são os verbos “renunciou”, “recusou” e “rejeitou”, que nos transmite ideia de repulsa, de “não querer algo” e também os advérbios que nos fala da maneira que a ação foi cometida “veementemente”, “energeticamente” e “impetuosamente”, ou seja, de modo intenso.

Essas questões linguísticas representam uma barreira no acesso à informação na *Web* atual. Para eliminar esta barreira a *Web Semântica* procura estabelecer maneiras de criar várias associações para cada informação, tanto relativas ao seu contexto quanto ao seu significado, de modo que possam ser reconhecidas por diferentes tipos de dispositivos.

Assim, é possível afirmar que a semântica é o estudo da função das palavras em determinado contexto. Desta forma, se o objetivo da *Web Semântica* é acrescentar semântica ao conteúdo da *Web*, essa semântica servirá para definir o sentido de um termo no contexto de um determinado documento.

Na última década muitas transformações ocorreram, no entanto, a *Web Semântica* ainda está muito aquém do seu ideal. Sendo que para se chegar a essa “*Web inteligente*”, como pode ser denominada, são necessárias muitas iniciativas que não são fáceis de serem concretizadas.

Souza e Alvarenga (2004) observam que, para atingir os propósitos da *Web Semântica*, é necessária uma padronização de tecnologias, de linguagens e de metadados descritivos de forma que todos os usuários da *Web* obedeçam a determinadas regras comuns e compartilhadas sobre como armazenar dados e descrever a informação armazenada e que esta possa ser utilizada tanto por humanos quanto por máquinas, de forma automática e sem ambiguidade. Mas será que é possível atingir os propósitos da *Web Semântica* de forma que alcance ‘todos’ os usuários da *Web*? É preciso levar em consideração as características pessoais dos usuários como a idade, por exemplo, pois a *Web* pode ser acessada por pessoas de todas as faixas etárias, desde uma criança muito pequena até uma pessoa bastante idosa. E por pessoas que tem bastante ou nenhuma habilidade com o uso das ferramentas tecnológicas oferecidas pela *Internet*. Outros fatores a serem considerados são a capacidade de compreensão e conhecimentos lógicos e abstratos dos indivíduos. Bem como a questão dos contextos socioculturais distintos de pessoas que vivem em diferentes regiões do mundo. Pois estas particularidades pessoais influenciam na maneira como as informações são tratadas, buscadas e acessadas na *Web*.

Barreto (2008, p. 9) salienta que novas tecnologias ainda enfrentam alguns problemas devido às características da *Web*, ou seja:

- a) A *Web* é distribuída: a *Web* não é constituída por uma autoridade central, isto constitui um entrave uma vez que é um produto composto por várias individualidades, a falta de um controle centralizado apresenta muitos desafios para dar significado às informações.
- b) A *Web* é dinâmica: O ritmo acelerado das mudanças das informações na *Web* representa um desafio a mais para qualquer tentativa de criar vocabulários padrão e fornecer semântica formal. Como o entendimento de um dado domínio muda tanto o vocabulário quanto a semântica podem ser refinados.
- c) A *Web* é abrangente: atualmente existem milhões de sites em funcionamento por todo o globo. Diante disso, a implementação de um sistema de reconhecimento semântico torna-se uma tarefa árdua.

No entanto, estas características não são passíveis de resolução, pois são intrínsecas da *Web*, na verdade a tendência é aumentar ainda mais esses problemas. Neste sentido, é necessário continuar avançando com novas tecnologias que consigam driblar esses problemas. O projeto *Web Semântica* traz no âmbito de suas inovações tecnológicas uma série de avanços que podem possibilitar melhorias significativas nos processos de organização e recuperação de informações na *Web*.

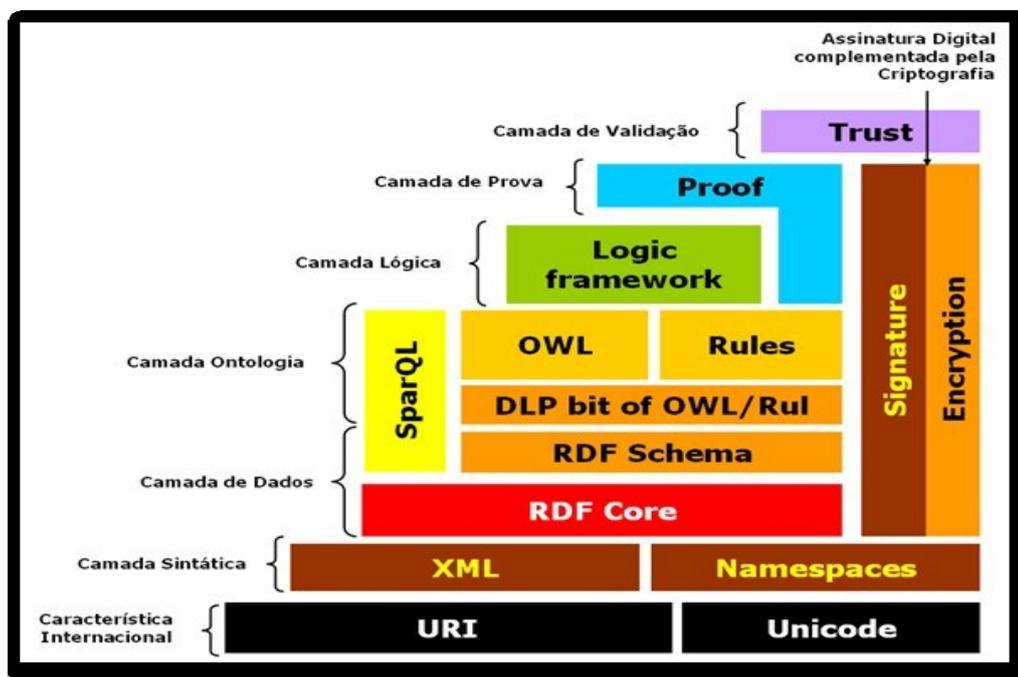
O projeto da *Web Semântica* foi proposto em 1998 por Tim Berners- Lee e segue em andamento. Desde então, muitos aprimoramentos foram realizados na *Web* atual. Exemplos disso são os novos motores de busca, interfaces inovadoras, criação de dicionários de sinônimos e a organização inteligente de conteúdos. Como os buscadores: Bing, UOL Busca, AOL, MSN, Ixquick, Clusty, entre outros. Essas melhorias visam à satisfação do usuário, porque desta forma o pesquisador ou usuário do sistema não precisará investir tanto tempo e empenho na *Internet* em busca da informação necessária, pois com as novas ferramentas e formas de organização da informação o usuário terá mais facilidade e sucesso na recuperação das informações adequadas a sua necessidade.

A intenção principal da *Web Semântica* é desenvolver tecnologias e linguagens que tornem a informação legível para as máquinas.

A finalidade passa pelo desenvolvimento de um modelo tecnológico que permita a partilha global de conhecimento assistido por máquinas. A integração das linguagens ou tecnologias *eXtensible Markup Language (XML)*, *Resource Description Framework (RDF)*, arquiteturas de metadados, ontologias, agentes computacionais, entre outras, favorecerão o aparecimento de serviços *Web* que garantam a interoperabilidade e cooperação (BARRETO, 2008, p. 20).

A figura 1 ilustra os diferentes módulos de uma arquitetura para a *Web Semântica*. O *W3C* definiu uma arquitetura para padronizar o desenvolvimento da *Web Semântica* e estas camadas desempenham funções que interagem umas com as outras.

Figura 1 - Arquitetura da *Web Semântica* proposta em 2005.



Fonte: Berners-Lee (2005), comentada por (SANTOS; ALVES, 2009).

Este trabalho está focado nas Camadas de Dados e Ontologia.

A Camada de Dados está diretamente relacionada com a representação, o processamento e a codificação dos metadados. O *RDF*, conforme será apresentado mais adiante neste capítulo, é uma linguagem para representação da informação na *Web*. O *RDF* também define um modelo de dados para descrição de semântica de dados para a inferência por computador.

Na Camada Ontológica, a ontologia é utilizada como um vocabulário, objetivando atribuir sentido e significado ao conteúdo dos documentos, atuando desta forma, como ferramenta de representação do conhecimento. A Ontologia também será tratada mais adiante neste capítulo.

Neste sentido a *Web Semântica* facilitará a pesquisa dos dados, ajudando a resolver problemas atuais, como a recuperação da informação pertinente a pesquisa em meio à sobrecarga informacional e a forma inadequada como os dados são estruturados e armazenados. O desenvolvimento da *Web Semântica* proporcionará um novo cenário de busca de informação e de interação com as pessoas.

Na sequencia serão apresentadas algumas ferramentas que fazem parte desse contexto. São ferramentas que já estão sendo implantadas para tornar a *Web* ‘mais inteligente’:

2.2.1 Dados Vinculados

A *Web Semântica* ou *Web dos Dados*, segundo *Lightbase* (2011), trata da nova especificação da *Internet* coordenada pelo *W3C* “onde não basta apenas apresentar os dados: é preciso fornecê-los de forma a serem mais facilmente encontrados e possam ser recuperados por outras ferramentas”. Segundo o *W3C* (2014) “é necessário organizar os dados disponíveis na *Web* em um formato padrão, acessíveis e gerenciáveis pelas ferramentas da *Web Semântica*”. Além disso, é necessário criar relações entre os dados. Ao fazer uma busca na *Web*, pode ser do interesse do usuário encontrar mais sobre o mesmo assunto dentro do contexto da sua pesquisa. Então é necessário classificar a informação relacionando-a a algum conteúdo. Dados Vinculados ou também mencionados de outra forma como (Dados *Linkados* ou Dados *Ligados*), é o nome que se dá às coleções de dados relacionados na *Web*.

Para desenvolver e obter Dados Vinculados, as tecnologias devem estar disponíveis para um formato comum (*RDF*), para possibilitar “o acesso aos dados existentes em bases de dados, sejam elas relacionais ou outras, como, por exemplo, as que estão estruturadas na forma de linguagens de Marcação (*XML* ou *HMTL*)” (*W3C*, 2014).

O *W3C* (2014) cita um exemplo de Dados Vinculados que é *DBPedia* (<http://wiki.dbpedia.org>) que, essencialmente, faz com que o conteúdo da *Wikipedia* fique disponível para *RDF*. Extrai informações da *Wikipedia* tornando-as acessíveis por meio de licenças livres (*Creative Commons Attribution-ShareAlike 3.0 License* e *GNU Free Documentation License*), além de anexar outros conjuntos de dados encontrados na *Web* aos dados da *Wikipedia*.

Quando tratamos do tema Dados Vinculados, normalmente encontramos outro tema bastante relacionado e ao mesmo tempo muito parecido, que é ‘Dados Abertos’. É importante ressaltar que ambos têm relação, porém apresentam conceitos diferentes. Pereira (2012) esclarece a diferença informando que, enquanto

a ideia de Dados Abertos se refere à publicação de informações e a garantia do acesso universal a elas, Dados Vinculados estão relacionados à conexão destes dados a outros conjuntos de dados.

Desta forma, pode-se dizer que ambos são importantes para o desenvolvimento da *Web Semântica*, visto que juntos proporcionam a integração de dados em larga escala na *Web*. Neste contexto, surgiu o conceito de dados abertos interligados - *linked open data* -, “definido como um conjunto de boas práticas para publicar, de forma aberta, e interconectar conjuntos de dados estruturados na *Web*” (SANTOS NETO et al. 2013, p. 82).

Arquivos, bibliotecas e museus poderão utilizar essa ferramenta para interligar os dados de seus catálogos e assim compartilhar informações.

2.3 Ontologias

Para entender o significado do termo ontologia é necessário recorrer a sua origem. O termo ontologia tem origem do grego *ontos*, ser, e *logos*, palavra e apresenta várias definições. Sendo que as primeiras definições surgiram na filosofia. No sentido filosófico a ontologia é empregada para descrever a existência do ‘ser’ no mundo. Breitman (2006, p. 30) apresenta o conceito filosófico de ontologia como “a ciência do que é, dos tipos de estrutura dos objetos, propriedades, eventos, processos e relacionamentos em todas as áreas da realidade.”

O termo ontologias vem sendo muito utilizado e discutido nas últimas três décadas. Comunidades como Inteligência Artificial, Representação do Conhecimento, Processamento de Linguagem Natural, *Web Semântica*, Engenharia de Software, entre outras. “Essa popularidade advém dos avanços tecnológicos que proporcionaram a massificação e o compartilhamento de informações digitais em âmbito global e a, conseqüente necessidade de organizá-las para então, recuperá-las” (SCHISSEL; BRÄSCHER, 2012, p. 126).

No contexto da Ciência da Computação e Ciência da Informação, o termo ontologia confere outro sentido. Utilizada com a finalidade de organizar e representar o conhecimento, a ontologia “se propõe a classificar as coisas em categorias, na perspectiva do sujeito e da linguagem do domínio” (VITAL; CAFÉ, 2011, p. 118). Breitman (2006, p. 30) cita a definição de ontologia de Gruber (1993) como a mais

encontrada na literatura da *Web Semântica*: “ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”. Neste conceito Ambrósio; Morais (2007, p.3) acrescentam que a

especificação formal quer dizer algo que é legível para os computadores, explícita são os conceitos, propriedades, relações, funções, restrições e axiomas explicitamente definidos. Conceitualização representa um modelo abstrato de algum fenômeno do mundo real e compartilhada significa conhecimento consensual.

Complementando a definição de Gruber (1993), conforme Vital; Café (2011, p. 119) “as ontologias, inerentes aos estudos da *Web Semântica*, objetivam o processamento automatizado da informação”.

O consórcio *W3C* define uma ontologia como “a definição dos termos utilizados para descrever e representar uma área do conhecimento”. E de uma forma bem resumida informa que ontologias devem fornecer descrições para os tipos de conceitos abaixo relacionados:

- a) Classes (ou “coisas”) nos vários domínios de interesse;
- b) Relacionamentos entre essas classes “coisas”;
- c) Propriedades (ou atributos) que essas “coisas” devem possuir.

Na esfera da Inteligência Artificial o que se denomina por ontologia “é considerado ontologia formal, ou seja, aquela que define vocabulário com o uso da Lógica” (MARCONDES; CAMPOS, 2008, p.112).

Do ponto de vista da Ciência da Computação, há diversas definições para o termo ontologia. Desta forma, Schissl e Bräscher (2012, p. 129) enfatizam que

o entendimento de ontologia como forma de especificação é fundamental na elaboração conceitual e construção da *Web Semântica*, pois o estabelecimento de limites a conceitos específicos e a definição das relações entre eles são essenciais para que a máquina possa inferir o significado da informação.

Dependendo da área do conhecimento, as ontologias podem variar em termos estruturais, no entanto, alguns componentes são básicos de uma ontologia. Entre eles, Hyvonen (2002, p.122 *apud* FEITOSA, 2006, p. 74) cita:

- a) classes - os conceitos genéricos de um vocabulário são definidos como classes. Por exemplo, a classe Tigre representa a categoria genérica de espécies de tigres;
- b) relação de superclasse – as classes são organizadas em hierarquias conceituais. Exemplificando, a classe Tigre é uma subclasse da classe dos carnívoros;
- c) propriedades de classe – classe podem ter propriedades, frequentemente denominadas “pistas”. As propriedades de uma

subclasse podem ser herdadas de sua superclasse, o que leva a uma economia em termos de representação e que torna possível a realização de inferências simples. Por exemplo: uma vez que carnívoros comem carne, e possuem dentes afiados, então os tigres também, visto que são carnívoros;

- d) características de propriedade (facetadas) – propriedades de classes podem, em si mesmas, possuir facetadas que as restringem. Por exemplo: as facetadas tipo de valor e cardinalidade podem ser associadas à propriedade *país*, para denotar que é um tipo de classe Pessoa e que o número de valores deve ser de pelo menos 2;
- e) indivíduos – um objeto individual sobre o qual trata uma ontologia é denominado de instância ou objeto. Cada objeto é uma instância de uma ou mais classes. Por exemplo: *Garfield* pode ser uma instância da classe *Gato* e também da classe *Gato de Cartoon*. A relação de pertinência é frequentemente denotada pela expressão “*é um*”;
- f) axiomas e restrições – uma ontologia baseada em lógica formal pode ter axiomas e restrições adicionais ou regras de inferência. Axiomas e regras podem ser escritos por meio de termos, funções, predicados, operadores, quantificadores, entre outros.

Do ponto de vista da representação do conhecimento, segundo Feitosa (2006, p.121),

uma ontologia não deve ser concebida apenas como um vocabulário informal, ou mesmo como uma linguagem de termos estruturados, mas requer uma possibilidade de interação algorítmica dos seus significados, e por sua vez, uma representação formal, cujo processamento dos significados pode ser realizado por máquinas.

Neste contexto, as ontologias vêm sendo utilizadas para estabelecer relações entre conceitos, objetos e classes de objetos, tomando suas propriedades essenciais para análise. Na *Web Semântica*, as ontologias procuram desta forma, “representar e descrever o conhecimento de certo domínio através da expressão formal dos relacionamentos entre os conceitos específicos de determinado campo científico, profissional ou técnico” (FEITOSA, 2006, p. 124).

Atualmente, com as TICs, grande parte do conhecimento produzido encontra-se armazenado na *Web*. Esse fato torna a busca de uma interpretação que seja possível tanto para humanos quanto para computadores um dos objetivos essenciais no desenvolvimento de ontologias. Nesse sentido, tal nível de organização, inteligível para humanos e máquinas, proporcionado em boa parte pela ontologia na representação do conhecimento, é a razão de ser da *Web Semântica*.

2.3.1 Resource Description Framework (RDF)

O *Resource Description Framework (RDF)* segundo Pollock (2010, p.151) “é uma especificação padrão de dados e modelagem utilizada para codificar metadados e informação digital”. Breitman (2005, p.20) define *RDF* como “uma linguagem declarativa que fornece uma maneira padronizada de utilizar o *XML* para representar metadados no formato de sentenças sobre propriedades e relacionamentos entre itens na *Web*.”

Pollock (2010) comenta que o futuro da *Web Semântica* gira em torno e é baseada predominantemente na força essencial da linguagem *RDF*. Essa visão se justifica pelo fato de que o *RDF* foi projetado segundo Breitman (2006, p. 20) “de modo a representar metadados de recursos *Web* de maneira legível e, sobretudo, processável por máquinas”. Com isso, entende-se que as buscas na *Web Semântica* futuramente serão mais fáceis e mais rápidas porque os computadores poderão realiza-las com a mínima interferência humana.

O framework em *RDF* “é uma combinação de protocolos com base em *Web (URI, HTTP, XML* e assim por diante) que são construídos e também a teoria de modelo formal (semântica) que define as relações possíveis entre itens de dados em *RDF*” (PALLOCK, 2010, p. 152).

Para facilitar o entendimento sobre *RDF*, Pollock (2010, p. 152) esclarece que “o conceito por trás do *RDF* é que podemos utilizar para descrever uma ‘coisa’ fazendo declarações sobre suas propriedades”. A ‘coisa’ é o recurso que podemos descrever, ou seja, qualquer coisa: livros, pessoas, lugares, produtos e assim por diante, desde que possuam um endereço *Web*.

Apesar de a informação na *Web* poder ser lida automaticamente, sua semântica não é determinada. Nesse sentido, “o *RDF* vai acrescentar meta informação a esses recursos, de modo a possibilitar às máquinas lidarem com eles de maneira inteligente” (BREITMAN, 2005, p.21).

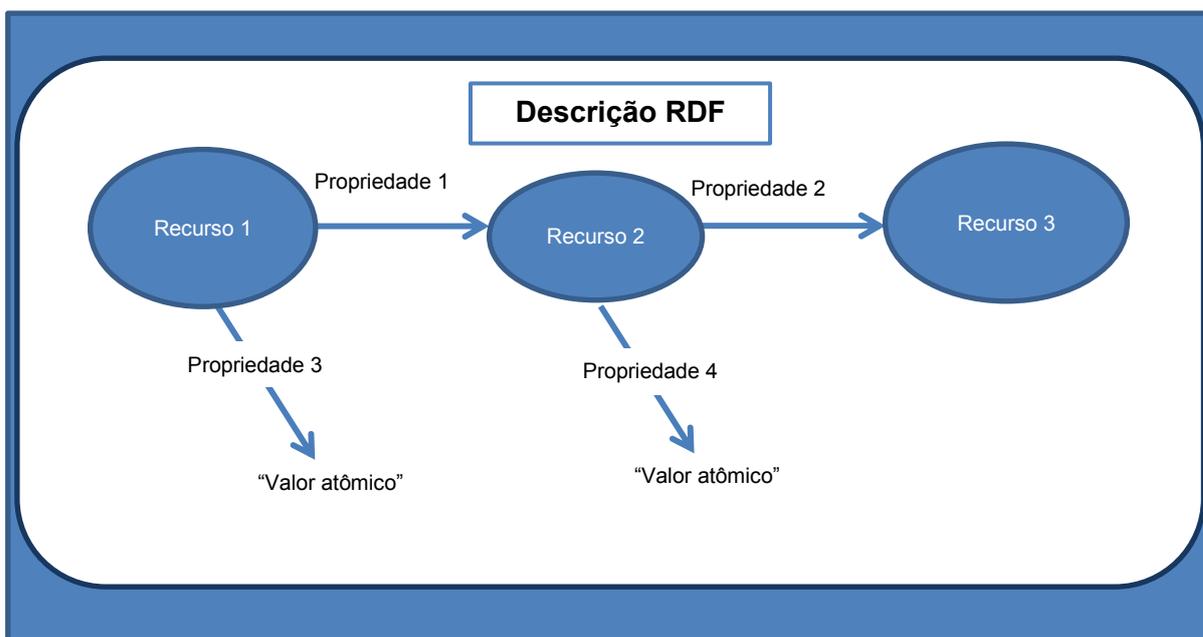
Outras informações pertinentes são apresentadas por Lourenço (2005, p. 67):

O *RDF* foi desenvolvido pelo *W3C (World Wide Web Consortium)*, com base nas pesquisas do *Dublin Core* e da *Arquitetura Warwick*. Assim, é dividido em duas aplicações: o *RDF Schema* e o *RDF Framework*, que é uma arquitetura de metadados. A parte descritiva deste padrão de metadados é chamada de *Esquema RDF* e define as propriedades dos recursos (título,

autor, assunto, tamanho etc), os tipos de recursos e suas semânticas. Este padrão providencia informação sobre a interpretação das declarações dadas em um modelo de dados *RDF* e é baseado na linguagem *XML*. O Esquema *RDF* constitui a base descritiva da Arquitetura *RDF*, ou modelo de dados.

*RDF*³ fornece um modelo para a descrição de recursos. Recursos têm propriedades (atributos ou características). Assim, *RDF* define um recurso como qualquer objeto que é exclusivamente identificável por um *Uniform Resource Identifier (URI)*. As propriedades associadas a recursos são identificados por tipos de propriedade e tipos propriedade têm valores correspondentes. Tipos de propriedade expressam as relações de valores associados a recursos. Uma coleção dessas propriedades que se refere ao mesmo recurso é chamada de descrição. No núcleo do *RDF* é um modelo independente de sintaxe para representar recursos e suas descrições correspondentes. A Figura 2 ilustra uma descrição genérica *RDF*.

Figura 2 - Modelo de dados *RDF*

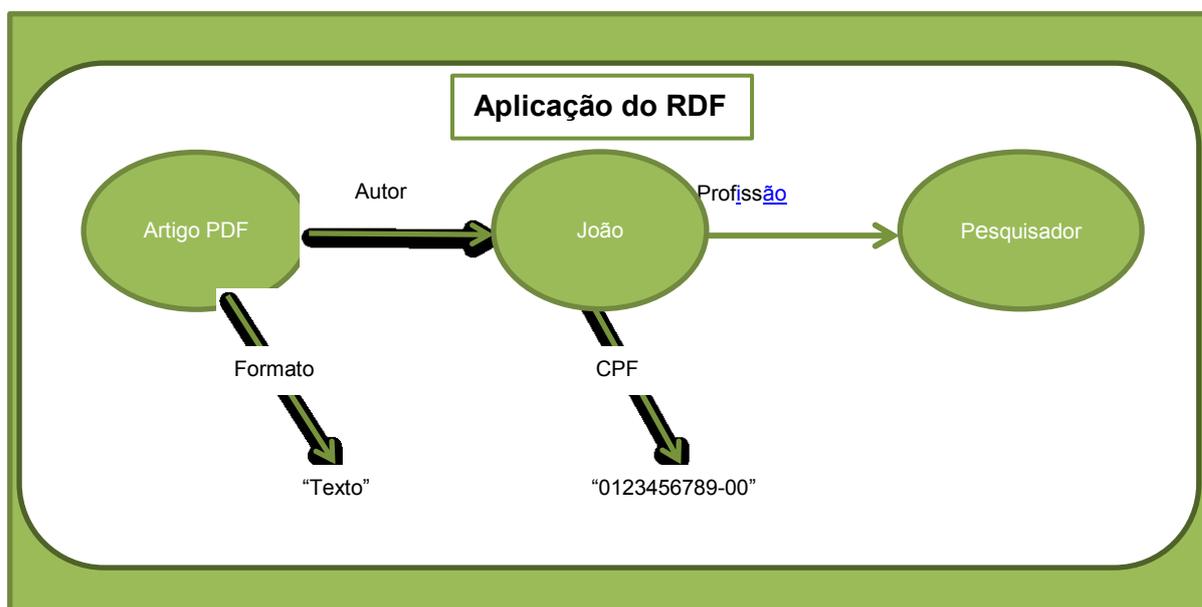


Fonte: a autora (adaptado de MILLER, 1998)

Na figura 3 temos a representação do modelo RDF aplicado a um cenário específico.

³ MILLER, Eric. Uma Introdução ao Resource Description Framework. **D-lib Revista**, Eua, p.1-10, maio 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 07 jun. 2014.

Figura 3 – Modelo de dados RDF aplicado a um cenário específico



Fonte: a autora (adaptado de MILLER, 1998)

A aplicação e a utilização do modelo de dados *RDF* pode ser ilustrada por meio de exemplos concretos. Considerando as seguintes afirmações:

- a) "O autor do documento 1 é João"
- b) "João é o autor do documento 1"

Para os seres humanos, estas declarações transmitem o mesmo significado (isto é, João é o autor de um documento particular). Para uma máquina, no entanto, estas cadeias são completamente diferentes. Usando um modelo triádico de recursos, propriedade-tipos e valores correspondentes, *RDF* tenta fornecer um método inequívoco de expressar semântica em uma codificação de leitura óptica.

O *RDF* na verdade é escrito em *XML*, utilizando uma linguagem chamada *XML/RDF*. "Esse fato facilita a troca de informações entre máquinas que utilizam aplicativos ou até mesmo sistemas operacionais diferentes" (BREITMAN, 2005, p. 23).

2.3.2 Web Ontology Language (OWL)

A *OWL*⁴ (*Web Ontology Language*) é uma linguagem para definir e instanciar ontologias na *Web*. Uma ontologia *OWL* pode incluir descrições de classes e suas respectivas propriedades e seus relacionamentos. *OWL* foi projetada para o uso por aplicações que precisam processar o conteúdo da informação ao invés de apenas apresentá-la aos humanos. Ela facilita mais a possibilidade de interpretação por máquinas do conteúdo da *Web* do que *XML*, *RDF* e *RDFS* (*RDF Schema*), por fornecer vocabulário adicional com uma semântica formal.

OWL é considerada uma tecnologia importante para a futura implementação da *Web Semântica*. Pois ela vem representando um papel importante em um número cada vez maior de aplicações, e vem sendo foco de pesquisa para ferramentas, técnicas de inferências, fundamentos formais e extensões de linguagem.

OWL foi projetada para disponibilizar uma forma comum para o processamento de conteúdo semântico da informação na *Web* e é atualmente uma recomendação do *W3C*, ou seja, um padrão. Ela foi desenvolvida para aumentar a facilidade de expressar semântica (significado) disponível em *XML*, *RDF* e *RDFS*. Logo, pode ser considerada uma evolução destas linguagens em termos de sua habilidade de representar conteúdo semântico da *Web* interpretável por máquinas. Uma vez que a *OWL* é baseada em *XML*, a informação pode ser facilmente trocada entre diferentes tipos de computadores usando diferentes sistemas operacionais e linguagens de programação.

Um exemplo exposto pelo *W3C* é a utilização de vocabulário para organizar o conhecimento. Bibliotecas, museus, jornais, portais governamentais, empresas, aplicações de redes sociais, e outras comunidades que gerenciam grandes coleções de livros, artefatos históricos, notícias, glossários de visita, entradas de blog, e outros itens podem agora usar vocabulários, usando formalismos padrão, para alavancar o poder de dados vinculados.

Recentemente o *W3C* publicou uma segunda versão do *OWL*, a *OWL 2* *Web Ontology Language*, informalmente *OWL 2*, é uma linguagem de ontologia para a *Web Semântica* com significado definido formalmente. *OWL 2* é uma extensão e

⁴ WIKIPÉDIA. OWL. Disponível em: <<http://pt.wikipedia.org/wiki/OWL>>. Acesso em: 07 jun. 2014.

revisão do *OWL Web Ontology Language* desenvolvido pelo Grupo de Trabalho Ontologia Web W3C e publicado em 2004 (designado por "CORUJA 1"). Como o OWL 1, OWL 2 é projetado para facilitar o desenvolvimento de ontologias e compartilhar através da *Web*, com o objetivo final de tornar o conteúdo *Web* mais acessíveis a máquinas.

2.3.3 Requisitos Funcionais para Registros Bibliográficos (*FRBR*)

A catalogação descritiva é uma das funções básicas do bibliotecário e neste contexto os *FRBR* – Requisitos Funcionais para Registros Bibliográficos representam um avanço significativo na área da representação bibliográfica. Os *FRBR* são o resultado de um reexame de práticas das normas de catalogação, que ocorreu sob a responsabilidade de um grupo de estudos da seção de Catalogação, Classificação e Indexação da IFLA. Tal estudo proporcionou reestruturar a forma de exibição dos elementos presentes nos registros bibliográficos, visando satisfazer as necessidades dos diversos usuários que utilizam estes registros.

De acordo com Moreno (2006, p. 17), “o modelo apresenta conceitos de **entidades, relacionamentos e atributos**⁵, baseado no modelo computacional Entidade-Relacionamento, lançando um novo olhar sobre o objeto bibliográfico, centrado no usuário e suas ações.”

A aplicação dos *FRBR* proporciona um volume maior de opções aos usuários que buscam informações nos registros bibliográficos. Visto que o modelo propõe o agrupamento de entidades com semelhanças em diversos níveis e relacionamentos de diversas naturezas.

Os *FRBR* têm como principais objetivos:

Primeiro, fornecer um quadro estruturado, claramente definido para relacionar dados registrados em registros bibliográficos às necessidades dos usuários destes registros. O segundo objetivo é recomendar um nível básico de funcionalidade para registros criados por entidades bibliográficas nacionais (IFLA, 1998, p. 7).

⁵ **Entidade** é aqui entendida como uma ‘coisa’ ou um ‘objeto’ no mundo real que pode ser identificada de forma unívoca em relação a todos os outros objetos. Uma entidade pode ser concreta ou abstrata.

Um **relacionamento** é uma associação entre uma ou várias entidades.

Atributos são as diversas características que um tipo de entidade possui, ou propriedades descritivas de cada membro de um conjunto de entidades. (CHEN, 1990, p. 21-24 *apud* MORENO, 2005, p. 34).

a) Entidades⁶

As 10 entidades definidas no *FRBR* estão divididas em 3 grupos.

Grupo 1 - entidades que são produto de trabalho intelectual ou artístico: **obra** (uma distinta criação artística ou intelectual), **expressão** (a realização artística ou intelectual de uma obra), **manifestação** (a materialização de uma expressão de uma obra) e **item** (um único exemplar de uma manifestação).

Grupo 2 – entidades responsáveis pelo conteúdo intelectual ou artístico, pela produção física e disseminação ou pela guarda das entidades do Grupo 1: **pessoa** (um indivíduo) e **entidade coletiva** (uma organização ou grupo de indivíduos e/ou organizações).

Grupo 3 - entidades que, juntamente com as entidades dos grupos 1 e 2, servem como **assunto** da entidade obra: conceito (uma noção abstrata ou ideia), **objeto** (uma coisa material), **evento** (uma ação ou ocorrência) e **lugar** (um local).

b) Atributos

Alguns dos atributos definidos no *FRBR* são: título, forma e data da obra; forma, data e idioma da expressão; título, edição, local e data de publicação, publicador e suporte da manifestação; identificador e origem do item; nome, datas e títulos de uma pessoa; nome, locais e datas associadas a uma entidade coletiva; termos representando conceitos, objetos, eventos e lugares.

c) Relacionamentos

Os relacionamentos descritos no *FRBR* são de três tipos. Relacionamentos entre as entidades do Grupo 1: uma obra é (pode ser) realizada por meio de várias expressões, uma expressão é materializada em várias manifestações e uma manifestação é exemplificada por vários itens (Figura 3).

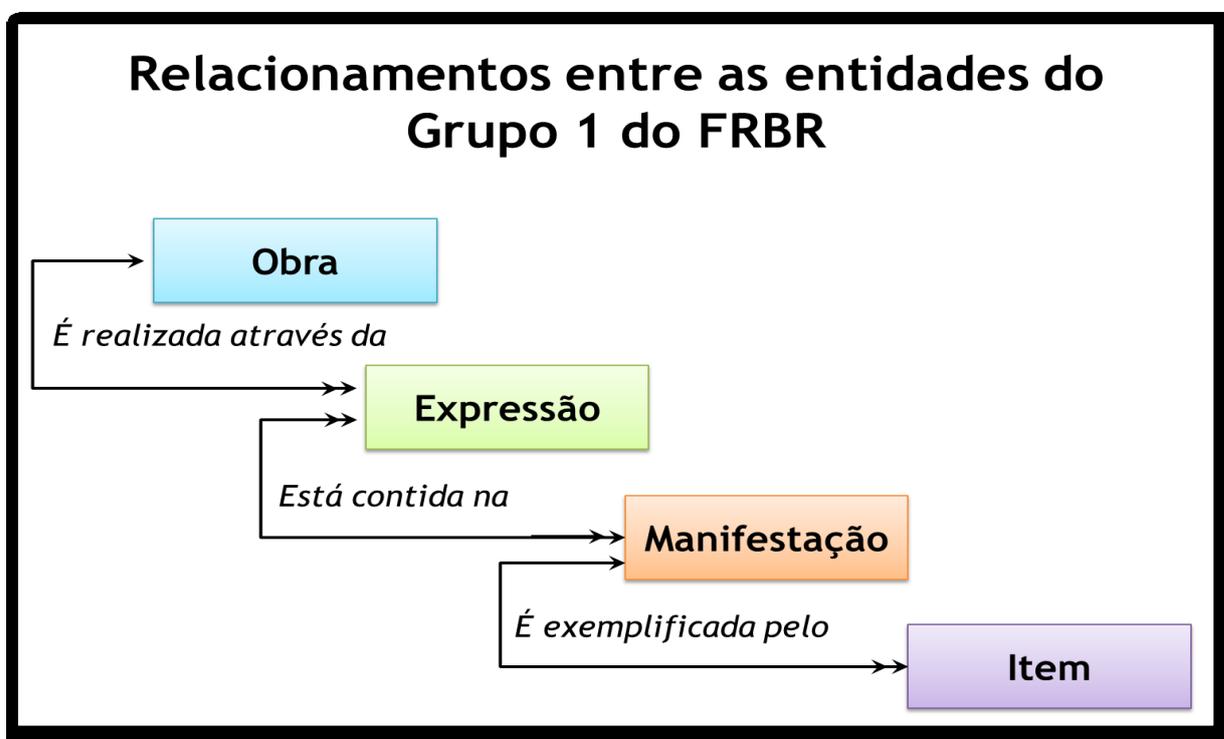
Relacionamentos entre as entidades dos Grupos 1 e 2: uma obra é criada por uma pessoa ou entidade coletiva, uma expressão é realizada por uma pessoa ou entidade coletiva, uma manifestação é produzida por uma pessoa ou entidade coletiva, e um item é guardado por uma pessoa ou entidade coletiva.

⁶ Fonte: ASSUNÇÃO, Fabricio Silva. O que é FRBR? 2012. Disponível em: <<http://fabricioassumpcao.com/2012/07/o-que-e-frbr.html#sthash.8K9FlaMZ.dpuf>>. Acesso em: 13 ago. 2014.

Relacionamentos de assunto: uma obra pode ter como assunto outra obra, uma expressão, uma manifestação, um item, uma pessoa, uma entidade coletiva, um conceito, um objeto, um evento ou um lugar.

Outros relacionamentos entre as entidades do Grupo 1 (alguns exemplos): uma obra é criada com base em outra obra, uma expressão é realizada por meio da tradução de outra expressão, um manifestação é uma reimpressão de outra manifestação, e um item é uma reprodução de outro item, etc.

Figura 4 - Relacionamentos entre as entidades do Grupo 1



Fonte: (ASSUNÇÃO, 2012)

Além das entidades, atributos e relacionamentos, o *FRBR* define também 4 tarefas dos usuário (ações realizadas pelo usuário durante a utilização de um catálogo ou de dados bibliográficos) e associam cada atributo e relacionamento com uma (ou mais) dessas tarefas. As tarefas são:

- a) **Encontrar** entidades que correspondem ao critério de busca do usuário (localizar uma única entidade ou um grupo de entidades em um arquivo ou banco de dados como o resultado de uma busca usando um atributo ou relacionamento da entidade);

- b) **Identificar** uma entidade (confirmar que a entidade descrita corresponde à entidade procurada ou distinguir entre duas ou mais entidades com características similares);
- c) **Selecionar** uma entidade apropriada às necessidades do usuário (escolher uma entidade que satisfaça os requisitos do usuário com respeito ao conteúdo, forma física, etc., ou para rejeitar uma entidade por ser inapropriada às necessidades do usuário);
- d) **Adquirir** ou obter acesso à entidade descrita (adquirir uma entidade por compra, empréstimo, etc., ou acessar uma entidade eletronicamente por uma conexão online com um computador remoto).

Os Requisitos Funcionais para Registros Bibliográficos (*FRBR*) podem possibilitar maior potencialidade nos relacionamentos entre registros bibliográficos e a recuperação da informação de forma mais efetiva nas solicitações de busca pelo usuário.

2.4 Metadados

Os metadados são dados sobre dados. No entanto, este conceito é muito vago. Pollock (2010, p. 118) acrescenta que os metadados são simplesmente formas de enriquecer os dados para que os sistemas de *software* possam interagir com a informação. Os metadados são informações sobre um documento: data, autor, título, data de publicação, editora, entre outras informações. E de acordo com Breitman (2005, p. 8) “servem para indexar páginas e sites na *Web Semântica*, permitindo que outros computadores saibam de que assuntos eles tratam.” No entendimento de Dias e Santos (2003, p. 83) os metadados são “utilizados para documentar e organizar de forma estruturada e padronizada as informações de documentos com o objetivo de facilitar e tornar mais efetiva a busca e a recuperação da informação na *Web*.”

No contexto da biblioteconomia, “metadado pode ser considerado como dado estruturado, que compartilha diversas características similares para a catalogação, e que descreve as características de um determinado recurso informacional” (FEITOSA 2006, p. 51).

Os padrões de metadados mais importantes e originários na área de Ciência da Informação de acordo com Feitosa (2006) são: *Dublin Core*, *TEI*, *AACR2*,

MARC e *GILS*. Esses padrões serão apresentados a seguir além de outros padrões de metadados tais como o *RDA* e o *RDF* que trouxeram maior impacto para *Web Semântica*.

2.4.1 Dublin Core

O padrão *Dublin Core* “foi criado pela *OCLC (On Line Computer Library)* e pelo *NCSA (National Center for Supercomputer Applications)*” (LOURENÇO, 2005, p. 63). Foi desenvolvido a partir da necessidade de uma Semântica para descrever recursos disponibilizados na *Internet*. Segundo Breitman (2005, p. 18), “do ponto de vista do bibliotecário seria suficiente criar um tipo de ‘cartão virtual’ para recursos disponibilizados na *Internet*”. Essa informação ajudaria as pessoas a descreverem seu material de maneira que facilitasse os navegadores a encontrarem o que estivessem procurando. Essa iniciativa surgiu durante um *workshop* em Dublin, Ohio, por este motivo ficou conhecida como *Dublin Core*. Segundo Dias; Santos (2003, p. 84),

sua finalidade é completar os seguintes objetivos: simplicidade de criação e manutenção; semântica de fácil compreensão; interação com padrões já existentes ou emergentes; escopo e aplicabilidade internacional; capacidade de extensão e interoperabilidade entre coleções e sistemas de indexação.

O padrão *Dublin Core* é largamente utilizado para descrever materiais digitais, tais como vídeo, som, imagem, texto e mídia composta como páginas da *Web*. Devido sua facilidade de manipulação e extensa capacidade de descrição dos recursos é um dos padrões mais utilizados. Constitui-se de 15 elementos qualificadores utilizados para anotação de itens de qualquer tipo. São eles:

1. Título: título do objeto
2. Criador: pessoa responsável pelo conteúdo do objeto
3. Assunto: tópico abordado pelo objeto
4. Descrição: descrição textual do conteúdo do objeto
5. Editora: entidade responsável pela disponibilização do objeto
6. Contribuidor: pessoa ou organização que contribui intelectualmente na criação do objeto
7. Data: data da criação ou publicação do recurso
8. Tipo: forma como o conteúdo é expresso

9. Formato: formato em que o documento é disponibilizado (*HTML*, *DOC*, *PDF*, etc).
10. Identificador: identificador único do objeto
11. Fonte: informação sobre as fontes de informação que contribuíram para a criação do conteúdo do objeto
12. Linguagem: Idioma que o objeto foi criado
13. Relação: relacionamentos com outros objetos
14. Cobertura: características temporais e espaciais
15. Direitos: informações sobre os direitos autorais do objeto

O *Dublin Core* é um padrão bem simples, como pode ser observado através do grupo de elementos qualificadores que o compõe.

Pollock (2010 p.142) relata que “bibliotecas e museus do mundo inteiro estão utilizando este conjunto de metadados das iniciativas da *Dublin Core* para codificar informações sobre seus bens de forma normativa para que sejam facilmente trocadas”. Significa que o *Dublin Core* é fundamental para permitir o acesso adequado aos recursos culturais e científicos através de bibliotecas, arquivos e museus.

2.4.2 Framework de Warwick

Um ano após a realização do *workshop* de Dublin foi realizado um 2º evento que aconteceu na Universidade de Warwick (Inglaterra) em abril de 1996. A Framework de Warwick “surgiu da necessidade de ampliar o *Dublin Core*, considerado muito simples, pois só disponibiliza um formato para descrição dos recursos” (BREITMAN, 2005, p. 19). Os representantes de diversos grupos de desenvolvedores se reuniram para propor um novo padrão para metadados, a *Framework de Warwick*. Neste evento foi abordada a necessidade de assegurar a compatibilidade com futuros desenvolvimentos e extensões dos esquemas de metadados. Foi proposta uma “integração de metadados a partir de esquemas diferentes em uma arquitetura de informação comum, capaz de garantir a interoperabilidade entre os metadados de diferentes tipos” (DANIEL JR; LAGOZE, 1997, p. 9). Houve consenso sobre a ideia de uma arquitetura de recipiente para o intercâmbio de vários pacotes de metadados. Então foi criada uma nova arquitetura,

que segundo Breitman (2005, p.17) é baseada no conceito de um container, que agrega vários tipos de metadados em pacotes separados.

A *Framework de Warwick* apesar de ter uma arquitetura bem definida, deixa em aberto algumas questões:

A primeira é relacionada à independência de sintaxe, pois cada um dos pacotes pode utilizar uma sintaxe diferente. Apesar de aumentar a flexibilidade do modelo, não garante que dois pacotes poderão trocar dados entre si. Outro problema é a Semântica utilizada. Nada no *framework* proposto garante que dois conjuntos de metadados possam estar utilizando um conceito com significados diferentes ou dois conceitos com o mesmo significado (BREITMAN, 2005, p. 20).

Para tratar estas dificuldades, surgiu um novo padrão, o *RDF – Resource Description Framework*. Desta forma, a *Framework de Warwick* nunca foi totalmente implementada.

2.4.3 Text encoding for Information Interchange (TEI)

O *TEI* é um consórcio que desenvolve coletivamente e mantém um padrão para a representação de textos em formato digital. Está baseado em “um conjunto de orientações que especificam os métodos de codificação para textos legíveis por máquina, principalmente na área de humanas, ciências sociais e linguísticas” (*TEI*, 2012).

Seus criadores, atentos às bibliotecas e aos princípios de descrição bibliográfica entenderam a necessidade do padrão *TEI* porque perceberam que os mesmos indivíduos que marcavam os textos eletrônicos criariam seus próprios metadados. Porém, conforme informa Lourenço (2005, p. 68), “os projetistas desenharam o *TEI* para atender a um catalogador treinado, baseando seus elementos descritivos no ISBD.” Da mesma forma que no formato *MARC*, a descrição dos registros é realizada intelectualmente e requer conhecimentos não somente sobre o formato, mas também sobre as regras de catalogação.

Ainda segundo Lourenço (2005, p.67), o *TEI* foi desenvolvido “com a meta de definir um jogo de diretrizes genéricas para a representação de documentos eletrônicos”. Resumidamente o *TEI* contém informação bibliográfica que viabiliza a recuperação da informação e a administração de compartilhamento de dados na utilização do recurso.

2.4.4 Anglo American Cataloguing Rules 2ª edição (AACR2)

O AACR2 é um código de catalogação e tem como objetivo a normalização da catalogação a nível internacional, auxiliando o tratamento da informação. Utiliza sistema de pontuação e a catalogação pode ser feita pelo suporte físico da obra.

Foi publicado em 1978, com três revisões em língua inglesa: 1988, 1998 e recentemente em 2002. A tradução de 2002 para o português teve revista a redação e numeração das regras e inclusão de novos exemplos (ANZOLIN, 2007, p.3).

O Código de Catalogação AACR2 está estruturado em 2 partes: a primeira parte trata da prestação de informações que descrevem o item que está sendo catalogado, e a segunda parte trata da determinação e estabelecimento de posições (pontos de acesso) ao abrigo do qual a informação descritiva deve ser apresentada para catalogar usuários, e com a tomada de referências aos cabeçalhos. Em ambas as partes as regras procedem do geral para o específico.

2.4.5 Machine Readable Cataloguing (MARC)

O MARC é o mais antigo dos padrões de metadados, “elaborado em 1960 pela *Library of Congress* para atender a necessidade de se ter um padrão para entrada de dados bibliográficos nos primeiros sistemas de automação de biblioteca que despontavam na época” (LOURENÇO, 2005, p. 62). Foi desenvolvido com a finalidade de catalogar eletronicamente o acervo das bibliotecas para permitir o empréstimo entre bibliotecas em redes integradas. Atualmente é o formato padrão adotado para entrada de dados bibliográficos em sistemas de automação bibliotecária e para a catalogação cooperativa. É amplamente utilizado pela comunidade bibliotecária e nos catálogos das bibliotecas.

O MARC é na verdade um formato padrão para gerenciamento de acervo eletrônico e intercâmbio entre bancos de dados bibliográficos. A entrada de dados em um registro MARC e a definição dos pontos de acesso deste registro são baseadas nas normas de catalogação do AACR2. “A descrição dos registros é realizada intelectualmente e requer conhecimentos sobre as regras de catalogação”

(FEITOSA, 2006, p. 57). Neste sentido o *MARC* cumpre a função de fornecer os recursos necessários para que um *software* de biblioteca administre corretamente o seu banco de dados.

2.4.6 Government Information Location Service (GILS)

O padrão *GILS* teve seus estudos iniciados em 1992, em parceria do *Office of Management and Budget*, do *National Archives e da Records Administration*, órgãos governamentais dos Estados Unidos. Tendo sido aprovado em 1994 pelo Departamento de Comércio, foi adotado como padrão Federal para processamento de informações do governo americano e em 1995 tornou-se o padrão de metadados obrigatório por lei em todos os departamentos governamentais dos Estados Unidos.

Sua intenção é auxiliar o público em geral a localizar e acessar tanto informações como fontes de informações. Para isso, define cerca de 70 atributos, denominados elementos centrais (*core elements*) além de aproximadamente mais 100 elementos herdados do protocolo Z39.50. Contudo, cada elemento de *GILS* pode ser:

- a) obrigatório ou opcional;
- b) repetível ou não-repetível;
- c) controlado ou não controlado; e
- d) elemento de agrupamento ou não-agrupamento.

Sendo assim *GILS* pode ser entendido como um serviço de coleta e disseminação de metadados; uma arquitetura para projeto físico e implementação e um esquema para descrição de metadados.

“O *GILS* utiliza como base a linguagem *SGML (Standard Generalized Markup Language)* e protocolo de comunicação de dados Z39.50 - protocolo de consulta que permite que os dados sejam recuperados de sistemas remotos” (LOURENÇO, 2005, p. 66).

2.4.7 Descrição de Recursos e Acesso (RDA)

RDA⁷ é um novo padrão de metadados que descreve os recursos mantidos nas coleções de bibliotecas, arquivos, museus e outras organizações de gestão da informação. Seu objetivo é fornecer um conjunto abrangente de diretrizes textuais e as instruções para a criação de metadados que abrange todos os tipos de conteúdo de recursos e meios de comunicação. Baseia-se nas Regras de Catalogação Anglo-americano desenvolvidas ao longo do século 20 e agora em sua segunda edição (*AACR2*), o desenvolvimento do modelo de *Requisitos Funcionais para Registros Bibliográficos (FRBR)* e a Declaração de Princípios Internacionais de Catalogação, formulado na primeira década do século 21. RDA essencialmente padroniza como o conteúdo de metadados é identificado, transcrito e geralmente estruturado, embora seja independente de qualquer codificação de metadados específico. RDA também identifica um conjunto geral de elementos de metadados, e em muitos casos, fornece um vocabulário controlado para o uso como o conteúdo de um elemento.

RDA⁸ oferece a bibliotecas o potencial de alterar significativamente a forma como os dados bibliográficos são criados e usados. Publicado em julho de 2010, RDA agora está disponível e algumas bibliotecas começaram a usá-lo para seu fluxo de trabalho de catalogação atual. RDA fornece um conjunto de orientações e instruções sobre a formulação de dados para apoiar a descoberta de recursos. RDA fornece um conjunto abrangente de orientações e instruções que abrangem todos os tipos de conteúdo e mídia. RDA procura criar um novo procedimento para descrever e proporcionar o acesso aos recursos de informação. Ao distinguir o padrão de catalogação do ISBD e o Padrão *MARC*, seus criadores esperam aumentar o número de contextos em que as descrições bibliográficas podem ser utilizadas e reutilizadas. Ambas as normas dependem de uma estrutura conceitual: o *RDF* triplo de sujeito, predicado e objeto, e o paradigma *FRBR* que subjaz RDA.

⁷ HILLMANN, Diane et al. RDA Vocabulários: Processos, Resultados, Uso. **D-lib Revista**, Eua, v. 16, n. 1, p.1-20, fev. 2010. Disponível em: <<http://dlib.org/dlib/january10/hillmann/01hillmann.html>>. Acesso em: 08 maio 2014.

⁸ D. Grant Campbell. . 2.011 RDA e RDF: Uma análise do discurso de dois padrões de descrição de recursos. Em Smiraglia, Richard P., ed. *Proceedings da North American Symposium on Organização do Conhecimento*, vol. 3. Toronto, Canadá, p. 207-16. Disponível em: <<http://journals.lib.washington.edu/index.php/nasko/article/download/12804/11285>>. Acesso em: 08 mai. 2014.

As estruturas de dados do RDA são originárias de um conjunto de condições administrativas e sociais específicas. RDA está em conformidade com a Declaração Internacional sobre Princípios de Catalogação Internacional Federação de Associações de Bibliotecas (IFLA). Seu objetivo declarado é o de permitir que os usuários encontrem, identifiquem, selecionem e obtenham os recursos de que precisam, localizem, confirmem, esclareçam e compreendam as entidades-chave, como pessoas, órgãos sociais e conceitos.

3 PROCEDIMENTOS METODOLÓGICOS

O presente trabalho caracteriza-se por ser uma pesquisa bibliográfica e documental, pois desenvolveu - se com base em material já elaborado constituído principalmente de livros, artigos científicos e outros documentos que fundamentam a pesquisa. O tipo de pesquisa descritiva também se aplica na busca dos objetivos propostos, pois se trata de um tipo que “visa à descrição das características de determinados fatos e fenômenos de determinada realidade”. (GIL, 2002).

A adoção dessa metodologia permitiu abordar os aspectos mencionados nos objetivos específicos definidos anteriormente e proporcionou a compreensão e concretização do objetivo geral proposto.

Os procedimentos metodológicos que nortearam a pesquisa foram seguidos da seguinte forma:

1. Levantamento bibliográfico: realizado em nível nacional e internacional em fontes bibliográficas primárias: livros, periódicos, anais de congresso, dissertações e teses; secundárias (Catálogo de bibliotecas, Base de Dados textuais e referenciais como: Scielo, Brapci, Periódicos Capes) da Ciência da Informação.
2. Levantamento documental: Também foram consultados os documentos disponíveis no site do *W3C*. Por exemplo, *Dublin Core*, *RDF*, etc. A seleção dos documentos seguiu os critérios de pertinência quanto aos assuntos relacionados ao tema da pesquisa, aos idiomas português, inglês e período de publicação limitado aos últimos dez anos, mas considerando algumas publicações importantes encontradas com pouco mais de dez anos de publicação.
3. Após a leitura dos textos e tendo como foco os objetivos explicitados, procedeu-se a análise e seleção dos fragmentos mais importantes e relevantes que embasaram a pesquisa e proporcionaram um maior entendimento sobre o tema proposto fornecendo subsídios para o alcance dos objetivos propostos.

O capítulo a seguir, apresenta a busca semântica na *Web*, desafios e perspectivas em relação à *Web Semântica*, técnicas e ferramentas que facilitam a busca na *Web*, bem como o papel do bibliotecário neste contexto. Após análise dos

assuntos acima citados, foram apresentados e discutidos os resultados das três etapas de pesquisa.

4 BUSCA SEMÂNTICA NA WEB: DESAFIOS E PERSPECTIVAS

Com o aumento exponencial da informação na *Web*, surgiu a necessidade de novas formas de organização para melhorar a recuperação da informação e tornar a busca mais efetiva. Neste sentido, a busca pelo aprimoramento das ferramentas e mecanismos de busca direcionados à localização e recuperação das informações é um tópico bastante importante e um grande desafio. A efetividade desses mecanismos de busca depende principalmente da maneira pela qual as informações são estruturadas e organizadas na *Web*. “Documentos podem ser estruturados e organizados de várias formas diferentes na *Web* e as ferramentas de busca têm que utilizar mecanismos de recuperação adequados para cada tipo de organização” (DIAS; SANTOS, 2003, p. 83).

As ferramentas de busca estão se atualizando e atribuindo mais recursos para tornar a busca na *Web* mais efetiva. Mesmo com a variedade de ferramentas disponíveis e apesar de estarem sempre em constante atualização em suas técnicas de busca, muitas vezes as ferramentas atuais ainda não conseguem atender de modo satisfatório seus usuários. Apesar de toda tecnologia algumas limitações ainda ocorrem. De acordo com Alves (2005, p.25),

os principais empecilhos estão relacionados com a crescente quantidade de informações disponibilizadas; com as próprias limitações das ferramentas de busca (suas técnicas de busca e indexação dos sites) e com a falta de representação (tratamento) adequado dos recursos informacionais na rede, tanto na representação por uma linguagem de marcação que possibilite melhor visualização do conteúdo do recurso, como também na construção de formas de representação convencionadas pela Biblioteconomia, catalogação e metadados.

Conjeturando sobre a colocação de Alves (2005) em relação aos problemas que dificultam as buscas na *Web*, a crescente quantidade de informações disponibilizadas é um fato que tende a aumentar, pois o crescimento de informações disponibilizadas na *Internet* é exponencial. No entanto, para que este problema seja solucionado ou minimizado é necessário organizar tal conteúdo, dar sentido e conexões às informações. Para tratar do problema da falta de representação adequada dos recursos informacionais na rede, é importante salientar a necessidade de integração e de uniformidade de linguagem e significado dos dados através da organização da informação na *Web*.

Em relação às formas de representação convencionadas pela Biblioteconomia, percebe-se que houve melhorias. Conforme informa Catarino e Souza (2012, p.84), “a representação descritiva vem sofrendo nos modelos convencionais de descrição bibliográfica, mudanças significativas que são provocadas pelas tecnologias de informação e comunicação.” O FRBR e o RDA, tópicos apresentados no capítulo 2 deste trabalho, são exemplos de desenvolvimento na área, que proporcionam uma melhoria na representação descritiva e vêm de encontro com a proposta de organização de recursos da *Web Semântica*. Porém, implantar estas mudanças na catalogação requeridas pelo FRBR e o RDA não é tarefa fácil.

Supõe-se que para incluir os tradicionais catálogos no ambiente da *Web Semântica*, ou seja, transformar os dados bibliográficos em dados vinculados na *Web* faz-se necessário modelar os dados já existentes que se utilizam de *MARC* ou *Dublin Core*, por exemplo, nos moldes do *RDF*. Deve-se, ainda, tornar o modelo conceitual *FRBR* e o código de catalogação *RDA*, recentemente publicado como sucessor do *AACR*, em base *RDF*; além de escrever os vocabulários controlados existentes em linguagens propostas pelo W3C (CATARINO; SOUZA, 2012, p. 86).

No entanto, observa-se que as ações que vêm sendo implantadas atuam “no sentido de permitir que os dados bibliográficos contidos nos catálogos bibliográficos possam ser inseridos em declarações nos moldes do *RDF*” (CATARINO; SOUZA, 2012, p. 89).

Em relação aos metadados, de acordo com informações apresentadas na literatura pertinente ao assunto, o fundamento da *Web Semântica* é o *RDF* que é uma forma de descrição dos metadados sobre os recursos da *Web*, tais como: título, autor, data de modificação de uma página *Web*, etc.

O relacionamento entre os itens representa um avanço na catalogação, assim como a vinculação dos dados representa uma melhoria para tornar a busca na *Web* mais eficiente. Entende-se desta forma, a partir das informações aqui apresentadas tanto em relação à *Web Semântica* quanto à Representação Descritiva, que fica clara a relação das atividades desenvolvidas em ambas as áreas, pois ambas preocupam-se em propiciar aos usuários formas mais produtivas e amplas de recuperação da informação.

As Ontologias, no contexto da *Web Semântica* (cf. capítulo 2), servem para representar e descrever o conhecimento de certo domínio através da expressão formal dos relacionamentos entre os conceitos específicos de

determinado campo científico, profissional ou técnico. Por exemplo, bibliotecas, museus, arquivos, podem agora usar as ontologias para organizar a informação na *Web*. Como exemplo de Ontologia na área da Ciência da Informação, pode-se citar o *FRBR*. Mais uma vez, percebe-se que a Representação Descritiva está intimamente ligada a *Web Semântica*, visto que, segundo Catarino e Souza (2012 p. 88), o *FRBR* é uma “Ontologia criada para a captura e representação da semântica subjacente das informações bibliográficas e para facilitar a integração, mediação e intercâmbio de informações bibliográficas e de museus.”

Considerada na literatura como um caminho para solucionar os problemas relacionados à recuperação da informação na *Internet*, a *Web Semântica* apresenta novas formas de organização dos recursos informacionais da *Web* e de acordo com seus idealizadores possibilitará expressar um maior significado das informações e proporcionar o desenvolvimento de ferramentas de busca mais efetivas.

O objetivo das aplicações com características semânticas é definir o significado de textos ou dados, criando vinculações para os usuários. É neste sentido que as ferramentas de busca estão intensificando seus esforços para possibilitar uma busca mais eficaz em relação à recuperação semântica da informação.

4.1 Principais ferramentas de busca na *Web*

Recuperar informações, ou localizar outros sites na *Internet*, de maneira eficiente, nem sempre é uma tarefa fácil para seus usuários. Para isso, estes podem recorrer a sites especializados em localizar outros sites, conhecidos como ferramentas de busca, isto é, catálogos de endereços de outros sites que existem na *Internet* (MORAIS; AMBROSIO, 2007).

As ferramentas de busca recebem várias terminologias na literatura científica como, buscadores, mecanismos de busca, serviços de busca, motores de busca, entre outros que, por uniformização, serão tratados neste trabalho, como “ferramentas de busca” (MONTEIRO, 2009). As ferramentas de busca mais comuns dividem-se em três tipos: Diretórios, Motores de Busca e Metabuscaadores.

a) Diretórios⁹

Diretórios foram os primeiros sistemas de busca criados. Também chamados de catálogos, tem como característica a categorização e organização em tópicos. Estes tópicos seguem uma estrutura lógica e são subdivididos por detalhamentos. Normalmente os diretórios são feitos por pessoas, diferente dos motores de busca que usam robôs como veremos mais adiante. Os principais diretórios são o DMOZ e o *Yahoo!*

Figura 5 – Diretório do DMOZ



Fonte: (DMOZ, 2014)

Os diretórios ainda são amplamente utilizados. As vantagens deles sobre as ferramentas de busca são:

1. Possibilidade de aprofundar ou avançar nas buscas;
2. Buscar apenas conteúdo selecionado. Por fazer parte de uma estrutura, não se corre o risco de obter resultados dúbios. Por exemplo, ao buscar por oficinas, o usuário terá como resultados categorias para oficinas de arte e oficinas mecânicas separadas;

⁹ Fonte pesquisada: TEIXEIRA, Paulo Rodrigo. **Marketing de Busca: O que são diretórios Web?**. 2007. Disponível em: <<http://www.marketingdebusca.com.br/o-que-sao-diretorios-web/>>. Acesso em: 13 nov. 2014.

3. Menor quantidade de sites irrelevantes. Como são catalogados por pessoas, a chance de haver conteúdo não relacionado ao que o usuário procura é muito menor.

Mas também existem desvantagens em relação às ferramentas de busca.

1. São mais lentas em atualização. Por dependerem de pessoas, demoram mais que os robôs das ferramentas de busca para serem atualizados;
2. Pode ainda não existir a hierarquia sobre o assunto. Com isto, fica difícil achar assuntos semelhantes;
3. Recuperam apenas sites das categorias.

Os diretórios têm uma importância muito grande no mercado de ferramentas de busca. Pois eles valem como base para diversas ferramentas de busca começarem suas buscas por links. Grandes ferramentas de busca, como o *Google*, utilizam do diretório *DMOZ* como apoio nas suas pesquisas. Fazem varredura no diretório de tempos em tempos, desta maneira conseguem manter suas bases de dados sempre atualizadas com sites novos.

Directory Mozilla (Dmoz), conhecido também por *Open Directory Project (ODP)* (em português: Projeto de Diretório Aberto), é um projeto que conta com a colaboração de voluntários que editam e categorizam páginas da *Internet*. Qualquer pessoa pode sugerir páginas que podem ser aprovadas ou não pelo editor equivalente da categoria de envio. Qualquer pessoa pode também se candidatar a editor do *Dmoz* (DMOZ, 2014).

Atualmente o *Yahoo!* integra sua busca e serviços patrocinados de busca com a *Microsoft*. Como parte dessa integração, o *Yahoo!* compartilha com a *Microsoft* a consulta de busca do usuário, endereço de IP, informações anônimas do navegador da *Web* (tais como um identificador exclusivo do navegador da *Web*) e outros dados anônimos (como dados demográficos, incluindo sexo e idade em anos inteiros). O *Yahoo!* formata os resultados fornecidos pelo serviço de busca *Bing* da *Microsoft* com o objetivo de oferecer ao usuário uma experiência de busca aprimorada (YAHOO!, 2014).

Ao fazer uma busca no Diretório do *Yahoo!* percebe-se que é um site muito interessante para quem busca informações sobre assuntos atuais, notícias, entretenimento, esportes, empregos, entre outros assuntos mais específicos. Por ser um Diretório, o *Yahoo!* categoriza hierarquicamente os assuntos em tópicos e este

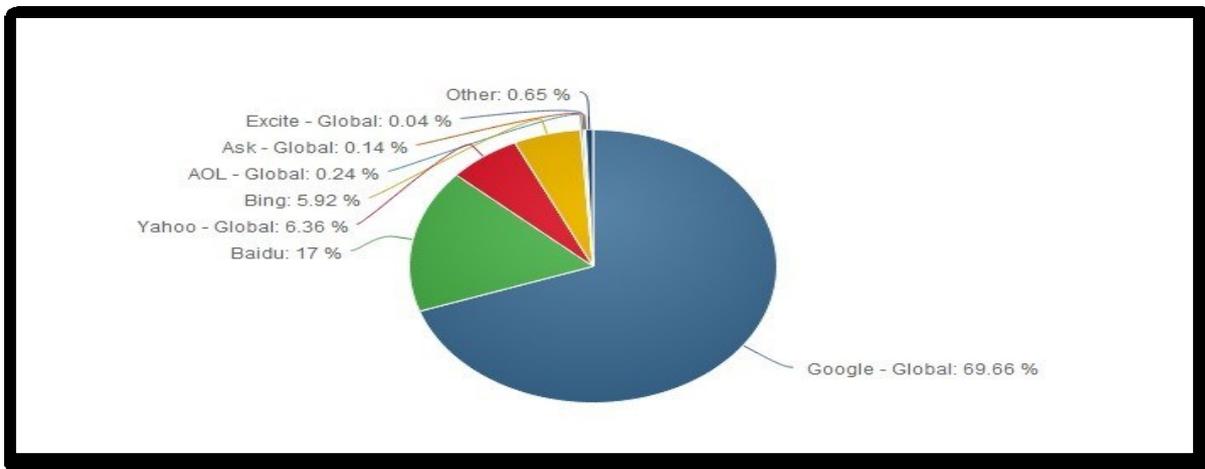
aspecto, facilita a busca na *Web*, pois o usuário pode partir direto para o assunto que lhe interessa. Outra vantagem percebida é a possibilidade de pesquisar por período que a informação foi publicada. Nesse sentido, dependendo do assunto a ser pesquisado, o usuário pode escolher o período a ser pesquisado de acordo com a sua necessidade de informação, seja ela mais recente ou não tão recente. Por outro lado, percebe-se algumas desvantagens. Os resultados podem ser tendenciosos, pois são classificados por pessoas, há menor quantidade de informações. Numa busca por assuntos mais específicos da área científica, por exemplo, o *Yahoo!* recupera poucas informações relevantes se for comparado ao *Google*.

b) Motores de busca

Motores de busca ou buscadores são conjuntos organizados de robôs que rastreiam a *Internet* em busca de páginas; índices e bases de dados que organizam e armazenam as páginas encontradas; e algoritmos para tratamento e recuperação das páginas. Eles permitem que seus usuários realizem buscas na *Internet*, principalmente através de palavras-chave. Exemplos de motores de busca são *Google*, *Yahoo!* e *Bing*.

Segundo pesquisas recente apresentada por Freire (2014) no site Tech Tudo, atualmente os mecanismos de busca mais conhecidos e utilizados em todo o mundo são: *Google*; *Baidu*; *Yahoo!*; *Bing*; *AOL*; *Ask*; *Other*; *Excite*. Conforme apresentação da Figura abaixo:

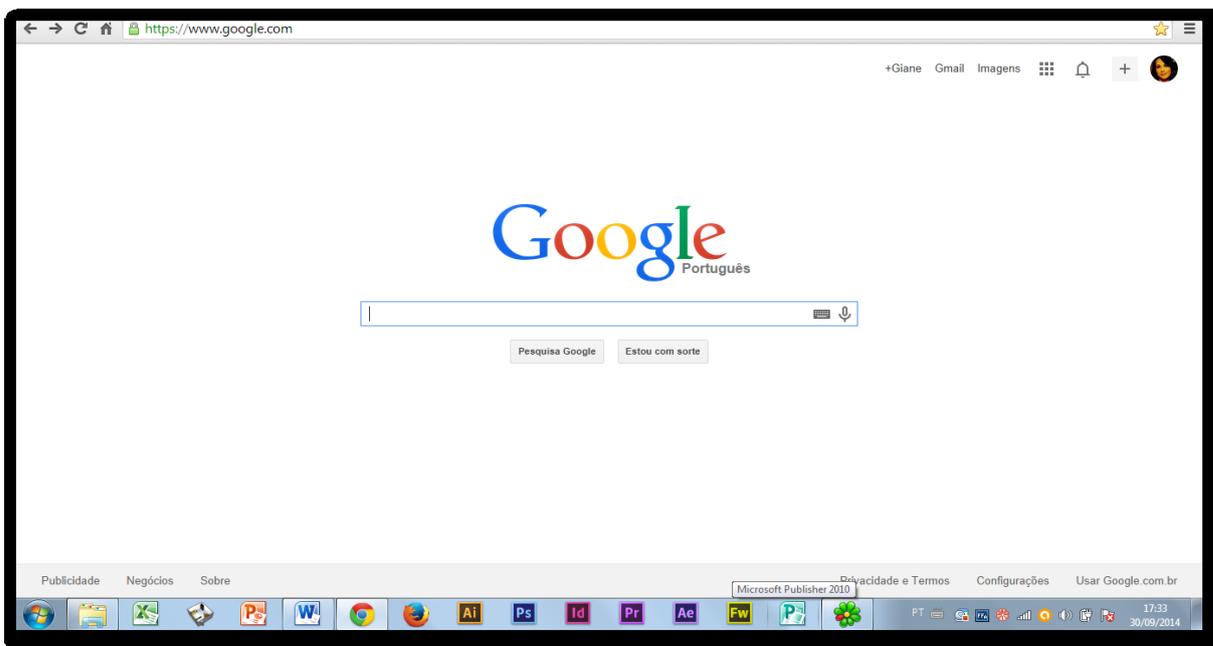
Figura 6 - Gráfico do *Netmarketshare* sobre uso mundial dos buscadores em 2014



Fonte: (Foto: Reprodução/Raquel Freire)

De acordo com o gráfico acima, o buscador *Google* tem o maior índice de utilização mundial dos buscadores atualmente. Por esta razão, esta pesquisa apresenta um detalhamento maior no serviço *Google*. A figura abaixo apresenta a página de acesso do buscador *Google*.

Figura 7 - Página de acesso Buscador *Google*



Fonte: (GOOGLE, 2014)

Conforme apresenta a figura acima, o buscador *Google* apresenta uma interface simples e clara. Provavelmente a combinação desses fatores ajudou o *Google* a ser o que é: o buscador mais utilizado em todo o mundo. Isso é tão notório que, hoje em dia, já existe um ditado popular muito utilizado quando alguém faz uma pergunta que é a seguinte frase: “joga no *Google*”, para responder a pergunta. Não parece, mas por trás da aparência simples da interface do *Google*, se esconde uma série de tecnologias. E essas tecnologias não surgiram de uma hora para outra, o *Google* foi se aperfeiçoando ao longo do tempo para tornar a busca mais efetiva.

O processo de pesquisa e a página de resultados envolvem muitos componentes, desta forma o *Google* atualiza constantemente suas tecnologias e sistemas para oferecer resultados melhores. Apesar de todos os avanços, o processo de busca do *Google* ainda deixa a desejar no aspecto semântico, pois nem sempre a recuperação da informação atende as necessidades do usuário em relação a sua intensão de busca.

Existem outros sistemas importantes que exigem ajustes e refinamentos constantes. Esta lista de projetos do *Google* oferece um vislumbre dos inúmeros aspectos da pesquisa no quadro a seguir:

Quadro 2 – Lista de projetos do buscador *Google*

Projetos	Aspectos da pesquisa
Respostas	Apresenta respostas imediatas e informações sobre assuntos como o tempo, resultados esportivos e fatos rápidos.
Preenchimento automático	Prevê o que o usuário pode estar procurando. Inclui a interpretação de termos com mais de um significado.
Livros	Encontra resultados em milhões de livros, incluindo visualizações e texto, em bibliotecas e editoras de todo o mundo.
<i>Google Instant</i>	Apresenta resultados imediatos enquanto o usuário digita.
Atualização	Mostra as últimas notícias e informações. Inclui a coleta de resultados adequados ao pesquisar por datas específicas.

Continuação...

Quadro 2 – Lista de projetos do buscador *Google* (continuação)

Projetos	Aspectos da pesquisa
Indexação	Usa sistemas para a coleta e o armazenamento de documentos na Web.
Imagens	Mostra resultados baseados em imagens com miniaturas para que o usuário possa escolher a página a visitar com apenas um olhar.
Painel do conhecimento	Fornecer resultados com base em um banco de dados de pessoas do mundo real, além de lugares, coisas e das conexões entre todos.
Celular	Inclui melhorias concebidas especificamente para dispositivos móveis como <i>tablets e smartphones</i> .
Notícias	Inclui resultados de jornais on-line e blogs de todo o mundo.
Interpretação de consultas	Extraí o significado mais profundo das palavras que o usuário digita.
Refinamentos	Oferece recursos como a pesquisa avançada, pesquisas relacionadas e outras ferramentas de pesquisa, todos destinados a ajudá-lo a refinar a pesquisa para o usuário.
SafeSearch (pesquisa segura)	Reduz a quantidade de páginas da Web, imagens e vídeos impróprios para menores nos resultados.
Métodos de pesquisa	Possibilita novas formas de pesquisar, como a pesquisa por imagem e a pesquisa por voz.
Qualidade do site e das páginas	Utiliza um conjunto de sinais para determinar a confiabilidade, reputação ou autoridade de uma fonte. Um desses sinais é o <i>PageRank</i> ¹⁰ , um dos primeiros algoritmos do Google, que examina os links entre as páginas para determinar sua relevância.

Continuação...

¹⁰ PageRank é uma família de algoritmos de análise de rede que dá pesos numéricos a cada elemento de uma coleção de documentos hiperligados, como as páginas da Internet, com o propósito de medir a sua importância nesse grupo por meio de um motor de busca.
<http://pt.wikipedia.org/wiki/PageRank>

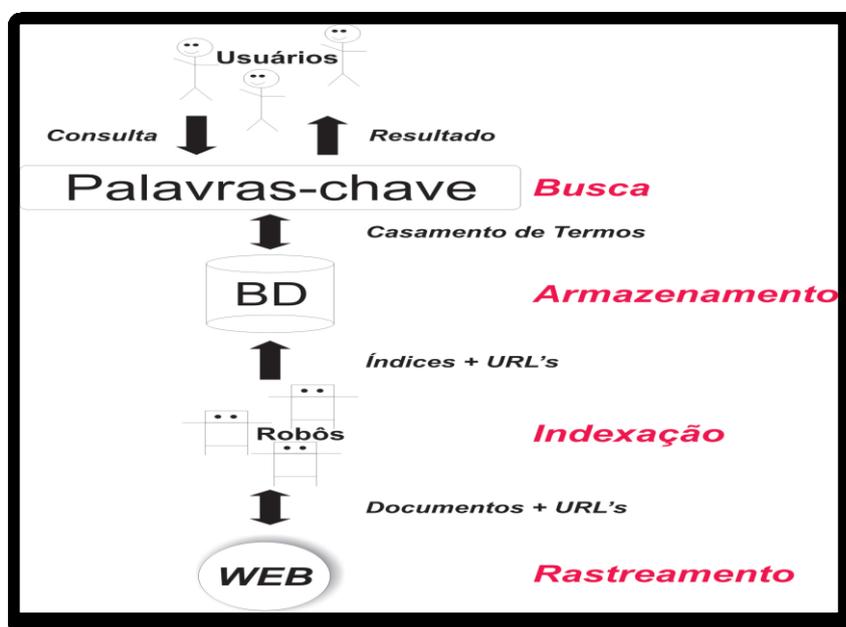
Quadro 2 – Lista de projetos do buscador *Google* (continuação)

Projetos	Aspectos da pesquisa
Snippets (Recorte)	Mostra pequenas visualizações de informações, como o título da página e um texto descritivo curto, para cada resultado da pesquisa.
Sinônimos	Reconhece palavras com significados semelhantes.
Ortografia	Identifica e corrige possíveis erros de ortografia e oferece alternativas.
Tradução e internacionalização	Refina os resultados de acordo com seu idioma e país.
Pesquisa universal	Mistura conteúdo relevante como imagens, notícias, mapas, vídeos e seu conteúdo pessoal em uma única página de pesquisa unificada.
Conteúdo de usuário	Fornecer resultados mais relevantes baseados na região geográfica e Histórico da Web, entre outros critérios.
Vídeos	Mostra resultados baseados em vídeo com miniaturas para que o usuário possa escolher rapidamente o vídeo a assistir.

Fonte: Por dentro da pesquisa (GOOGLE, 2014)

Quando um usuário realiza uma busca, o mecanismo de busca procura o termo em sua base de dados e fornece os resultados desta pesquisa ao usuário. Todas essas funções realizam-se em um site da *Web* conforme representação da figura abaixo:

Figura 8 - Visão geral dos mecanismos de busca



Fonte: (ALVES, 2005)

A busca informacional na *Web* aumentou muito nos últimos tempos. O grande número de usuários das ferramentas de busca pode ser explicado principalmente pelo fato destas ferramentas serem fáceis de usar, e serem geralmente rápidas no tempo de resposta, apesar de consultar milhões de páginas em seu banco de dados para selecionar um resultado. Além disso, a informação está muito mais acessível, pois pode ser acessada através de dispositivos móveis, *Smartphones*, *Tablets*, *Ultrabooks*, entre outros, que com certeza, geram facilidade e comodidade aos internautas que buscam a informação.

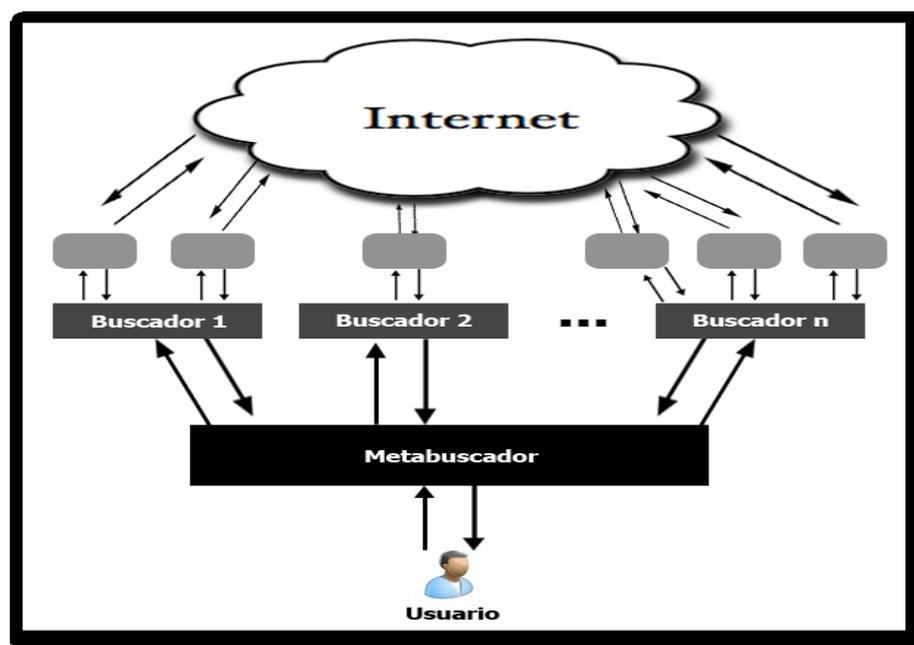
Porém, o desafio do usuário é encontrar a informação desejada e que realmente satisfaça a sua necessidade informacional. Por exemplo, numa simples pesquisa sobre “flor” (como planta) no *Google*, o resultado da busca apresentou o conceito e imagens de flor. No entanto, apresentou outras informações que não estavam de acordo com a intenção de busca, como por exemplo: nome de música com o termo “flor”, nome de floriculturas, etc. Percebe-se que apesar de todos os avanços, a recuperação da informação baseada na intenção do pesquisador ainda precisa melhorar.

Pelo fato das ferramentas de busca realizarem rastreamento e indexação de forma automática, estas conseguem armazenar grande quantidade de páginas. Isto, na maioria das vezes, significa mais informações sobre as quais seus usuários realizam pesquisas. A atualização das informações no seu banco de dados também é considerada mais eficiente.

c) Metabusador ou metapesquisador

É uma ferramenta de busca na *Internet* que se utiliza de outras para fornecer seus resultados. Portanto, ao se utilizar de um metabuscador o usuário estará pesquisando simultaneamente em diversos sites de busca, tais como o *Google* e o *Yahoo*, tornando mais fácil a localização do item procurado. A figura abaixo representa o processo de busca em um metabuscador:

Figura 9 – Processo de busca utilizando um metabuscador



Fonte: *wikimedia commons* (2011)

No quadro abaixo são apresentados alguns exemplos de metabuscadores, ferramentas recomendadas para pesquisas complexas e aprofundadas sobre um determinado tema.

Quadro 3 – Exemplos de Metabuscadores

Metabuscadores	Combinação de motores de busca
Dogpile (dogpile.com)	Pesquisa simultaneamente os bancos de dados do Google, do Yahoo! e do Bing.
Metacrawler (metacrawler.com).	A lista inclui <i>Google, Yahoo, MSN Busca</i> , entre outros. Os sites de busca incluídos na pesquisa por palavra chave são consultados um a um, através de suas próprias ferramentas de pesquisa e os principais resultados de cada um deles são apresentados ao usuário de forma prática, em uma única tela.
Sidekiq (sidekiq.com)	Compila uma quantidade impressionante de ferramentas, além de facilitar a pesquisa avançada do <i>Google</i> , pré-inserindo códigos no campo de busca que podem ajudar bastante o usuário a encontrar arquivos específicos, como apresentações, livros e músicas.
Whos Talkin (whostalkin.com)	Se a intenção é saber o que as pessoas estão falando em tempo real sobre um determinado tema, o <i>Who's Talkin (whostalkin.com)</i> é o endereço certo. Ele é capaz de realizar buscas em sites como <i>Twitter, WordPress, Facebook</i> e <i>MySpace</i> , entre dezenas de outros, e mostra apenas o que foi publicado no momento em que a pesquisa foi feita

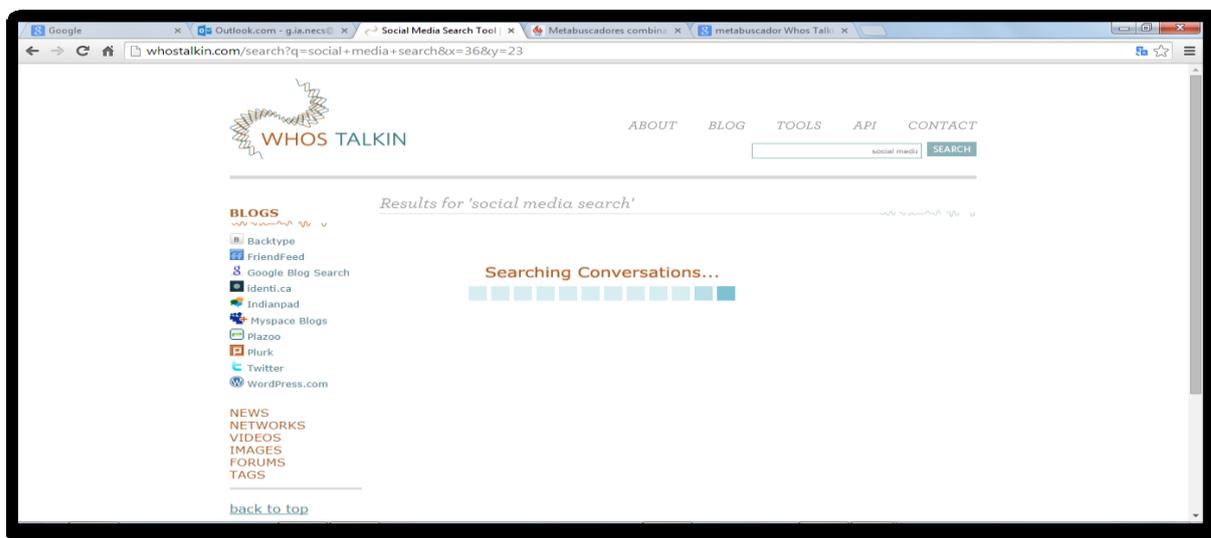
Fonte: (METABUSCADORES... 2010)

Ao se utilizar metamotores de busca, deve-se levar em conta as diferentes estratégias de busca de cada um desses mecanismos. Pois muitas vezes, a ordem dos resultados é apresentada ignorando a relevância da obra e é exibida de acordo com outros critérios. Obtêm-se uma maior cobertura, porém, os resultados são pouco precisos.

Os metabuscadores são pouco utilizados em relação aos motores de busca e os diretórios. Não se sabe exatamente o porquê, no entanto, acredita-se que muitos usuários desconhecem a existência dessas ferramentas específicas de busca na *Web* ou não sabem utilizá-las.

A figura abaixo representa a página de acesso do metabuscador *Whos Talkin*:

Figura 10 – Página de acesso do metabuscador *Whos Talkin*



Fonte: Whos Talkin (2014)

Conforme representação da figura acima, a metabusca ou metapesquisa é um sistema de busca na *Web* que permite ao usuário a pesquisa em vários mecanismos simples de busca ao mesmo tempo. Os metabuscadores funcionam da mesma forma que outros motores de busca, com a vantagem de padronizar a pesquisa do usuário, tornando-a mais abrangente e mais rápida, por expandir o número de fontes buscadas. E são muito eficientes na busca por termos muito específicos e difíceis de localizar.

4.1.1 Buscadores Semânticos

O desenvolvimento da *Web Semântica* é impulsionado por tecnologias que melhoram consideravelmente a recuperação da informação. Neste sentido, a busca semântica é uma aplicação da *Web Semântica* para melhorar o resultado da pesquisa. A busca de informações baseadas na intenção do pesquisador e significado contextual dos termos de pesquisa é definida como pesquisa ou busca semântica. (JONH, 2012, tradução nossa).

Um ser humano pode compreender a intenção e o contexto de uma pergunta, mas os motores de busca e *softwares* de sistemas a princípio não foram desenvolvidos com essa capacidade. No entanto, os motores de busca estão se aperfeiçoando cada vez mais para atenderem as necessidades dos pesquisadores, apresentando resultados mais relevantes e úteis. Estão se tornando mais inteligentes.

A busca semântica faz parte, na atualidade, das funcionalidades dos motores de busca atuais, enquadrados na terceira Geração da *Web*, “onde as informações são recuperadas com base na **semântica** das informações contidas nos documentos, ou seja, os dados são recuperados com base em informações estruturadas semanticamente” (BARRETO, 2008, p. 3, grifo nosso).

A Marcação Semântica dos dados é apontada por Souza (2006) como uma aplicação importante na melhoria dos sistemas de recuperação da informação na *Web* porque proporciona melhor representação dos documentos, torna as informações legíveis por máquinas, melhora o agrupamento dos documentos. Os melhores exemplos desta vertente são:

as tecnologias exploradas no contexto da *Web Semântica*, com vistas ao projeto e à implementação de padrões de metadados, que adicionem aos dados informações significativas sobre seus contextos, marcando-os semanticamente; e mecanismos de busca que levem em conta estes dados marcados (SOUZA, 2006, p.171).

Google e o *Bing* são exemplos de buscadores que estão se desenvolvendo para possibilitar resultados de pesquisa com base no significado semântico dos termos (JONH, 2012, tradução nossa).

Numa pesquisa tradicional, quando realizamos a busca de um termo no *Google*, obtemos uma sequência de diversos resultados com temas relacionados a ele, mesmo que a nossa intenção seja apenas encontrar a definição do termo. Na pesquisa semântica, essa busca retornaria resultados mais precisos, melhorando consideravelmente o resultado da pesquisa. Para que isso ocorra, de acordo com John (2012, tradução nossa), um motor de busca inteligente deve considerar vários fatores para fornecer os resultados mais relevantes e úteis. Por exemplo, considerando o fator atualidade: se a Copa do Mundo estivesse se encerrando hoje e alguém estivesse pesquisando ‘qual país venceu a Copa do Mundo?’, o sistema de busca semântica deveria ser capaz de compreender a consulta e dar resultados relevantes baseado em notícias atuais. Outro exemplo, baseado no fator sinônimo:

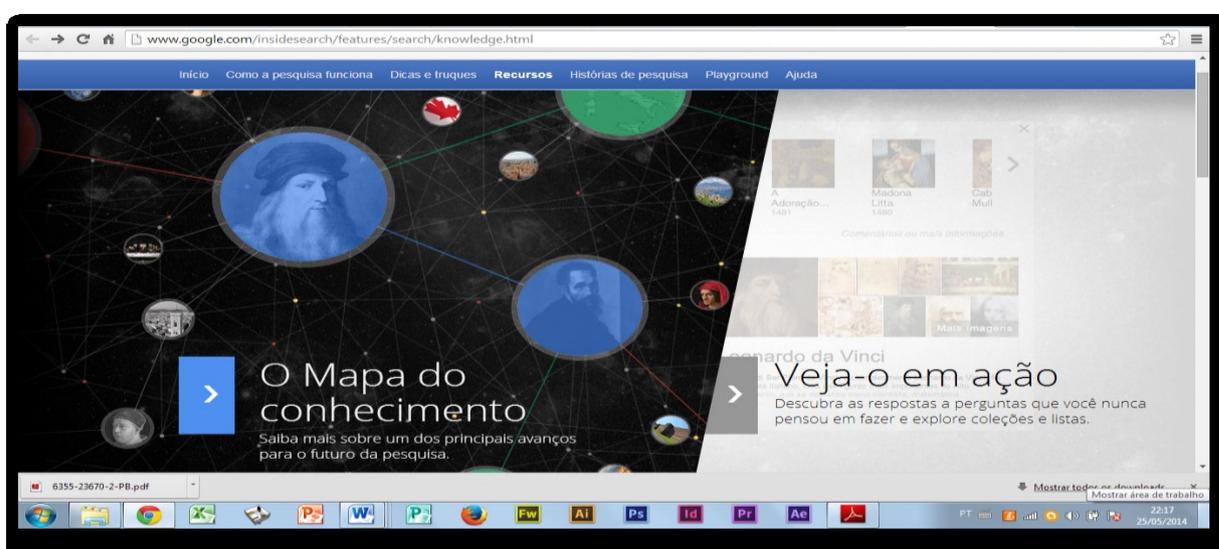
um motor de busca semântico deve ser capaz de compreender os sinônimos e dar mais ou menos os mesmos resultados de pesquisa para a palavra que possui sinônimos que o usuário procurar. Por exemplo, ao tentar procurar por 'maior montanha' e 'montanha mais alta', o usuário poderia obter praticamente os mesmos resultados uma vez que ambos têm o mesmo significado nesta consulta em particular, mesmo que "maior" e "mais alta" possa significar coisas diferentes em diferentes casos.

Dentre os buscadores focados em processos semânticos, destaca-se o “*Google Knowledge Graph* (Mapa do conhecimento), lançado recentemente nos Estados Unidos”. (SIQUEIRA, 2013, p.60). O Mapa do Conhecimento do *Google* baseia-se em componentes de Inteligência Artificial. Siqueira (2013, p.60) explica como funciona o processamento semântico das informações do Mapa do Conhecimento:

A partir dos termos de busca, a metodologia irá subtrair ‘entidades’, compará-las às armazenadas nas suas bases de dados, e disponibilizar como resultado as informações relevantes. O diferencial é que as palavras não são tomadas isoladamente em fragmentos de textos e links, mas correlacionadas à semântica dos significados que perfazem a entidade. Isso significa que o conceito de entidade implica a sua percepção existencial enquanto um objeto ligado aos seus atributos de espaço, data, local, cultura, bem como seus relacionamentos com outras entidades.

A figura 10 apresenta uma imagem do Mapa do Conhecimento do *Google*:

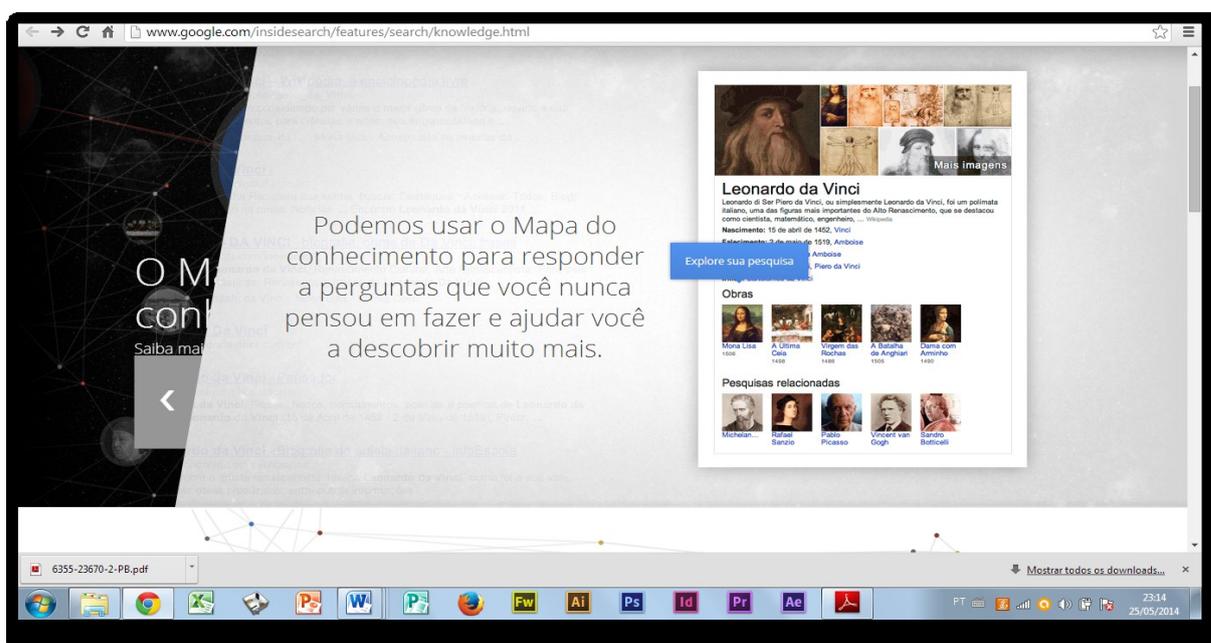
Figura 11 - Mapa do Conhecimento



Fonte: (Google, 2014).

O objetivo do Mapa do conhecimento é construir um enorme mapa de elementos do mundo real e as conexões entre eles, de modo a oferecer resultados mais relevantes para o usuário. O mapa do conhecimento envolve a coleta de informações sobre objetos do mundo real. Esses objetos podem ser pessoas, livros, filmes e muitas outras coisas. Por exemplo, se o interesse da pesquisa for pesquisar sobre os pintores mais importantes do Renascimento. Porém, o usuário só conhece Leonardo Da Vinci, então ele poderá iniciar a sua pesquisa por 'Da Vinci'. O resultado da pesquisa apresentará informações que o ajudarão a explorar o tópico mais amplo de pintores renascentistas. Verá algumas das obras mais famosas da era, como a Mona Lisa, e descobrirá outros pintores da época, como Michelangelo e Rafael. O resultado da pesquisa desse exemplo está representado na figura abaixo:

Figura 12 - Resultado da pesquisa no Mapa do Conhecimento.



Fonte: (Google, 2014)

Além do Mapa do Conhecimento do *Google* existem outros **buscadores semânticos**¹¹. A seguir são apresentados alguns deles:

¹¹ BUSCADORES SEMÂNTICOS: Serviços on-line que tentam 'interpretar' os termos que o usuário procura. Folha de S. Paulo. 2012. Disponível em: <<http://www1.folha.uol.com.br/fsp/tec/34737-buscadores-semanticos.shtml>>. Acesso em: 27 out. 2014.

- a) *Wolfram Alpha* (wolframalpha.com) - uma das fontes de pesquisa do Siri, assistente virtual do *iPhone 4S*, tem como ponto forte a resolução de problemas matemáticos. Se a pesquisa for feita por um termo comum, ele gera um resumo utilizando como base outros sites, como o *IMDB* e a *Wikipédia*.

Este buscador semântico é muito interessante para buscar dados estatísticos. Ao fazer uma busca na sua caixa de pesquisa com a frase 'qual a idade média da população brasileira?', percebeu-se que o buscador interpretou exatamente aquilo que se desejava pesquisar, ou seja, inferiu a intenção de busca. Desta forma, apresentou como resultado o número que corresponde à idade média da população brasileira (uma estimativa). Além deste resultado, apresentou outros números relativos à pesquisa (população, densidade populacional, expectativa de vida, crescimento populacional) como mostra a figura abaixo:

Figura 13 – Resultado da busca semântica no Wolfram Alpha

population	202 million people (world rank: 5 th) (2014 estimate)
population density	23.8 people/km ² (people per square kilometer) (world rank: 188 th) (2010 estimate)
population growth	0.852 %/yr (world rank: 141 st) (2013 estimate)
life expectancy	73.9 years (world rank: 116 th) (2013 estimate)
median age	28.6 years (world rank: 105 th) (2009 estimate)

Fonte: (WOLFRAM ALPHA, 2014)

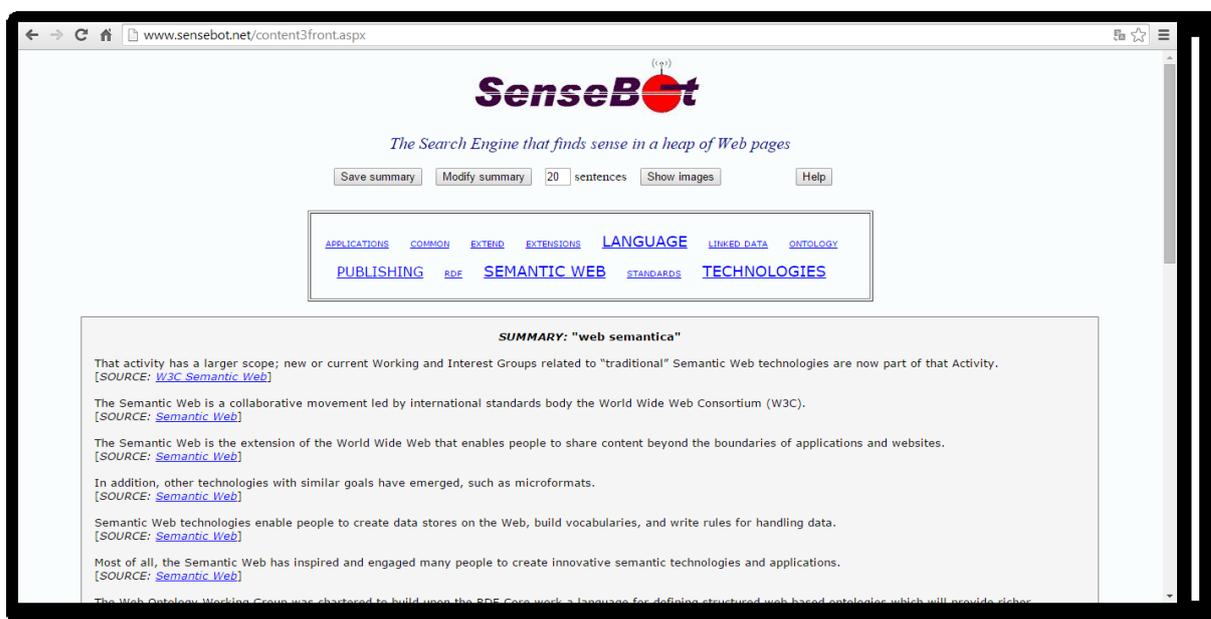
Além disso, em algumas pesquisas a resposta também apresenta gráficos. O buscador também apresenta abaixo de sua caixa de pesquisa uma guia chamada **exemplos** que relaciona tópicos que podem ser pesquisados como, por exemplo: Unidades e Medidas, Tempo e Meteorologia, Alimentos e Nutrição, Saúde e Medicina, Dados Socioeconômicos, entre outros. Esses tópicos auxiliam a pesquisa, pois o usuário pode iniciar a sua busca partindo logo para o assunto que lhe interessa.

- b) *SenseBot* (sensebot.net) - usa os motores de busca do *Yahoo!*, do *Google* ou do *Bing*. Ele usa mineração de texto para analisar páginas

da *Web* e identificar seus conceitos semânticos fundamentais. Em seguida, realiza sumarização de conteúdo para produzir um resumo coerente;

Ao fazer uma busca sobre '*Web Semântica*', o buscador apresentou vários resumos sobre o tema da pesquisa. O resumo apresentou-se como uma sequência de frases. Cada frase continha uma referência (um *link*) à página *Web* de origem que foi encontrado. A partir dos resumos, elegeu-se o documento que melhor atendia a intensão de busca. O buscador permite salvar o resumo em um arquivo de texto ou ainda modificar o resumo após alterar alguns parâmetros na página. Também exibe miniaturas de imagens e vídeos relevantes no topo do resumo. Outra vantagem interessante é que o buscador apresenta uma **nuvem semântica** dos conceitos acima do sumário, que permite orientar o foco dos resultados conforme ilustra a figura abaixo.

Figura 14 – Tela do Sensebot com a nuvem semântica



Fonte: (SENSEBOT, 2014)

Tendo em vista a explosão informacional, os mecanismos de busca estão potencializando seus recursos na solução de problemas, com a finalidade de melhorar a qualidade do acesso e do tempo envolvido na busca informacional.

As técnicas mais comuns utilizadas na organização da informação pelas ferramentas de busca como a indexação de palavras-chave, a análise dos links das

páginas e o *PageRank* abrem espaço para novas tecnologias. Neste sentido, as pesquisas se intensificam com ênfase para o processamento semântico da informação, “com potencialidades para buscas e integração de dados em sistemas de recuperação de informação, desenvolvimento de padrões de análise, modelos de pergunta e resposta, e esquemas” (GRUBER, 1993, p. 5).

Juntamente com a promessa do Mapa do Conhecimento, o futuro dos mecanismos de pesquisa, busca e recuperação da informação na *Web* de acordo com Siqueira (2013, p. 63) “provavelmente estará no aprofundamento das possibilidades funcionais de relacionamentos, da utilização de modelos de inteligência artificial e dos pressupostos da *Web Semântica*”.

Os buscadores se tornaram ferramentas imprescindíveis na obtenção de informações na rede. Principalmente pelas facilidades que as tecnologias oferecem, os buscadores representam um papel importante na qualidade do acesso e do tempo envolvido na busca informacional.

4.2 Técnicas que facilitam a Busca na Web

a) Recursos do *Google* ¹²

1. Organização de informações por meio da indexação ¹³

A *Web* é como uma biblioteca pública em constante expansão, com bilhões de livros e nenhuma administração centralizada. Essencialmente, o *Google* reúne as páginas durante o processo de rastreamento e cria um índice para que o usuário saiba exatamente como encontrar o que procura. De forma muito semelhante ao índice na parte de trás de um livro, o índice do *Google* inclui informações sobre as palavras e onde podem ser encontradas. Quando o usuário realiza uma pesquisa no nível mais básico, os algoritmos do *Google* procuram os termos de sua pesquisa no índice para localizar as páginas adequadas.

Conforme descrito no site do *Google* (2014), o processo de pesquisa torna-se muito mais complexo a partir daí. Quando o usuário procura por “cães”, ele

¹² GOOGLE. Por dentro da pesquisa. Disponível em:
<<http://www.google.com.br/insidesearch/>>.

Acesso em: 22 ago. 2014.

¹³ Indexação é o tratamento dado a uma página rastreada, antes que esta seja armazenada. Este tratamento irá auxiliar na futura localização desta página pelo mecanismo de busca. (MORAIS; AMBROSIO, 2007, p. 12)

não quer ver uma página com a palavra "cães" escrita centenas de vezes. Ele provavelmente quer imagens, vídeos ou uma lista de raças. Os sistemas de indexação do *Google* levam em consideração vários aspectos diferentes das páginas, como sua data de publicação, seu conteúdo em termos de fotos e vídeos, e muito mais.

Com o **Painel do conhecimento**, o *Google* vai além da mera associação de palavras-chave para compreender melhor as pessoas, lugares e coisas que importam para o usuário.

2. Localização de informações por meio da indexação

O *Google* usa programas conhecidos como "rastreadores da *Web*" para descobrir páginas disponíveis publicamente. O rastreador mais conhecido é chamado de "*Googlebot*". Os rastreadores analisam as páginas da *Web* e seguem os links contidos nelas, como o usuário faz ao navegar na *Internet*. Eles avançam de link em link e transmitem aos servidores do *Google* os dados destas páginas da *Web*.

O processo de rastreamento começa com uma lista de endereços da *Web* de rastreamentos anteriores e mapas de sites fornecidos pelos proprietários de *Websites*. Ao acessar esses *Websites*, os rastreadores do *Google* procuram por links para outras páginas a visitar. O programa dá atenção especial a estes novos sites, a alterações em sites existentes e a *links* inativos.

Programas de computador determinam os sites a rastrear, a frequência do rastreamento e o número de páginas a buscar em cada site.

3. Algoritmos

Existem milhares ou milhões de páginas com informações úteis para uma consulta típica. Os algoritmos são os processos e fórmulas de computador que transformam as perguntas em respostas. Atualmente, os algoritmos do *Google* utilizam mais de 200 sinais ou "pistas" diferentes para adivinhar o que o usuário realmente procura. Esses sinais incluem coisas como os termos em *Websites*, a atualização do conteúdo, a região do usuário e o *PageRank*.

b) Marcação semântica dos dados na origem

As tecnologias exploradas no contexto da *Web* semântica, voltadas para o projeto e à implementação de padrões de metadados, são os melhores exemplos deste seguimento.

O desenvolvimento de novas possibilidades de marcação semântica dos dados utilizando-se metalinguagens, criando registros de metadados acoplados aos próprios documentos com termos amplamente consensuais e não ambíguos, para que esses possam ser mais facilmente manipulados e identificados por computadores e outros dispositivos e, como consequência, pelos usuários (SOUZA, 2006, p. 168)

A finalidade dessa tecnologia é fazer com que o buscador entenda da mesma forma que nós humanos entendemos o mundo.

4.3 Papel do Bibliotecário

Atualmente o mundo vivencia grandes mudanças em termos sociais políticos e culturais. O *boom* da informação na *Internet* e as novas tecnologias da informação e comunicação fez surgir à sociedade da informação, “pautada pela produção, disseminação e uso dos conteúdos informacionais” (OLIVEIRA; SILVEIRA, 2010, p. 2).

O progresso tecnológico é evidente e de acordo com Silva e Cunha (2002, p. 77) “atua, principalmente, como facilitador no processo comunicacional. Agora é possível processar, armazenar, recuperar e comunicar informação em qualquer formato, sem interferência de fatores como distância, tempo ou volume.”

Conseqüentemente a sociedade da Informação apresenta mudanças para todos os setores e para acompanhar essas transformações econômicas e sociais as organizações e os profissionais de diversas áreas necessitam de atualização e capacitação, aliando de forma positiva a sua rotina, essas mudanças.

Nesta conjuntura, em que a mudança tecnológica é a regra, buscar condições para ancorar a preparação do profissional do futuro requer uma estratégia diferenciada. Este profissional deverá interagir com máquinas sofisticadas e inteligentes, será um agente no processo de tomada de decisão. Além disso, o seu valor no mercado será estimado com base em seu dinamismo, em sua criatividade e em seu empreendedorismo. Todos esses fatores evidenciam que só a educação será capaz de preparar as pessoas para enfrentar os desafios dessa nova sociedade (SILVA; CUNHA, 2002, p. 78).

No contexto bibliotecário, a inserção das TICs tem provocado o surgimento de novas formas de trabalho, novos serviços, como as bibliotecas digitais, tendo como alvo um novo tipo de usuário. Consequentemente os profissionais da área, ou seja, os bibliotecários são desafiados a mudar seu perfil e sua forma de atuação assim como atualizar-se constantemente para acompanhar as mudanças atuais. O papel mais importante do bibliotecário na atualidade parece ser o de gerenciador da informação. A importância dessa tarefa pode ser assim comentada:

o grande problema desse século é a superabundância de informação. Então, se não possuímos sistemas e estratégias adequadas de acesso à informação ou estivermos despreparados para acessá-las, de que servirá tanta informação? Do que servirá a tecnologia, se a maioria das pessoas não saberá utilizá-la ou não terá acesso a elas? Os computadores e os sistemas inteligentes de processamento de dados podem até assumir parte dessa tarefa. No entanto, a organização e a manipulação de toda essa informação requerem instruções, e aqui é que o bibliotecário poderá contribuir (SILVA; CUNHA, 2002, p. 78).

A maneira como as pessoas interagem, trabalham e se organizam é influenciada pela utilização das tecnologias da informação. Nessa perspectiva o bibliotecário dispõe de um novo espaço de trabalho: a *Web*, que requer modificações nas formas de organização do conhecimento e tratamento da informação.

Considera-se, portanto, que um profissional formado em Biblioteconomia deve ter um ótimo conhecimento das práticas da Organização da Informação e saber relacioná-las com o as novas técnicas de Representação da Informação e do Conhecimento que estão sendo desenvolvidas na *Web*. Neste sentido, acredita-se que um dos caminhos para o aprimoramento da formação dos profissionais bibliotecários é a oferta de disciplinas voltadas para este saber, no sentido de proporcionar a capacitação destes profissionais para atuar na aplicação dos conceitos e práticas da Organização da Informação na *Web*.

A propagação do uso das tecnologias a serviço da informação transpõe barreiras físicas e institucionais. A necessidade de familiarizar-se com os mecanismos virtuais existentes tornou-se natural tanto quanto saber a localização de um livro na estante da biblioteca e o uso da *Internet* passou a ser um instrumento obrigatório de trabalho do bibliotecário (SILVA, 2005).

É importante ressaltar que o bibliotecário continua tendo sua importância no tratamento e organização da informação. Bem como na recuperação, disponibilização e disseminação da informação. O caráter do trabalho do profissional bibliotecário não mudou. A transformação está relacionada às novas práticas impostas pela aplicação das técnicas e ferramentas que fazem parte do desenvolvimento da *Web*. Portanto, não se trata de um novo profissional, mas de um bibliotecário com mais qualificação profissional, com novas habilidades e competências para atuar com eficiência neste novo cenário.

Não é possível afirmar que o mercado de trabalho atual está necessitando ou requisitando este profissional da informação com novas habilidades e competências. Seria necessário fazer uma pesquisa de mercado para fornecer a resposta. No entanto, é evidente que as mudanças estão acontecendo. Portanto, é necessário acompanhá-las e adaptar-se a elas para não ficar despreparado. Fazendo uma analogia com a Biologia, é como na lei de sobrevivência do mais apto, vence aquele que melhor se adapta as mudanças e está mais bem preparado para os desafios do mercado de trabalho.

Segundo Oliveira e Silveira (2010, p.8),

As oportunidades convergem para prestação de novos serviços de informação ao usuário através da *Web* tais como: acessibilidade universal de material; padrão iniciado de serviço inter-bibliotecário; lista de livros e leitura online, base de dados local; balcão de referência virtual; passeio virtual; formulários web; catalogação cooperativa e apoio a educação a distância.

Neste caso, o bibliotecário formado precisa estar em constante atualização com as novas tecnologias, com as novas práticas do fazer bibliotecário e com as ferramentas de busca na *Web*. Podem-se citar algumas ações neste sentido:

- a) Educação continuada na instituição e à distância (cursos de atualização na área);
- b) Aquisição de mais um idioma;
- c) Domínio das tecnologias da informação;
- d) Formação de mestres e doutores;
- e) Acompanhamento do estágio de desenvolvimento da *Web Semântica* e suas aplicações.

Teotônio (2011) observa que o bibliotecário não deve temer os serviços das ferramentas de busca, deve reconhecer a mudança do universo informacional e

modificar os serviços com base na opinião dos usuários. Além disso, os bibliotecários devem entender que o futuro das bibliotecas será guiado pelo fato de que seus usuários acessam, consomem e criam conteúdos.

O cenário atual se caracteriza por usuários diversos e independentes que utilizam muitos recursos tecnológicos, como *laptops*, *smartphones*, *tablets*, entre outros dispositivos. O que mais se evidencia no perfil dos novos usuários é o fato deles estarem cada vez mais satisfeitos com formas alternativas de busca e acesso da informação na *Internet*. Dessa forma, percebe-se que os usuários diminuíram o uso dos recursos da biblioteca para desenvolver suas pesquisas. O problema não é o acesso ao recurso físico da biblioteca, ou o fato de ir até ela, mas até que ponto a *Internet* é utilizada adequadamente para busca, recuperação ou comunicação da informação.

A quantidade de informações disponível nunca foi tão grande. Conforme Cunha (2003, p. 43) “Apesar de todos os sistemas de gestão desenvolvidos, dispomos atualmente de poucos instrumentos realmente eficazes para filtrar a informação pertinente, no momento certo, no enorme fluxo disponível no ciberespaço.” A quantidade de informação disponível na *Web* sugere a necessidade dos bibliotecários serem intermediários em alguns casos, analisando e avaliando as fontes, de forma a conciliar necessidades com recursos.

Tendo em vista a grande dificuldade de se recuperar informação precisa e confiável na *Web*, surge a *Web Semântica*, propondo uma revolução, um novo modelo de representação da informação e do conhecimento. Nessa perspectiva, o bibliotecário com seus conhecimentos adquiridos relacionados à representação descritiva podem contribuir para o seu desenvolvimento, pois as técnicas de representação da informação utilizadas pela biblioteconomia podem colaborar como fundamento para o desenvolvimento de estudos relacionados às tecnologias criadas no desenvolvimento da *Web Semântica*. Assim como “o conhecimento dos bibliotecários no gerenciamento de metadados os torna importantíssimos parceiros nos grupos de desenvolvimento de padrões da *Web Semântica*” (ALVAREZ, 2014, p. 4224).

Contudo, o grande desafio dos bibliotecários no atual cenário é a constante atualização e familiarização com as novas tecnologias aplicadas no desenvolvimento da *Web* que permitam ao profissional oferecer novos serviços e acompanhar a demanda da sociedade neste contexto.

5 CONCLUSÕES

O objetivo da *Web Semântica* é solucionar o problema da falta de informação com significado a respeito do conteúdo organizado na *Web* através de novas ferramentas e tecnologias. Nos últimos anos, houve muito progresso no desenvolvimento de técnicas e ferramentas para permitir o crescimento da *Web Semântica*, conforme foi apresentado no capítulo 2. Algumas tecnologias já estão sendo implantadas na *Web*, tais como, o XML - uma linguagem de marcação que permite incluir informações significativas a respeito de palavras ou termos do documento, o RDF - um modelo padrão para o intercâmbio de dados, que padroniza a criação de metadados na *Web* permitindo maior intercâmbio e reutilização de descrições de termos e palavras, e as Ontologias que disponibilizam um vocabulário específico de conhecimento e descrevem os termos e seus relacionamentos.

Porém, a *Web Semântica* enfrenta alguns desafios no seu desenvolvimento. As ontologias, por exemplo, são apontadas como uma das principais dificuldades encontradas no processo de implantação da *Web Semântica*. O problema com as ontologias é que são muito difíceis de criar, implementar e manter. No caso do *FRBR*, por exemplo, que é uma ontologia da Biblioteconomia, para sua implantação, entende-se que é necessário modelar os dados bibliográficos já existentes que se utilizam de *MARC* ou *Dublin Core*, por exemplo, nos moldes do RDF. Deve-se, ainda, tornar o *FRBR* e o *RDA*, em base *RDF*. Esse processo certamente envolve muito tempo e esforço dos profissionais da área.

Outro grande desafio está relacionado à estruturação dos dados na *Web*, pois apesar de crescer exponencialmente a cada dia, a *Web* atual apresenta-se de forma desestruturada. Isso ocorre porque grande parte das páginas disponíveis na *Web* é direcionada para troca entre pessoas e não para serem processadas automaticamente. Para tornar a *Web* estruturada é importante ter uma grande quantidade de dados disponível em um formato padrão, acessível e gerenciável por ferramentas da *Web Semântica*. Além do acesso aos dados da *Web*, é necessário, ainda, criar relações entre os dados. E essa é uma tarefa que exige muito tempo e empenho de grupos de estudos como o W3C, instituições e usuários envolvidos no projeto da *Web Semântica*.

Apesar destes desafios, algumas ferramentas de busca estão se aperfeiçoando para tornar suas buscas mais precisas. O *Google*, por exemplo, já

lançou o buscador semântico “Mapa do Conhecimento” fundamentado em inteligência artificial para fornecer respostas baseadas na intensão de busca do pesquisador e tornar a recuperação da informação mais eficiente.

A grande quantidade de informações disponível na *Web* organizada de forma inadequada implica a necessidade dos bibliotecários serem intermediários da informação. O papel mais importante do bibliotecário na atualidade parece ser o de gerenciador da informação.

O bibliotecário com seus conhecimentos adquiridos podem contribuir para o desenvolvimento da *Web Semântica*, pois as técnicas de representação da informação utilizadas pela biblioteconomia podem colaborar como fundamento para o desenvolvimento de estudos relacionados às tecnologias criadas no seu desenvolvimento. No entanto, é imprescindível que estes profissionais se mantenham atualizados e comecem a se familiarizar mais com as técnicas e ferramentas que estão sendo criadas e aplicadas no processo de desenvolvimento da *Web Semântica* para não ficarem despreparados para atender as demandas da sociedade.

É fato que a maioria dos usuários tem dado prioridade às buscas na *Web*, acessando cada vez menos os tradicionais catálogos das bibliotecas para atender suas necessidades informacionais. Com a idealização da *Web Semântica*, espera-se que futuramente os usuários possam acessar os mecanismos de busca e ter acesso aos catálogos das bibliotecas através da vinculação dos dados na *Web*. Da mesma forma, acredita-se que as buscas se tornem mais simples e precisas. Que os computadores interpretem os resultados e refinem a busca, tirando a responsabilidade dos usuários de fazer este papel e poupando seu tempo. Neste sentido, tanto o trabalho do profissional da informação, quanto a recuperação das informações pelos usuários serão otimizados.

O implemento da *Web Semântica* começou recentemente e certamente ainda vai levar mais alguns anos até que atinja o seu ideal. Portanto, ainda há um longo caminho a percorrer antes que a idealizada *Web Semântica* se torna realidade.

Espera-se que essa pesquisa tenha contribuído para com os profissionais da Ciência da Informação por proporcionar um referencial teórico sobre o tema *Web Semântica* e suas tecnologias, que se apresentam como caminhos para atingir o

objetivo de proporcionar uma busca mais efetiva e a recuperação semântica dos recursos informacionais na *Web*.

REFERÊNCIAS

ALVAREZ, Edgar Bisset. Bibliotecas e Web Semântica: análise sobre o estado deste relacionamento. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, Belo Horizonte. **Anais...** . Belo Horizonte: UFMG, 2014. p. 4219 - 4227. Disponível em:

<<http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>>. Acesso em: 07 nov. 2014.

ALVES, Rachel Cristina Vesú. **Web Semântica**: uma análise focada no uso de metadados. 2005. 180 f. Dissertação (Mestrado) - Curso de Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005. Disponível em:

<http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/alves_rcv_me_mar.pdf>. Acesso em: 30 ago. 2014

ANZOLIN, Heloisa Helena. Atualizações em AACR2. In: ENCONTRO NACIONAL DOS USUÁRIOS DA REDE PERGAMUM, 9., 2007, Curitiba. **Anais...** . Curitiba: Pontifícia Universidade Católica do Paraná, 2007. p. 1 - 59. Disponível em:

<http://cdij.pgr.mpf.mp.br/sistema-pergamum/ix-encontro-nacional/20_04_2007/Curso AACR2.pdf>. Acesso em: 21 abr. 2014.

ASSUNÇÃO, Fabricio Silva. **O que é FRBR?** 2012. Disponível em:

<<http://fabricioassumpcao.com/2012/07/o-que-e-frbr.html#sthash.8K9FlaMZ.dpuf>>. Acesso em: 13 ago. 2014.

BARRETO, Luís José Frederico. **Protótipo de Procura Semântica aplicado a um Motor de Busca Documental**. 2008. 102 f. Dissertação (Mestrado) - Curso de Engenharia Eletrotécnica e de Computadores Maior Automação, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, 2008. Disponível em:

<<http://repositorio-aberto.up.pt/bitstream/10216/61861/1/000142705.pdf>>. Acesso em: 23 maio 2014.

BERNERS-LEE, Tim. **Information Management**: A proposal. CERN: 1989.

Disponível em: <<http://www.w3.org/History/1989/proposal.html>>. Acesso em: 29 mar. 2014.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic Web.

Scientific American, 2001. Disponível em:

<www3.unisa.it/uploads/3845/semantic_web-berners_lee.pdf>. Acesso em: 29 mar. 2014.

BREITMAN, Karin Koogan. **Web Semântica**: a Internet do futuro. Rio de Janeiro: LTC, 2006. xvii, 190 p.

CATARINO, Maria Elisabete; SOUZA, Terezinha Batista de. A representação descritiva no contexto da Web Semântica. **Transinformação**. 2012, vol.24, n.2, pp. 77-90. Disponível em : <<http://www.scielo.br/pdf/tinf/v24n2/a01v24n2.pdf>>. Acesso em: 18 abr. 2014.

CAMPBELL, D. Grant. 2011. RDA e RDF: Uma análise do discurso de dois padrões de descrição de recursos. Em Smiraglia, Richard P., ed. **Proceedings da North American Symposium on Organização do Conhecimento**, vol. 3. Toronto, Canadá, p. 207-16. Disponível em: <<http://journals.lib.washington.edu/index.php/nasko/article/download/12804/11285>>. Acesso em: 08 mai. 2014.

CUNHA, Miriam Figueiredo Vieira da. O papel social do bibliotecário. **Encontros Bibli**. 2003, v8, n. 15, p. 41-46. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2003v8n15p41>>. Acesso em: 02 dez. 2014.

DANIEL JUNIOR, Ron; LAGOZE, Carl. Ampliando o quadro Warwick: De Metadados Containers para Objetos Ativos Digitais. **D-lib Revista**, Eua, p.1-10, nov. 1997. Disponível em: <<http://www.dlib.org/dlib/november97/daniel/11daniel.html>>. Acesso em: 24 abr. 2014.

DIAS, Tatiane Domingos; SANTOS, Neide. *Web Semântica: conceitos básicos e tecnologias associadas*. **Cadernos do IME – Série Informática**, v. 14, jun de 2003. Disponível em: <<http://www.e-publicacoes.uerj.br/index.php/cadinf/article/view/6619/4734>>. Acesso em 05 maio 2014.

DMOZ. 2014. Disponível em: <<http://www.dmoz.org/>>. Acesso em: 13 nov. 2014.

FEITOSA, Ailton. **Organização da Informação na Web: das tags à Web Semântica**. Brasília, DF: Thesaurus, 2006. 131p.

FREIRE, Raquel. **Baidu ou Google? Veja qual buscador é melhor e mais fácil de usar**. 2014. Disponível em: <<http://www.techtudo.com.br/dicas-e-tutoriais/noticia/2014/08/baidu-ou-google-veja-qual-buscador-e-melhor-e-mais-facil-de-usar.html>>. Acesso em: 31 ago. 2014.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 175 p. Disponível em: <<http://pt.notices-pdf.com/como-elaborar-projetos-de-pesquisa-antonio-carlos-gil-pdf.html>>. Acesso em: 21 maio 2014.

GOOGLE. 2014. Disponível em: <<http://www.google.com>>. Acesso em: 24 ago. 2014.

GOOGLE. **Mapa do Conhecimento**. 2014. Disponível em: <<http://www.google.com/insidesearch/features/search/knowledge.html>>. Acesso em: 25 maio 2014.

GOOGLE. **Por dentro da pesquisa**. 2014. Disponível em: <<http://www.google.com.br/insidesearch/>>. Acesso em: 22 ago. 2014.

GRUBER, Tom. **What is na ontology? 1993**. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>> Acesso em: 18 abr. 2014.

HILLMANN, Diane et al. RDA Vocabulários: Processos, Resultados, Uso. **D-lib Revista**, Eua, v. 16, n. 1, p.1-20, fev. 2010. Disponível em: <<http://dlib.org/dlib/january10/hillmann/01hillmann.html>>. Acesso em: 08 maio 2014.

IFLA Study Group on the Functional Requirements for Bibliographic Records. **Functional requirements for bibliographic records**: final report. München : K. G. Saur, 1998.

JONH, Tony. **O que é a pesquisa Semântica e como ela funciona com a pesquisa do Google**. 2012. Disponível em: <<http://www.techulator.com/resources/5933-What-Semantic-Search.aspx>>. Acesso em: 17 maio 2014.

LIGHTBASE. **Web Semântica**: Dados Lincados. 2011. Disponível em: <<http://rdf.lightbase.cc/semantica#topo>>. Acesso em: 28 abr. 2014.

LOURENÇO, Cíntia de Azevedo. **Modelagem de dados como ferramenta de análise de padrões de metadados em bibliotecas digitais**: o padrão de metadados brasileiro para teses e dissertações segundo o modelo entidade-relacionamento. 2005. 161 f. Tese (Doutorado em Ciência da Informação) – Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/EARM-6ZGNZC/doutorado___c_ntia_de_azevedo_louren_o.pdf?sequence=1>. Acesso em 20 abr. 2014.

MARCONDES, Carlos Henrique; CAMPOS, Maria Luiza de Almeida. Ontologia e Web Semântica: o espaço da pesquisa em Ciência da Informação. **PontodeAcesso**, v. 2, n. 1, 2008. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/2669>>. Acesso em: 12 abr. 2014.

METABUSCADORES combinam forças de vários sites de pesquisa. **Folha de S. Paulo** - Informática, 24 mar. 2010. Disponível em: <<http://www1.folha.uol.com.br/tec/2010/03/712777-metabuscadores-combinam-forcas-de-varios-sites-de-pesquisa.shtml>>. Acesso em: 11 out. 2014.

MILLER, Eric. Uma Introdução ao Resource Description Framework. **D-lib Revista**, Eua, p.1-10, maio 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 07 jun. 2014.

MONTEIRO, Silvana Drumond. As Múltiplas Sintaxes dos Mecanismos De Busca No Ciberespaço. **Informação & Informação**, Londrina, v. 14, n. especial, p.68-102, 2009. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/2027>>. Acesso em: 27 ago. 2014.

MORAIS, Edison Andrade Martins; AMBRÓSIO, Ana Paula L. Ontologias: conceitos, usos, tipos, metodologias, ferramentas e linguagens. **Instituto de Informática/UFG**, dez. 2007. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-07.pdf>. Acesso em: 12 Abr. 2014.

MORENO, Fernanda P. **Requisitos Funcionais para Registros Bibliográficos – FRBR**: um estudo no catálogo da Rede Bibliodata. Brasília, 2006. Dissertação (mestrado em Ciência da Informação) - Universidade de Brasília. Disponível em: <<http://repositorio.bce.unb.br/handle/10482/2565>>. Acesso em: 21 out. 2011.

MORENO, Fernanda P.; ARELLANO, Miguel A. M. Requisitos Funcionais para Registros Bibliográficos – FRBR: uma apresentação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 3, n. 1, p. 20-38, jul./dez. 2005. Disponível em: <<http://eprints.rclis.org/handle/10760/10241>>. Acesso em: 29 set. 2011.

OLIVEIRA, Lais Pereira de; SILVEIRA, Carlos Eduardo da. Desafios do bibliotecário frente às novas tecnologias da informação e comunicação. In: ENCONTRO REGIONAL DE ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO, GESTÃO E CIÊNCIA DA INFORMAÇÃO, 12., 2010, Porto Alegre. **Anais...** . Porto Alegre: UFRGS, 2010. p. 1 - 15. Disponível em: <<https://pt.scribd.com/doc/48656094/Desafios-do-bibliotecario-frente-as-novas-tecnologias-da-informacao-e-comunicacao>>. Acesso em: 07 nov. 2014.

PEREIRA, Danielle. **Dados Abertos e Dados Linkados**. 2012. Open Data for Development- OD4D. Disponível em: <<http://platform.od4d.org/article?locale=pt&uri=http%3A%2F%2Fplatform.od4d.org%2Fposts%2F97>>. Acesso em: 28 abr. 2014.

PICKLER, Maria Elisa Valentim. Web Semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**. Belo Horizonte, v. 12, n. 1, abr. 2007 . Disponível em: <<http://www.scielo.br/pdf/pci/v12n1/05.pdf> >. Acesso em: 27 mar. 2014.

POLLOCK, Jeffrey T.. **Web Semântica para Leigos**. Rio de Janeiro: Alta Books, 2010. 401 p.

SANTOS NETO, Antonio Laurindo dos et al. Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. **Transinformação**, Campinas, v. 25, n. 1, p.81-87, abr. 2013. Disponível em: <<http://www.scielo.br/pdf/tinf/v25n1/a08v25n1>>. Acesso em: 24 maio 2014.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; ALVES, Rachel Cristina Vesú. Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0. **DataGramaZero**, Rio de Janeiro, v. 10, n. 6, p.1-10, dez. 2009. Disponível em: <http://www.dgz.org.br/dez09/F_I_art.htm>. Acesso em: 07 jun. 2014.

SCHISSL, Marcelo; BRÄSCHER, Marisa. Ontologia: ambiguidade e precisão. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. esp.1, p. 125-141, 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp1p125/22729>> . Acesso em: 27 abr. 2014.

SEMANTIC WEB. 2014. Disponível em: <http://semanticweb.org/wiki/main_Page>. Acesso em: 01 abr. 2014.

SENSEBOT. **Busca semântica**. 2014. Disponível em: <<http://www.sensebot.net/content3front.aspx>>. Acesso em: 12 nov. 2014.

SIQUEIRA, Ivan Cláudio Pereira. Mecanismos de busca na *Web*: presente, passado e futuro. **Ponto de Acesso**, Salvador, v. 7, n. 2, p.47-67, ago. 2013. Disponível em: <www.portalseer.ufba.br/index.php/revistaici/article/download/.../6136>. Acesso em: 24 maio 2014.

SILVA, Edna Lucia da; MENEZES, Estera Muszcat. **Metodologia da pesquisa e elaboração da dissertação**. 4. ed. Florianópolis: UFSC, 2005. 138 p. Disponível em: <https://projetos.inf.ufsc.br/arquivos/Metodologia_de_pesquisa_e_elaboracao_de_teses_e_dissertacoes_4ed.pdf>. Acesso em: 25 out. 2014.

SILVA, Edna Lúcia da e CUNHA, Miriam Vieira da. A Formação Profissional no Século XXI : Desafios e dilemas . *Ciência da Informação*. 2002, vol.31, n.3, pp. 77-82. Disponível em: <<http://www.scielo.br/pdf/ci/v31n3/a08v31n3.pdf>>. Acesso em: 02 dez. 2014.

SILVA, Fabiano Couto Corrêa da. *Bibliotecários especialistas: guia de especialidades e recursos informacionais*. Brasília: Thesaurus, 2005.

SOUZA, Renato Rocha. Sistemas de Recuperação de Informações e Mecanismos de Busca na Web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, p.161-173, 10 jul. 2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n2/v11n2a02.pdf>>. Acesso em: 23 maio 2014.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e SUAS contribuições para a Ciência da Informação. **Ciência da Informação**, Brasília, v 33, n. 1, abril 2004. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652004000100016&lng=en&nrm=iso>. Acesso em: 27 mar. 2014.

TEIXEIRA, Paulo Rodrigo. *Marketing de Busca. O que são diretórios Web?*. 2007. Disponível em: <<http://www.marketingdebusca.com.br/o-que-sao-diretorios-web/>>. Acesso em: 13 nov. 2014.

TEOTÔNIO, Mara Karoline Lins. Bibliotecário 2.0: novos desafios na era da sociedade em rede. *Ciência da Informação*, Brasília, v. 1, n. 4, p.34-39, jul. 2011. Disponível em: <<http://periodicos.unb.br/index.php/RICI/article/view/6160/5430>>. Acesso em: 05 dez. 2014.

TEXT ENCODING INITIATIVE - TEI. [ca 2012]. Disponível em: <<http://www.tei-c.org/index.xml>>. Acesso em: 21 abr. 2014.

VITAL, Luciane Paula; CAFÉ, Ligia Maria Arruda. Ontologias e Taxonomias: diferenças. **Perspectivas em Ciência da Informação**, v. 16, n. 2, 2011. Disponível

em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/200>>. Acesso em: 12 abr. 2014.

WIKIMEDIA COMMONS. **File:Meta-search-es.png**. 2011. Disponível em: <<http://commons.wikimedia.org/wiki/File:Meta-search-es.png?uselang=pt-br>>. Acesso em: 13 set. 2014.

WIKIPEDIA. **Semântica**. 2014. Disponível em: <<http://pt.wikipedia.org/wiki/Sem%C3%A2ntica>>. Acesso em: 01 abr. 2014.

WIKIPEDIA. **OWL**. Disponível em: <<http://pt.wikipedia.org/wiki/owl>>. Acesso em: 07 jun. 2014.

W3C. **Semantic Web**. 2014. Disponível em: <<http://www.w3.org/standards/semanticweb/>>. Acesso em: 01 abr. 2014

W3C. **OWL 2 Web Ontology Language Document Overview** (segunda edição). 2012. Disponível em: <<http://www.w3.org/TR/owl2-overview/>>. Acesso em: 07 jun. 2014.

W3C. Semantic Web. **Query**. 2014. Disponível em: <<http://www.w3.org//standards/semanticweb/query>>. Acesso em: 25 out. 2014.

WHOS TALKIN. **Metabusador**. 2014. Disponível em: <<http://whostalkin.com/search?q=social+media+search&x=36&y=23>>. Acesso em: 13 set. 2014.

YAHOO!. 2014. Disponível em: < <https://br.yahoo.com>>. Acesso em: 28 ago. 2014.