

ROGÉRIO APARECIDO SÁ RAMALHO

Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação

**Marília
2006**

ROGÉRIO APARECIDO SÁ RAMALHO

Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, Campus de Marília, como requisito parcial para obtenção do título de Mestre em Ciência da Informação. Área de concentração: Informação, Tecnologia e Conhecimento, linhas de pesquisa: Informação e Tecnologia; Organização da Informação.

Orientação: Dr^a. Silvana Aparecida Borsetti Gregorio Vidotti

Co-orientação: Prof^a Dr^a. Mariângela Spotti Lopes Fujita

**Marília
2006**

Ramalho, Rogério Aparecido Sá
R165w Web Semântica : aspectos interdisciplinares da gestão
de recursos informacionais no âmbito da Ciência da
Informação / Rogério Aparecido Sá Ramalho. – Marília,
2006.
120 f. ; 30 cm.

Dissertação (Mestrado em Ciência da informação) –
Faculdade de Filosofia e Ciências, Universidade Estadual
Paulista, 2006.

Bibliografia: f. 111-120.

Orientadora: Prof^ª Dr^ª Silvana Aparecida Borsetti
Gregorio Vidotti

Co-Orientadora: Prof^ª Dr^ª Mariângela Spotti Lopes
Fujita

1. Web Semântica. 2. Recuperação da Informação.
2. Ontologia. 3. Gestão de recursos informacionais.
4. Internet. Autor. II. Título.

CDD: 004.67

ROGÉRIO APARECIDO SÁ RAMALHO

Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação

Banca Examinadora:

Dr^a. Silvana Aparecida Borsetti Gregorio Vidotti
Programa de Pós-Graduação em Ciência da Informação
Universidade Estadual Paulista – UNESP – Campus de Marília

Dr. Marcos Luiz Mucheroni
Programa de Pós-Graduação em Ciência da Computação
Centro Universitário Eurípides de Marília - UNIVEM

Dr^a. Plácida Leopoldina Ventura Amorim da Costa Santos
Programa de Pós-Graduação em Ciência da Informação
Universidade Estadual Paulista – UNESP – Campus de Marília

Marília, 22 de Fevereiro de 2006.

Dedico este trabalho aos meus Pais Maria e Raul Ramalho, que sempre me apoiaram incondicionalmente em todos os aspectos de minha Vida, ensinando-me a fazer de cada limitação um incentivo e de cada fraqueza uma força, por meio de seus próprios exemplos de vida.

À minha querida Celina, companheira de todos os momentos, mesmo quando não estava fisicamente ao meu lado, que foi fundamental para a realização deste trabalho, transmitindo-me sempre Apoio, Amor e Confiança.

“Ando devagar porque já tive pressa e levo este sorriso, porque já chorei demais. Hoje me sinto mais forte, mais feliz quem sabe, só levo a certeza de que muito pouco eu sei, eu nada sei ...”

”Cada um de nós compõe a sua história, cada ser em si, carrega o dom de ser capaz... De ser Feliz!”

(Almir Sater e Renato Teixeira)

AGRADECIMENTOS

A Deus, por TUDO que tem me proporcionado, dando-me a cada dia novas lições de humildade e concedendo-me coragem suficiente para aprender com a Vida.

Às professoras Silvana Aparecida Borsetti Gregorio Vidotti e Mariângela Spotti Lopes Fujita, pelos exemplos de profissionalismo, por todos os ensinamentos transmitidos, pelos incentivos constantes e pela confiança depositada ao acreditarem no projeto de um jovem sonhador, acolhendo-me na área de Ciência da Informação.

Aos professores Marcos Luiz Mucheroni e Plácida Leopoldina Ventura Amorim da Costa Santos, pelos apontamentos e contribuições dadas no decorrer desta pesquisa e na qualificação, que foram de grande valia para conclusão deste trabalho.

Aos Professores do Departamento de Ciência da Informação da UNESP, Campus de Marília, por todas as lições recebidas, não limitadas a conhecimentos acadêmicos, e por persistirem no desenvolvimento de um ensino de qualidade.

Aos meus colegas de Mestrado, os quais participaram ativamente da produção coletiva do conhecimento que culminou neste trabalho, pelo companheirismo, amizade e os ótimos momentos que compartilhamos, em especial aos amigos: Jane Monção, Maria de Lourdes, Vera Boccato e Willy Dantas.

Ao meu querido irmão Raul Augusto pelo apoio e incentivo constante e por tudo que ensinaste a este teu irmão caçula.

Ao meu grande Amigo William Yonenaga, pessoa fantástica, com a qual tive a grata satisfação de conviver desde a época da graduação, o meu reconhecimento pelo companheirismo e incentivo constante.

Ao meu amigo Airton Althman, à Luana Sales da Universidade Federal Fluminense e a todos os “colegas virtuais” que tive a oportunidade de manter contato ao longo desta pesquisa, pelas instigantes discussões que tanto contribuíram para a realização deste projeto.

Aos meus amigos e colegas da Fundação Dracenense de Educação e Cultura, pelo apoio para a realização deste trabalho. À professora Zenaide Branco pela revisão do texto.

Aos profissionais da informação atuantes na biblioteca da UNESP, Campus de Marília, pela excelência no atendimento, em especial à minha Amiga Maria Luzinete Euclides, pela amizade e carinho.

A todos os funcionários da UNESP, Campus de Marília, pela atenção e simpatia que sempre demonstraram no exercício de suas funções, em especial aos atuantes no Setor de Pós-graduação da UNESP: Edna, Yara, Márcia, Aline e Andréia, por terem sempre me auxiliado de forma amigável e cortês.

Enfim, a todos que direta ou indiretamente contribuíram para a elaboração deste trabalho o meu muito obrigado!

*“É difícil dizer o que é impossível,
pois a fantasia de ontem
é a esperança de hoje
e a realidade de amanhã”.*

(Robert H. Goddard, 1882 - 1945)

RAMALHO, Rogério Aparecido Sá. *Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação*. 2006. 120 f. Dissertação (Mestrado em Ciência da Informação). Faculdade de Filosofia e Ciências – Universidade Estadual Paulista, Marília, 2006.

RESUMO

No âmbito da gestão de recursos informacionais os modelos e métodos de organização e recuperação de informações sempre estiveram condicionados às tecnologias utilizadas, de modo que com desenvolvimento e intensificação da utilização das tecnologias digitais uma nova gama de possibilidades vem sendo incorporada aos processos de produção, armazenamento, representação e recuperação de informações, atingindo um estágio em que os modelos clássicos de organização e recuperação de informações precisam ser (re)pensados sob diferentes perspectivas, pois os mesmos não parecem ser capazes de solucionar os problemas identificados no ambiente Web, evidenciando a necessidade de desenvolvimento de novas tecnologias que permitam otimizar a recuperação de informações em ambientes digitais. Nesse sentido, os estudos relacionados ao projeto Web Semântica vêm destacando-se como uma nova perspectiva no desenvolvimento de tecnologias que possibilitem um aumento na qualidade e relevância das informações recuperadas, a partir do desenvolvimento de instrumentos que permitam descrever formalmente, em um formato que possa ser processado por máquinas, os aspectos semânticos inerentes aos recursos informacionais, contribuindo para a identificação e contextualização das informações disponíveis no ambiente Web. Deste modo, a proposição deste trabalho é a realização de um estudo teórico e metodológico de caráter interdisciplinar acerca do projeto Web Semântica, buscando favorecer a “desmistificação” dos conceitos e tecnologias subjacentes e avaliar em que medida a área de Ciência da Informação pode contribuir para sua concretização, ressaltando os possíveis reflexos destas novas abordagens tecnológicas em seu *corpus* teórico. Assim, apresenta-se um levantamento bibliográfico acerca do desenvolvimento da Internet e os principais conceitos e tecnologias relacionados ao projeto Web Semântica, contextualizando-os a partir dos aportes teóricos da área de Ciência da Informação. Verificou-se que os conceitos e tecnologias subjacentes ao projeto Web Semântica podem ser considerados como uma renovação ou desdobramento dos métodos convencionais de representação, organização e recuperação de informações, apontando a possibilidade de contribuições da área de Ciência de Informação para o desenvolvimento do projeto Web Semântica, devido ao seu embasamento teórico referente a formas de representação e as práticas profissionais identificadas em seu campo de atuação, sendo evidente o caráter interdisciplinar que delinea o *corpus* teórico dos estudos relacionados ao projeto Web Semântica, apresentando-se como um campo fértil para pesquisas e indicando a possibilidade de desenvolvimento de novos métodos de organização e recuperação de informações, construídos por meio de esforços interdisciplinares que favoreçam a junção da fundamentação teórica inerente à área de Ciência da Informação com as novas tecnologias emergentes da área de Ciência da Computação.

Palavras-chave: Web Semântica; Recuperação de Informação; Ontologia; Gestão de Recursos Informacionais; Internet.

RAMALHO, Rogério Aparecido Sá. *Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação*. 2006. 120 f. Dissertation (M.Sc. Information Science). Faculdade de Filosofia e Ciências – Universidade Estadual Paulista, Marília, 2006.

ABSTRACT

In the scope of the information resource management, the models and methods of organization and retrieval of information were always conditioned to the used technologies, so that with the development and intensification of digital technology uses, a new scale of possibilities has been incorporated to the production process, storage, representation and retrieval of information, reaching a stage where the classic models of organization and retrieval of information need to be (re)thought under different perspectives, because they don't seem to be able to solve the problems identified in the Web environment, becoming evident the need of development of new technologies that allow to optimize the retrieval of information in digital environment. In this way, the studies related to the Semantic Web project have been detaching as a new perspective in the development of technologies that enable an increase in the quality and relevance of the recovered information through the development of instruments that allow describing them formally, in a format that can be processed by machines. The semantic aspects that are inherent to the information resources contribute to the identification and contextualization of the available information in the Web environment. In this way, the proposition of this research is the accomplishment of a theoretical and methodological study of interdisciplinary characteristic about the Semantic Web project, aiming to identify its theoretical basis, favoring the "demystification" of the concepts and subjacent technologies, and evaluating in what stage the Information Science area can contribute to its concretization, becoming evident the possible reflexes of these new technological approaches in its theoretical corpus. So a bibliographic review about the development of the Internet and the main concepts and technologies inherent to the Web Semantic project is presented, contextualizing them from the theoretical basis of the Information Science area. It was verified that the concepts and technologies subjacent to the Web Semantic project can be considered as a renovation or an unfolding of the conventional methods of representation, organization and retrieval of information, pointing the possibility of contribution in the Information Science area to the development of the Web Semantic project, because of its theoretical basis referent to the representation forms and the professional practices identified in its action field, becoming evident the interdisciplinary feature that delineate the theoretical corpus of the studies related to the Web Semantic project, presenting it as a fertile field to researches and pointing the possibility of development of new methods of organization and retrieval information, built through interdisciplinary efforts that favor the joint of the theoretical basis inherent to the Information Science area with the new emergent technologies of the Computer Science area.

Keywords: Semantic Web; Information Retrieval; Ontology; Information Resources Management; Internet.

LISTA DE FIGURAS

FIGURA 1 – Capa da proposta original da World Wide Web	22
FIGURA 2 – “A Roda de Leitura”	26
FIGURA 3 – Partes que constituem um URL	33
FIGURA 4 – Exemplo de uma comunicação na Web	33
FIGURA 5 – Estruturação dos recursos e links na Web atual e na Web Semântica	40
FIGURA 6 – Estrutura de recuperação de informações na Web Semântica	41
FIGURA 7 – Estrutura para o desenvolvimento de novas aplicações no âmbito da Web Semântica	42
FIGURA 8 – A utilização de Agentes no âmbito da Web Semântica	43
FIGURA 9 – Arquitetura da Web Semântica proposta em 2000	44
FIGURA 10 – Arquitetura da Web Semântica proposta em 2002	45
FIGURA 11 – Arquitetura da Web Semântica proposta em 2005	47
FIGURA 12 – Sugestão de alteração na Arquitetura proposta para Web Semântica ...	51
FIGURA 13 – “Espectro Funcional” da Web Semântica	52
FIGURA 14 – Diagrama de uma ontologia de um domínio acadêmico	58
FIGURA 15 – Processo de Análise de um Documento XML	63
FIGURA 16 – Estrutura de um “elemento simples” em XML	64
FIGURA 17 – Codificação de um documento em XML	65
FIGURA 18 – Codificação XSD do documento apresentado na figura 17	66
FIGURA 19 – Diagrama do Esquema apresentado na figura 17	67
FIGURA 20 – Trecho da Codificação de um documento em XML	69
FIGURA 21 – Grafo de uma declaração RDF	71
FIGURA 22 – Grafo de uma declaração RDF Composta	72
FIGURA 23 – Representação de uma declaração RDF utilizando a Sintaxe XML	73
FIGURA 24 – Trecho de código baseado em RDFS	75
FIGURA 25 – Trecho de código OWL indicando os Namespaces utilizados	77
FIGURA 26 – Trecho de código OWL referente ao cabeçalho de uma ontologia	78
FIGURA 27 – Declaração de Classes e Subclasses em OWL	80
FIGURA 28 – Declaração de uma propriedade do tipo Objeto	81

FIGURA 29 – Declaração de uma propriedade do tipo Datatype	82
FIGURA 30 – Declaração de uma propriedade do tipo Datatype representando a instância de um objeto	82
FIGURA 31 – Níveis de complexidade estrutural dos Vocabulários Controlados	90
FIGURA 32 – An Ontology Spectrum	93
FIGURA 33 – The ontology spectrum: Weak to strong semantics	95
FIGURA 34 – An extract from the UKAT	99
FIGURA 35 – Relacionamentos da figura 34 apresentados a partir do SKOS	100
FIGURA 36 – Espectro Funcional das camadas da Web Semântica	101

LISTA DE TABELAS

TABELA 1 – Relação dos 20 países com maior número de usuários na Internet	30
TABELA 2 – Datatypes definidos pelo W3C para utilização em OWL	81

SUMÁRIO

1	INTRODUÇÃO	15
2	A INTERNET	20
2.1	Evolução histórica da Internet	21
2.2	A evolução da Internet no Brasil	27
2.3	Estrutura e componentes da Web	31
3	WEB SEMÂNTICA	37
3.1	Definições	37
3.2	Arquitetura da Web Semântica	44
3.3	Ontologias na Web Semântica	54
3.4	Linguagens de Representação de Recursos Informativos	60
3.4.1	XML	60
3.4.2	RDF	70
3.4.3	OWL	75
4	CONTEXTUALIZANDO A WEB SEMÂNTICA NO ÂMBITO DA CIÊNCIA DA INFORMAÇÃO	84
4.1	Ontologias como instrumento de representação	87
4.2	A Web Semântica sob o prisma da Ciência da Informação	98
5	CONSIDERAÇÕES FINAIS	106
5.1	Pesquisas Futuras	110
	REFERÊNCIAS	111

1 INTRODUÇÃO

É inegável o fato de que as técnicas e tecnologias sempre assumiram um importante papel no processo civilizatório, de modo que é cada vez mais evidente a influência que, as mesmas, exercem na sociedade contemporânea, a qual vem sendo denominada por inúmeros autores como “Sociedade da Informação”¹.

Segundo Burnham (2004), as transformações por que passa o mundo globalizado são influenciadas, direta ou indiretamente, pelo avanço das tecnologias da informação, de modo que a velocidade com que trafegam os dados nas “infovias” e a extensão do alcance das informações disponibilizadas contribuem para uma mudança na compreensão da relação tempo-espço.

Nesse contexto, a Internet, uma infra-estrutura de redes, servidores e canais de comunicação que possibilita a troca de informações em nível global, popularizada principalmente a partir do desenvolvimento da *World Wide Web*, ou simplesmente Web, tem favorecido um aumento exponencial na quantidade de informações disponíveis, permitindo que qualquer pessoa possa desempenhar o papel de produtor ou consumidor de informações, independentemente de fronteiras geográficas.

Conforme afirma Vidotti (2001, p.44):

Podemos pensar na Internet como uma grande biblioteca, ou como um ambiente hipermídia coletivo, no qual os usuários são agentes ativos do processo de armazenamento, indexação, recuperação e disseminação de documentos eletrônicos hipertextuais, um ambiente auto-organizado em permanente mutação.

No âmbito da gestão de recursos informacionais os modelos e métodos de organização e recuperação de informações sempre utilizaram as tecnologias vigentes. Desse modo, a intensificação da utilização das tecnologias digitais vem favorecendo uma nova gama de possibilidades de desenvolvimento de instrumentos que possibilitem otimizar os processos de produção, armazenamento, representação e recuperação de informações.

¹ A discussão acerca desta denominação, seu caráter ideológico ou não, pertinência e outros elementos de sua discussão fogem aos objetivos desta pesquisa.

Assim, considerando a representação da informação como elemento fundamental para a garantia de qualidade na recuperação, apresenta-se como desafio a necessidade de singularização contextual na reconstrução do conhecimento, com a determinação de requisitos de qualidade e relevância das informações, a partir da utilização de categorias que permitam organizar, de maneira eficiente, o “oceano” de dados disponíveis, possibilitando a identificação da informação que realmente interessa ao usuário em um contexto adequado, pois devido ao aumento na quantidade de recursos informacionais disponíveis e a maneira como tais recursos são representados e estruturados no ambiente digital, torna-se uma tarefa cada vez mais complexa a recuperação de informações, possibilitando, inclusive, que informações relevantes, em um contexto específico, deixem de ser utilizadas devido às dificuldades encontradas no processo de recuperação.

Deste modo, verifica-se que os modelos clássicos de organização e recuperação de informações precisam ser (re)pensados sob diferentes perspectivas, pois os mesmos não parecem ser capazes de solucionar os problemas identificados no ambiente Web, evidenciando a necessidade de desenvolvimento de novas tecnologias que possibilitem otimizar a recuperação de informações em ambientes digitais.

Tradicionalmente os recursos informacionais são disponibilizados no ambiente Web a partir de formatos de apresentação que têm como objetivo principal possibilitar uma interface adequada para os usuários, limitando-se a tarefa de exibir informações, para que estas possam ser visualizadas e interpretadas por usuários humanos. Quanto ao projeto Web Semântica, a idéia é pensar nas máquinas para que estas possam servir aos humanos de maneira mais eficiente, apresentando-se como uma nova perspectiva no desenvolvimento de tecnologias de representação e recuperação de informações, focando-se no desenvolvimento de instrumentos que permitam descrever formalmente os aspectos semânticos inerentes aos recursos informacionais, com o intuito de possibilitar um aumento na qualidade e na relevância das informações recuperadas, a partir da identificação e contextualização de maneira automatizada das informações disponíveis no ambiente Web.

Sendo assim, apresenta-se como proposição deste trabalho um estudo teórico e metodológico de caráter interdisciplinar acerca do projeto Web Semântica, buscando identificar seu embasamento teórico e avaliar em que medida a área de Ciência da

Informação pode contribuir para sua concretização, ressaltando os possíveis reflexos destas novas abordagens tecnológicas em seu *corpus* teórico.

Nesta perspectiva, destaca-se a necessidade de estudos que favoreçam a elaboração de novas tecnologias e métodos mais adequados para o tratamento e organização de informações no meio digital, assim como o fornecimento de subsídios teóricos para que tais tecnologias não sejam desenvolvidas a partir de um “vazio conceitual”, com o único objetivo de atender demandas de mercado, possibilitando que os avanços tecnológicos possam ser devidamente sedimentados no campo teórico.

Assim, a partir da interação das linhas de pesquisas: Informação e Tecnologia e Organização da Informação, do Programa de Pós-Graduação em Ciência da Informação da Unesp de Marília, o objetivo desta pesquisa é contribuir para uma “desmistificação” dos conceitos e tecnologias subjacentes ao projeto Web Semântica, com o intuito de contribuir para um melhor entendimento das tecnologias relacionados, identificando pontos de convergência entre as novas abordagens computacionais e os métodos de análise, síntese, organização e representação de recursos informacionais utilizados tradicionalmente no âmbito da área de Ciência da Informação e indicando possibilidades de desenvolvimento de novos métodos de organização e recuperação de informações, que favoreçam a junção da fundamentação teórica inerente à área de Ciência da Informação com as novas tecnologias emergentes da área de Ciência da Computação.

Nesse contexto, utilizou-se como metodologia para realização desta pesquisa uma análise exploratória e descritiva da literatura relacionada ao projeto Web Semântica, identificada principalmente em estudos da área de Ciência da Computação, porém não se limitando a informações provenientes desta área do conhecimento.

Deste modo, adotou-se como abordagem inicial a seleção de documentos que permitissem um melhor entendimento dos conceitos inerentes ao projeto Web Semântica, levando-se em consideração critérios de pertinência e atualidade de seus respectivos conteúdos, devido principalmente à abrangência das tecnologias relacionadas.

Neste sentido, por ser tratar de um tema em evolução, optou-se pelo não aprofundamento em detalhes tecnológicos, apresentando considerações e exemplos das principais tecnologias utilizadas com o intuito de facilitar uma melhor compreensão do projeto Web Semântica de uma forma integral.

Posteriormente, procurou-se estabelecer relacionamentos que permitissem uma maior sustentação teórica aos conceitos identificados, contextualizando-os a partir dos aportes teóricos da área de Ciência da Informação.

Sendo assim, a realização desta pesquisa justifica-se devido à crescente demanda de estudos interdisciplinares, que respeitando as especificidades de cada área do conhecimento, possam auxiliar na evolução do conhecimento humano de uma forma integral, baseados na crença da unidade do conhecimento por sobre as fronteiras das ciências, verificando-se a possibilidade de maiores contribuições da área de Ciência da Informação no desenvolvimento de tecnologias de organização e representação de informações no meio digital.

Desta maneira, de acordo com o objetivo proposto, a presente pesquisa foi estruturada a partir de cinco capítulos mutuamente complementares, de modo que além do presente capítulo introdutório, os capítulos seguintes podem ser descritos sucintamente da seguinte forma:

No próximo capítulo apresenta-se um breve levantamento bibliográfico acerca dos principais fatos e personalidades que marcaram a evolução da Internet no âmbito global e nacional, descrevendo de modo sucinto os principais componentes de sua estrutura.

No capítulo 3, intitulado sugestivamente como Web Semântica, busca-se identificar a fundamentação teórica inerente ao projeto Web Semântica, a partir dos recentes estudos apresentados principalmente na área de Ciência da Computação, descrevendo quais as principais tecnologias necessárias para sua concretização e como estas se relacionam na estrutura de camadas proposta pelo W3C, contextualizando as ontologias como instrumentos de modelagem semântica de objetos e conceitos no âmbito da representação de recursos informacionais e descrevendo as principais linguagens e formatos computacionais padronizados pelo W3C para o desenvolvimento da Web Semântica: XML, RDF e OWL.

No capítulo 4, busca-se contextualizar o projeto Web Semântica no âmbito da área de Ciência da Informação, verificando quais as possíveis contribuições desta área do conhecimento e indicando quais os desafios e reflexos destas novas abordagens no fazer profissional da área de Ciência da Informação.

Finalmente no capítulo 5 são apresentadas as considerações finais desta pesquisa, com o intuito de contribuir para uma maior discussão a respeito do papel incorporado pelas tecnologias de informação na sociedade contemporânea, contextualizando o caráter interdisciplinar que delinea o projeto Web Semântica e apontando possíveis contribuições identificadas a partir do referencial teórico da área de Ciência da Informação, evidenciando assim como um campo fértil para pesquisas, indicando inclusive pesquisas futuras, pois conforme afirma Wiener (1970, p.27): “São estas regiões fronteiriças da ciência que oferecem as mais ricas oportunidades ao investigador qualificado”.

2 A INTERNET

O homem sempre aspirou desenvolver meios que permitissem reunir e disponibilizar de maneira integral todo o conhecimento da humanidade, tal intenção pode ser verificada em fatos históricos como a construção da Biblioteca de Alexandria², criada por Ptolomeu I, o Mundaneum³, idealizado por Paul Otlet e Henri de La Fontaine, pelo Memex⁴, de Vannevar Bush, pelo projeto Xanadu³, de Theodore Nelson, e, de certo modo, mais recentemente pelo desenvolvimento da Internet, a qual diferentemente dos anteriores, constituiu-se a partir de um conjunto de tecnologias originárias da combinação de estudos independentes, de modo que não existe um consenso a respeito de um nome que possa representar a sua idealização ou desenvolvimento de forma integral.

Nessa perspectiva o presente capítulo apresenta um breve levantamento bibliográfico acerca dos principais fatos históricos que marcaram a evolução da Internet e apresenta algumas considerações a respeito das principais tecnologias e conceitos relacionados, descrevendo de modo sucinto os principais componentes de sua estrutura.

² Segundo Mey (2004) a Biblioteca de Alexandria constituiu o “centro de excelência” cultural do mundo, durante seis séculos, reunindo estudiosos, sábios e artistas, não havendo dúvidas sobre sua importância na história da humanidade.

³ Criado em Bruxelas, em 1920, representando a materialização do ideal universalista de síntese do saber, a qual teria sido precedida por outras iniciativas dos mesmos idealizadores com propósitos semelhantes: o Instituto Internacional de Bibliografia, a “Classificação Decimal Universal”, o “Repertório Bibliográfico Universal” e o Repertório Iconográfico Universal. (RAYWARD, 1997)

⁴ O Memex foi concebido como uma máquina capaz de estocar grande quantidade de informações de forma fácil e que permitisse uma rápida recuperação; e o projeto Xanadu foi imaginado a partir da idéia de uma imensa rede de informações acessível em tempo real, contendo todo o saber literário e científico do mundo, de modo que, apesar de ambos os projetos não terem obtido o êxito esperado, significaram avanços importantes, conforme será apresentado ao longo deste capítulo.

2.1 Evolução histórica da Internet

Um dos primeiros fatos marcantes na breve história da Internet ocorreu em 1969, quando o pesquisador Leonard Kleinrock estabeleceu a primeira comunicação entre dois computadores conectados a então denominada *Advanced Research Project Agency Network*, (ARPANET), uma rede experimental financiada pela *Defende Advanced Research Projects Agency* dos EUA, (DARPA), enviando uma mensagem de um computador localizado na Universidade da Califórnia, em Los Angeles (UCLA), para outro localizado no Instituto de Pesquisa de Stanford (SRI). Desde então, Leonard Kleinrock é considerado como o inventor dos princípios básicos da comunicação por pacotes de dados, uma das bases para o funcionamento da Internet. (GILLIES e CAILLIAU, 2000; KLEINROCK, 1969)

Tendo surgido em plena guerra fria, o objetivo inicial da ARPANET era descentralizar as informações militares, pois como não possuía um centro definido, nem mesmo uma rota única para o envio de informações, os computadores conectados poderiam se comunicar mesmo em caso de um ataque inimigo que prejudicasse o restante da estrutura da rede. Foi então, a partir da década de 1970, que os estudos relacionados com a ARPANET possibilitaram o desenvolvimento de redes acadêmicas, tais como: HEPNet, Usenet, Eunet, Bitnet, entre outras. De modo que os princípios que regiam a ARPANET constituíam a base de todas as redes.

Nesse contexto, devido ao crescimento da quantidade de redes de computador, assim como o aumento da quantidade de computadores conectados, o protocolo de comutação de pacotes desenvolvido para a ARPANET o *Network Control Protocol* (NCP), utilizado para a troca de dados na rede, apresentava uma série de limitações tornando-se então inadequado. Assim, em 1980 Vinton Cerf em parceria com Robert Kahn publicaram a especificação para os protocolos: *Transmission Control Protocol* (TCP) e *Internet Protocol* (IP), dando origem ao conjunto de protocolos TCP/IP, o qual ainda é utilizado na Internet.

Foi em 1985 que a *National Science Foundation* (NSF) estabeleceu a NSFNet, uma rede criada com o propósito de interligar várias universidades, tomando uma série de medidas que se apresentaram posteriormente fundamentais para o desenvolvimento da Internet, entre elas a obrigatoriedade da utilização do padrão TCP/IP, o que possibilitou a junção de várias redes. Pouco tempo depois a NSFNet iria constituir a “coluna vertebral” da própria Internet dos EUA.

No final da década de 1980, mais especificamente em 1989, Tim Berners-Lee, pesquisador do *European Organization for Nuclear Research Center* (CERN), apresentou uma proposta de elaboração de um ambiente gráfico para a utilização da Internet que possibilitasse a difusão de textos, imagens estáticas, áudios, vídeos e animações de forma integrada, tal ambiente foi denominado como *World Wide Web* (WWW), ou simplesmente Web. (BERNERS-LEE, 1989)

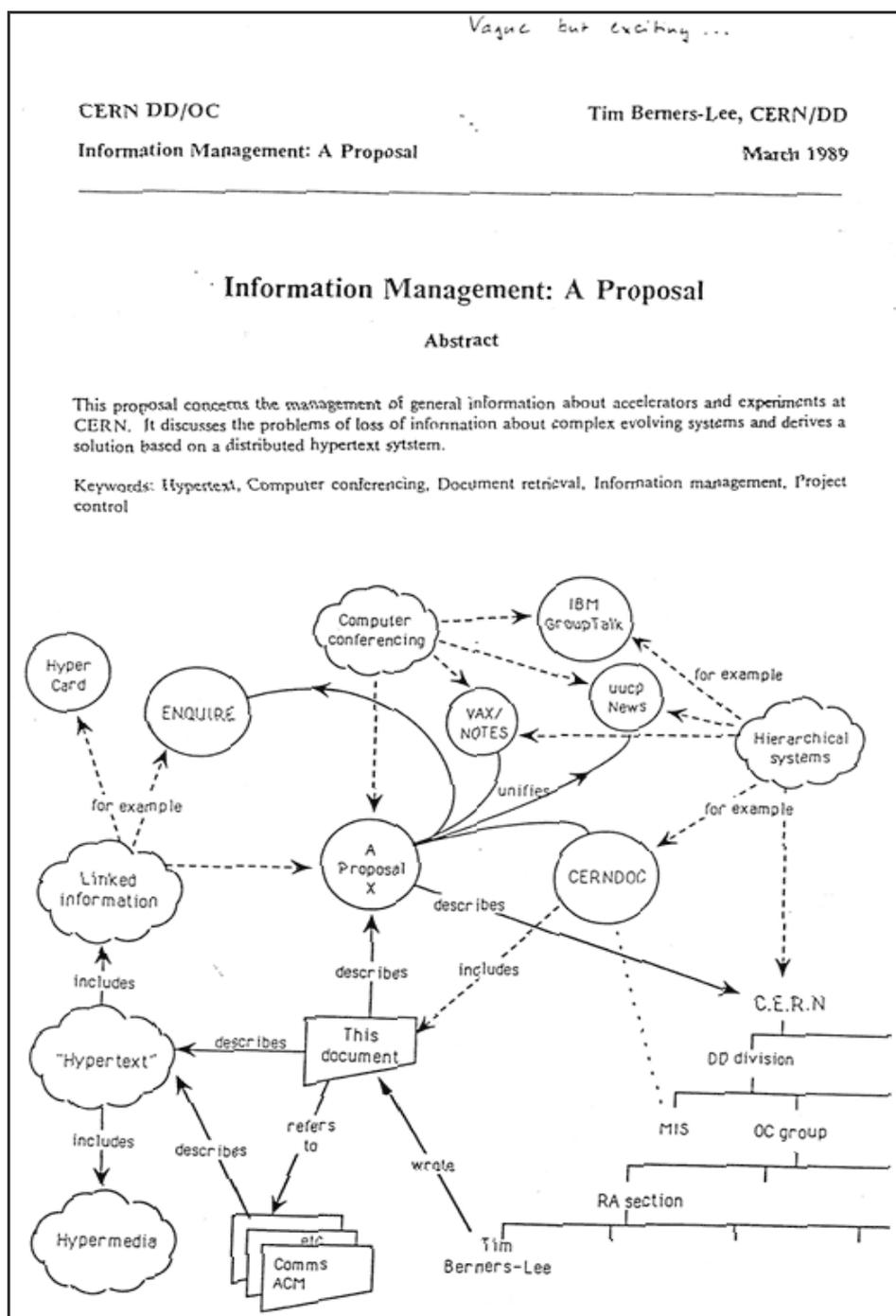


FIGURA 1 – Capa da proposta original da *World Wide Web*
 Fonte: Gillies e Cailliau, 2000, p.181.

Conforme afirma Engelschall (2001, p. 8, tradução nossa)

Depois da aprovação da idéia por Mike Sendall, chefe de Tim. Berners-Lee, iniciou-se o desenvolvimento de um browser e editor de hipertexto GUI⁵ usando a linguagem de desenvolvimento NeXTStep. Berners-Lee deu o nome de “WorldWideWeb” para este software e mais tarde rebatizou-o como “Nexus”, para evitar confusões entre tal software e o espaço abstrato de informação. Depois o projeto foi desenvolvido no CERN durante dois anos, a World Wide Web (WWW) rapidamente tornou-se o primeiro sistema de hipertexto global e a abreviação WWW entrou na consciência pública.

Deste modo, com o desenvolvimento de *browsers* capazes de acessar conteúdos disponibilizados em hipertexto, a utilização da Internet tornou-se uma tarefa mais agradável e intuitiva, o que viabilizou a sua utilização para fins comerciais. Assim, a Internet inicialmente circunscrita no mundo acadêmico e militar, transformou-se em uma promissora rede mundial de computadores interligados ao redor do mundo, a qual constitui atualmente uma das mais ricas fontes de conteúdos informacionais.

Segundo Krishnamurthy e Rexford (2001, p.4) o termo hipertexto foi cunhado por Theodore Nelson em 1965 para exprimir a idéia de escrita/leitura não linear em um sistema computacional. Foi nesta época, quando os primeiros sistemas de tele-informática militares acabavam de ser instalados e que os computadores possuíam poucas funcionalidades, não sendo capazes sequer de processar textos, que Nelson criou o projeto Xanadu, imaginando uma imensa rede de informações acessível em tempo real, contendo todo o saber literário e científico do mundo, a qual milhares de pessoas poderiam se conectar para ler, escrever, comentar, interagir e estudar, utilizando-se de todos os recursos nela disponíveis, compostos não somente de textos, como também de imagens e sons.

Segundo Beiguelman (2003), o próprio Nelson apresentou a idéia de hipertexto, em sua obra *Literary Machines*, como uma escrita não seqüencial, um texto com vários caminhos que permite que os leitores façam escolhas, e que é mais facilmente lido numa tela interativa. Deste modo, Nelson foi um dos pioneiros no desenvolvimento de “mundos virtuais” e desenho de hipertextos, podendo-se considerar que ele descreveu as janelas antes mesmo da casa existir.

⁵ A sigla GUI é a abreviação de *Graphical User Interface*, Interface Gráfica do Usuário, (que é a base da interface da maioria dos sistemas operacionais atuais).

Theodore Nelson foi um dos discípulos de Vannevar Bush, o qual também assumiu uma importância marcante na história do hipertexto, quando publicou em 1945 o clássico ensaio intitulado *As We May Think*, profetizando o advento de mecanismos que permitissem um melhor processamento, registro, transporte e distribuição de informações, mecanismos os quais podem ser comparados com os atuais sistemas de gestão de recursos informacionais. (BUSH, 1945)

Bush era matemático e responsável por uma agência de desenvolvimento e pesquisa científica do governo Norte-Americano. Coordenava o trabalho de mais de seis mil cientistas, assim, uma das questões enfrentadas por Bush era o volume crescente de dados que deviam ser armazenados e organizados de tal forma que permitisse a outros pesquisadores a utilização destas informações de maneira rápida e eficiente. Nesta perspectiva, Bush imaginou e descreveu, de maneira detalhada, uma máquina capaz de estocar grande quantidade de informações, de modo fácil e que permitisse uma rápida recuperação. Tal engenho batizado por Bush como Memex (*Memory Extension*) e foi concebido para suprir as "falhas da memória humana", por meio da utilização de recursos mecânicos. (LÉVY, 1993)

O Memex nunca foi construído, embora Bush o enxergasse como uma extensão natural das tecnologias existentes em 1945. Contudo, os conceitos relacionados a tal engenho visionário constituíram o substrato de pesquisas posteriores, como por exemplo, os estudos de Ted Nelson. Assim, como o Memex, o projeto Xanadu também não obteve o êxito esperado, no entanto considera-se que este último representa um avanço em relação ao Memex de Bush, e que de certo modo os conceitos subjacentes a tais projetos se concretizam a partir do desenvolvimento da Internet. A Internet, porém, limita-se aos recursos tecnológicos hoje existentes, não alcançando, por exemplo, a amplitude da proposta de Bush no sentido de permitir ao leitor, com o puro e simples apertado de um botão, sobrepor trajetos àqueles já existentes, criando associações entre lexias de hipertextos distintos.

Nesse contexto, Santos (1997) comenta a respeito da idéia de “absoluta novidade” do paradigma hipertextual defendida por alguns teóricos, a qual pode ser considerada como na realidade uma renovação ou desdobramento daquilo que a produção literária impressa, e anteriormente a tradição oral, já traziam consigo.

Conforme apresenta Lévy (1999, p.118), em seu livro *Cibercultura*, quando comenta a respeito dos reflexos culturais proporcionados pelo desenvolvimento do hipertexto:

Ele nos leva, de fato, à situação existente antes da escrita – mas em outra escala e outra órbita – na medida em que a interconexão e o dinamismo em tempo real das memórias on-line tornaram novamente possível, para os parceiros de comunicação, compartilhar o mesmo contexto, o mesmo imenso hipertexto vivo.

Segundo Lévy (1993), tecnicamente um hipertexto consiste numa rede composta de nós ligados por conexões. Os nós podem ser palavras, páginas, imagens ou partes de imagens, seqüências sonoras, referência a documentos complexos que podem ser eles mesmos hipertextos. Deste modo, pode-se considerar um hipertexto como um texto que possui marcações, que, ao serem clicadas, fornecem ao usuário outros dados sobre as informações relacionadas ao texto, dados os quais que no contexto da Internet podem estar na mesma página onde o texto é apresentado ou em qualquer parte do mundo. Tais marcações podem estar contidas não só em textos escritos, mas também em sons, imagens e animações, dando origem assim, a outros termos relacionados, como hipermídia e hiperdocumento, que generalizam para todas as categorias de signos os princípios da experiência hipertextual.

Nesta perspectiva, alguns pesquisadores encontram uma origem ainda mais remota para a idéia de hipertexto, observando uma das gravuras renascentistas apresentadas no livro *Le diverse et artificieuse machine* (1588), atribuído a Agostino Ramelli, apresentada a seguir na Figura 2.

Segundo Lara Filho (2003), a "roda de leitura" proposta por Ramelli tinha como finalidade permitir a consulta de vários volumes impressos simultaneamente sem que o leitor precisasse sair do lugar, bastando para tal girar a roda até o livro desejado. Deste modo, pode-se considerar Ramelli como um dos primeiros visionários do ambiente Web, porém é inegável que a idéia de Theodore Nelson, quando cunhou o termo hipertexto no contexto computacional, alterou radicalmente sua noção de espaço e amplitude em uma escala antes inimaginável, ou conforme afirma Lévy (1999, p.247), indo ao encontro de uma cultura do futuro, baseada no conceito de “universal sem totalidade”.

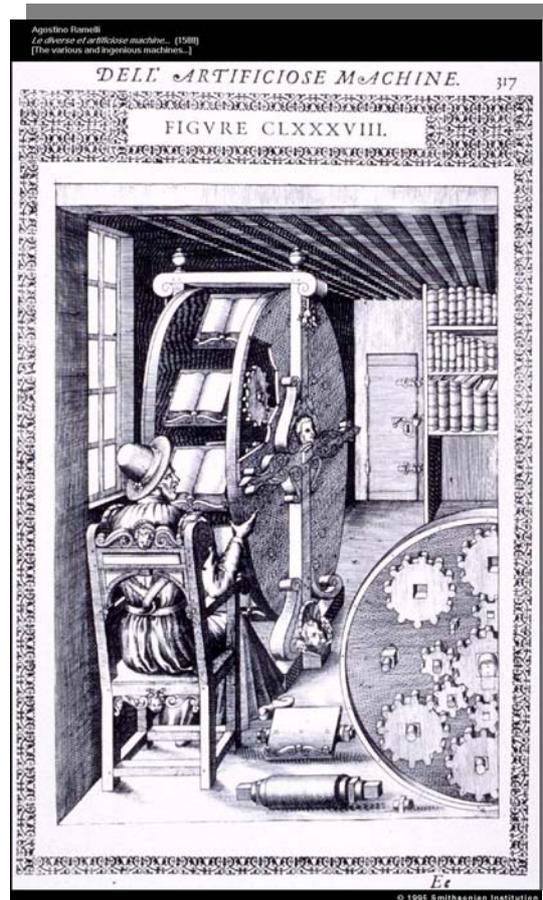


FIGURA 2 – “A Roda de Leitura”

Fonte: <http://www.scienceandsociety.co.uk/results.asp?image=10423516>

Acesso em: 31/08/2005

Outro pesquisador de importância histórica relevante, e que deve ser mencionado, é Douglas Engelbart, o qual reunindo sua experiência técnica às emergentes linhas de pesquisas em informática de sua época, apresentou anteriormente, concebeu a possibilidade de exposição de informações em uma tela. Executando um projeto junto ao *Stanford Research Institute*, Engelbart expôs a concepção de que seria possível ampliar as potencialidades intelectivas do ser humano a partir do alargamento dos horizontes mentais de construção do conhecimento. Assim para comprovar tal afirmação ele desenvolveu ferramentas tecnológicas que viabilizassem a intermediação entre o agente humano e o computador, de modo que em 1968 apresentou pela primeira vez na curta história da Computação uma tela com múltiplas janelas de trabalho, possibilitando manipulá-la, com a ajuda de um novo dispositivo, o qual batizou de *mouse*, utilizando tal dispositivo como uma extensão tecnológica do usuário, atuando sobre a virtualidade da imagem atualizada na tela do computador. (ENGELBART e ENGLISH, 1968)

Deste modo, pode-se afirmar que os processos técnicos assumiram uma relevante importância para o desenvolvimento da Internet, porém não se deve reduzir a história da Internet a uma história da técnica, afinal as técnicas sempre tiveram um papel fundamental no processo civilizatório, apesar de nas últimas décadas algumas correntes intelectuais adotarem a concepção de técnica como algo fundamentalmente estranho ao homem.

Evidentemente que as técnicas apontam perspectivas renovadoras e abrem portas para o futuro, permitindo novas visões de mundo antes inimagináveis, porém deve-se lembrar que a fusão entre técnica e cultura faz parte da própria experiência humana desde a primeira imagem gravada em pedra por um pintor das cavernas, sendo inegável a relação entre a atividade técnica e a visão simbólica das relações homem/mundo. Conforme afirma Machado (1993, p.11), “[...] é impensável uma época de florescimento cultural sem um correspondente progresso das suas condições técnicas de expressão, como também é impensável uma época de avanços tecnológicos sem conseqüências no plano cultural”.

Assim, verifica-se que o ambiente Web, e conseqüentemente a Internet, apresentam-se em constante evolução. Nessa perspectiva, em outubro de 1994 foi fundado no *Massachusetts Institute of Technology, Laboratory for Computer Science*, (MIT/LCS), com a colaboração do CERN, o *World Wide Web Consortium (W3C)*, um consórcio mundial liderado por Tim Berners-Lee que reúne empresas, instituições acadêmicas, profissionais e cientistas com o objetivo comum de desenvolver e padronizar novas tecnologias que possibilitem a evolução do ambiente Web, estendendo gradativamente suas funcionalidades. (JACOBS, 2006)

2.2 A evolução da Internet no Brasil

A Internet tem influenciado, a cada dia mais, o cotidiano de pessoas em todo o planeta, possibilitando evoluções em vários setores da sociedade. No Brasil tal realidade não é diferente, de modo que desde seus primórdios a Internet já despertava o interesse de pesquisadores brasileiros, constituindo o que alguns autores denominam como a primeira fase da evolução da Internet no Brasil, a qual se restringiu ao âmbito acadêmico. Conforme relata Pires (2005, p. 3):

As primeiras iniciativas para criar a rede e a infra-estrutura necessária para o desenvolvimento do ciberespaço no Brasil foram empreendidas por instituições públicas não comerciais, compostas eminentemente por instituições universitárias e de pesquisa, foram elas: o Laboratório Nacional de Computação Científica (LNCC) no Rio de Janeiro, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e o da Universidade Federal do Rio de Janeiro (UFRJ).

Tais iniciativas justificavam-se no interesse dos pesquisadores brasileiros em uma maior integração com pesquisadores de outras partes do mundo, principalmente com pesquisadores norte-americanos, possibilitando a ampliação de intercâmbios e a colaboração científica a partir do uso das tecnologias de informação e comunicação. Assim, na década de 1980 a Internet brasileira “viveu” um período de “gestação”, impulsionado pelo meio acadêmico, de modo que em 1987 pesquisadores e técnicos da Embratel se reuniram na Universidade de São Paulo, (USP), para discutir a montagem de uma rede que interligasse universidades brasileiras e internacionais. (STANTON, 1998)

Nesse contexto, em setembro de 1988 no Laboratório Nacional de Computação Científica (LNCC), anteriormente pertencente ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e atualmente subordinado ao Ministério da Ciência e Tecnologia (MCT), foi estabelecida a primeira conexão de uma instituição científica brasileira à rede BITNet (*Because It's There Network*), por meio de uma conexão com a Universidade de Maryland, em *College Park*. Em novembro do mesmo ano, uma segunda conexão foi estabelecida pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) às redes Bitnet e HEPNet, ligando-a ao *Fermi National Laboratory*, em Chicago, e em maio de 1989 a Universidade Federal do Rio de Janeiro (UFRJ) estabeleceu uma conexão de acesso à rede Bitnet interligando-se à UCLA (*University of Califórnia*), em Los Angeles. (PIRES, 2005)

Assim, a partir do reconhecimento da importância da utilização de redes de computadores para a comunidade acadêmica brasileira, surgiu a necessidade de coordenar tal infra-estrutura. Deste modo, em setembro de 1989 foi criada, pelo Ministério da Ciência e Tecnologia, a Rede Nacional de Pesquisa (RNP), com o objetivo inicial de construir uma infra-estrutura de rede nacional de âmbito acadêmico, disseminando o uso de redes de computadores no país e divulgando os serviços que a Internet poderia oferecer, estimulando, com tais ações, a formação de uma consciência

acerca da importância estratégica da Internet, de modo que a RNP tornou-se uma referência nacional em aplicação de tecnologias Internet. (STANTON, 1998)

Outro fato marcante da breve história da Internet brasileira data de fevereiro de 1991 quando, após aumentar a capacidade de seu sistema, a FAPESP estabeleceu a primeira conexão nacional utilizando o padrão TCP/IP, permitindo assim, a administração de domínios “.br” e a distribuição de endereços IP para todo o país. Deste então, a FAPESP passou a ser a regulamentadora da Internet brasileira, de modo que até hoje administra os domínios nacionais. (LUTFI, 1999)

Deste modo, a Internet gradativamente expandia-se pelo Brasil, porém restringindo-se ao meio acadêmico, então, com o intuito de popularizar a Internet em outros setores da sociedade. A partir de 1994, a RNP decidiu buscar apoio de empresas de telecomunicações, fato que segundo o coordenador da RNP nesta época Tadao Takahashi, citado por Lutfi (1999), gerou comentários do tipo: “vendilhões do templo”, mas que foi essencial para a transição para uma Internet aberta em nível nacional.

Assim, em 1995 teve início o processo de abertura da Internet no país, por meio da EMBRATEL e da RNP, sendo neste mesmo ano criado o Comitê Gestor da Internet no Brasil (CGI.br), a partir de uma iniciativa conjunta do Ministério da Ciência e Tecnologia e do Ministério das Comunicações, com o objetivo de coordenar a implantação do acesso à Internet nacional. (WIRTH, 2002)

Desde então, a FAPESP passou a dividir seu poder de regulamentação e administração da Internet brasileira com o Comitê Gestor da Internet do Brasil. Nesta época a maioria das universidades, entre públicas e privadas, já estavam conectadas à Internet, de modo que parte da população já aguardava ansiosamente a possibilidade de conectar-se à rede utilizando-se de acesso doméstico. Tal fato começou a se concretizar a partir de 1996, depois da autorização do então ministro das Comunicações Sérgio Motta. Assim, a Internet brasileira deixou de ser privilégio de universidades e passou a ser explorada comercialmente, ou seja, provedores poderiam cobrar pelo acesso à rede. (LUTFI, 1999)

Deste modo, iniciou-se o que alguns pesquisadores denominam como a segunda fase da evolução da Internet no Brasil, impulsionada pelo setor comercial, de modo que a partir de 1996 o número de usuários da Internet no Brasil cresceu exponencialmente. Segundo o Instituto Brasileiro de Opinião Pública e Estatística (IBOPE), o número de usuários passou de 700 mil em 1996 para cerca de 3,5 milhões em 1999 e segundo recentes pesquisas da *Internet World Stats*, em 2005, o Brasil encontra-se entre os 10 países do mundo com maior número de usuários conectados à Internet, ficando à frente inclusive de países como Canadá e Rússia, com aproximadamente 22,3 milhões de usuários, o que representa aproximadamente 2,4 % do total de usuários conectados no planeta, conforme apresentado a seguir na tabela 1.

TABELA 1 – Relação dos 20 países com maior número de usuários na Internet

	País	Usuários de Internet *	População (Estimada em 2005) *	% da População Incluída	% de usuários no Planeta
1	Estados Unidos	202,89	296,21	68,5 %	21,6 %
2	China	103,00	1.282,20	7,9 %	11,0 %
3	Japão	78,05	128,14	60,9 %	8,3 %
4	Alemanha	47,13	82,73	57,0 %	5,0 %
5	Índia	39,20	1.094,87	3,6 %	4,2 %
6	Reino Unido	35,81	59,89	59,8 %	3,8 %
7	Coréia do Sul	31,60	49,93	63,3 %	3,4 %
8	Itália	28,61	58,61	48,8 %	3,0 %
9	França	25,62	60,62	42,3 %	2,7 %
10	Brasil	22,32	181,82	12,3 %	2,4 %
11	Rússia	22,30	144,00	15,5 %	2,4 %
12	Canadá	20,45	32,05	63,8 %	2,2 %
13	Espanha	15,56	43,44	35,8 %	1,7 %
14	Indonésia	15,30	219,31	7,0 %	1,6 %
15	México	14,90	103,87	14,3 %	1,6 %
16	Taiwan	13,80	22,80	60,5 %	1,5 %
17	Austrália	13,79	20,51	67,2 %	1,5 %
18	Holanda	10,81	16,32	66,2 %	1,2 %
19	Polônia	10,60	38,13	27,8 %	1,1 %
20	Malásia	9,51	26,50	37,9 %	1,1 %

* Aproximações na escala de milhões de Pessoas.

Fonte: *Internet World Stats*, 2005. Disponível em:
<<http://www.internetworldstats.com/top20.htm>> Acesso em: 02/09/2005

Como se pode observar na Tabela 1, o Brasil não está alheio à evolução da Internet, porém, de acordo com pesquisas recentes, apenas cerca de 12 % da população brasileira tem acesso à Internet. Deste modo, um de seus maiores desafios, assim como dos demais países do mundo, é conectar as pessoas que ainda não têm acesso à rede, fato que sustenta o discurso de Alkalimat (1997), citado por Buchanan (1999), que afirmou que o dilema do século XXI seria a dicotomia entre as maravilhas tecnológicas oferecidas a uns poucos e uma realidade de fome, pobreza e exclusão por parte da maioria da população mundial.

2.3 Estrutura e componentes da Web

Segundo Daum e Merten (2002, p.6), o desenvolvimento da Web só foi possível a partir da definição de padrões abertos. Embora o conceito de padrão aberto não tenha se originado da Internet, tornou-se algo imprescindível para sua evolução, talvez pelo fato da Internet ter se popularizado muito no meio acadêmico antes de chegar ao grande público, o que de certo modo contribuiu para a criação de um “espírito comunitário” muito forte na rede.

Na indústria da informática, quando uma empresa desenvolve um sistema ou tecnologia é possível prevenir-se contra cópias da concorrência, bastando registrar sua criação, declarando seu padrão como proprietário e não disponibilizando todas as especificações para o público. Um padrão aberto é exatamente o oposto, quando o desenvolvedor de uma determinada tecnologia disponibiliza toda a especificação publicamente, de modo que qualquer pessoa possa copiar, desenvolver e comercializar equipamentos ou softwares baseados em tal tecnologia.

O fato é que na realidade há uma razão técnica fundamental para a estrutura da Internet se apresentar baseada em padrões abertos: a sua proposta tem como princípio a interconexão de redes de diversos tipos. Nesta perspectiva, ao utilizar padrões publicamente disponíveis para todos, tornou-se muito mais fácil a adaptação das diferentes redes e dos diferentes equipamentos utilizados, criados pelos mais variados fabricantes, pois caso tais padrões fossem proprietários, pertencentes a determinadas empresas, as mesmas teriam que desenvolver equipamentos e aplicativos que suprissem as necessidades de todos os ambientes de rede utilizados, fato que provavelmente inviabilizaria a idéia de uma Internet global.

Entre os padrões abertos que modelam a Internet, e de acordo com o enfoque desta pesquisa, pode-se destacar o *Hypertext Transfer Protocol* (HTTP), o qual pode ser entendido sucintamente como um conjunto de regras definidas para a comunicação entre os componentes da WEB. Segundo Krishnamurthy e Rexford (2001, p183), “[...] o HTTP foi proposto por Tim Berners-Lee em março de 1990, nos laboratórios do CERN, como um mecanismo suficientemente poderoso para acessar recursos em qualquer lugar da Internet e para ajudar a navegar entre eles por meio de links de hipertexto”. Assim, o HTTP constitui o modo mais comum de se transferir recursos na Web, definindo o formato das mensagens trocadas entre os componentes, quando, por exemplo, um usuário acessa uma página Web.

Nesse contexto, outro componente fundamental da Internet é o *Uniform Resource Identifier* (URI), Identificador Único de Recurso, um padrão para a codificação de nomes dos recursos e seus respectivos endereços na Internet. Segundo Berners-Lee et al. (1994) e Fielding (1995), um URI é um padrão conjunto que abarca os conceitos de *Uniform Resource Locator* (URL) e do *Uniform Resource Name* (URN), de modo que pode ser representado por qualquer um destes, ou por ambos.

Fazendo uma analogia de um recurso disponível no ambiente Web com um livro armazenado em uma biblioteca, pode-se considerar o URN de um recurso como o número ISBN de um livro, os quais fornecem uma identificação exclusiva, porém não oferecendo informações a respeito de onde o livro/recurso pode ser obtido. Do mesmo modo, pode-se considerar que o código que identifica onde um livro está localizado em meio ao acervo de uma biblioteca desempenha a mesma função do URL de um recurso no ambiente Web, indicando o local onde o livro/recurso pode ser obtido. Segundo Krishnamurthy e Rexford (2001, p.183), o modo mais popular de apresentação de um URI é utilizando um URL, o qual pode ser considerado como uma cadeia de caracteres formada por componentes padronizados. Informalmente, pode-se considerar que um URI é formado basicamente por três partes, conforme apresentado a seguir na Figura 3.



FIGURA 3 – Partes que constituem um URL

Assim, o URL “*http://www.marilia.unesp.br/images/biblioteca.gif*” identifica a localização exata de um recurso na Internet. No caso, uma imagem cujo nome é “*biblioteca.gif*”, que pode ser acessada utilizando-se o protocolo da Web “*http*” e que está localizada em “*www.marilia.unesp.br/images*”. Nesse contexto deve-se ressaltar que em um ambiente computacional o domínio “*www.marilia.unesp.br*” é representado pelo seu respectivo endereço IP, no caso “*200.145.141.1*”, conforme apresentado na Figura 3.

Desta maneira, verifica-se que para acessar um recurso disponível na Web é necessário identificar o seu URL, de modo que o *browser* ou navegador, software utilizado para acessar conteúdos da Web, possa localizar o recurso solicitado por meio do URL fornecido, e então realizar a transferência dos dados entre algum computador-servidor conectado à Internet e o computador-cliente que solicitou o recurso, sendo o protocolo HTTP o mais utilizado para realizar tal processo, conforme apresentado na figura 4.

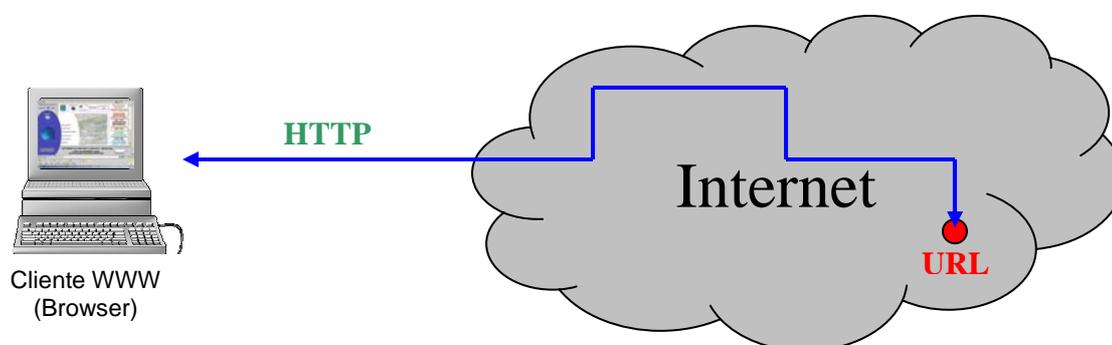


FIGURA 4 – Exemplo de uma comunicação na Web

Para a disponibilização de recursos informacionais no ambiente Web são utilizados documentos informacionais comumente chamados de páginas Web. Tais páginas são desenvolvidas utilizando-se linguagens computacionais denominadas como “linguagens de marcação”, sendo que uma linguagem de marcação é um tipo de linguagem computacional que, ao invés de utilizar comandos e funções para a realização de tarefas, utiliza “rótulos”, ou tecnicamente o que se denomina como “tags”, para “marcar” o conteúdo de um documento, possibilitando sua formatação e a incorporação de ligações, “links” de hipertexto, para outros recursos da Web.

A exemplo da maioria das tecnologias utilizadas no ambiente Web as linguagens utilizadas para o desenvolvimento de páginas Web são constituídas a partir de padrões abertos, e apresentam-se como ferramentas para tornar disponíveis os recursos informacionais no ambiente Web.

Nesse contexto, pode-se destacar a *Hypertext Markup Language* (HTML), a qual é considerada como a linguagem padrão para o desenvolvimento de páginas Web e sua origem confunde-se com a própria criação da *World Wide Web*. Tal linguagem possibilita a representação dos dados de forma simples, priorizando o modo de apresentação dos conteúdos, porém, não possibilita a descrição de estruturas semânticas ou o desenvolvimento de qualquer tipo de restrições de integridade.

A linguagem computacional HTML é derivada da linguagem *Standard Generalized Markup Language* (SGML), uma metalinguagem computacional utilizada para escrever outras linguagens, que é mais genérica e muito complexa, contendo mais de 500 páginas de especificações. A HTML utiliza o padrão ASCII para a representação dos conteúdos em formato de hipertexto.

Segundo Prasad (2003), o ASCII, *American Standard Code for Information Interchang*, é um padrão para a troca de dados proposto pelo *American National Standards Institute* (ANSI) em 1963 e aprovado em 1968, o qual se baseia no alfabeto romano e foi desenvolvido com o objetivo de padronizar a forma pela qual os computadores representam, em formato numérico, letras, números, acentuações e os mais variados sinais.

O padrão ASCII é considerado como um formato de texto simples, que pode ser identificado por qualquer computador. Inicialmente, tal padrão foi adotado para ser utilizado no desenvolvimento de aplicações voltadas para a Web, porém verificou-se posteriormente que o mesmo possui uma série de limitações.

Nesse contexto, deve-se ressaltar que a linguagem HTML possui um conjunto fixo de *tags* já pré-determinados, não permitindo a inclusão de novos. Com o intuito de exemplificar tal afirmação, basta considerar-se, por exemplo, que para “marcar” um trecho de texto informando que é um parágrafo e como tal deve ser tratado, e exibido pelo programa navegador, em HTML deve-se utilizar os *tags* <P> e </P> no início e no final de cada parágrafo, respectivamente. Assim, quando o programa navegador encontra um trecho de texto “cercado” por esses *tags*, interpreta-o como um parágrafo, exibindo-o de maneira adequada.

Apesar de sua simplicidade, a linguagem HTML permite a utilização de grande número de recursos, o que favoreceu sua popularização, de modo que a maioria das páginas Web atuais foram desenvolvidas utilizando tal linguagem. Segundo Araújo (2003, p. 48), tal fato se deve principalmente aos seguintes fatores:

- Ser extremamente simples;
- Possuir estilo próprio para a formatação de documentos;
- Possibilitar que as ligações de hipertexto sejam criadas facilmente;
- Apresentar suporte ao uso de formulários;
- Permitir a interação Homem-máquina;
- Possibilitar a programação simples, através do CGI (*Common Gateway Interface*);

Um dos problemas da HTML é que ela trabalha com forma e estrutura ao mesmo tempo, usando *tags* tanto para marcar trechos do conteúdo do documento quanto para informar ao programa navegador como tais trechos devem ser exibidos. Fato que complica o entendimento do código ao mesmo tempo em que também o limita, pois é possível apenas a utilização de *tags* que já estejam pré-definidos na especificação da linguagem. Outro fator de limitação da linguagem HTML é que ela impossibilita a atribuição de significados aos conteúdos das páginas, fazendo com que os motores de busca, *search engines*, recuperem um grande volume de dados que não são relevantes para os usuários finais, dificultando a recuperação de informações pertinentes em meio ao “oceano” de dados disponíveis no ambiente Web. Bosak (1997) denomina tais limitações como falta de “extensibilidade” da Linguagem.

Deste modo, verifica-se atualmente uma grande demanda de estudos relacionados ao desenvolvimento de mecanismos de representação de recursos informacionais. Nessa perspectiva, com o intuito de possibilitar o desenvolvimento de novas tecnologias que favoreçam a descrição dos aspectos semânticos inerentes aos recursos informacionais disponíveis no ambiente Web e a incorporação de formalismos lógicos que permitam aos computadores executar tarefas mais sofisticadas de maneira automatizada, originou-se o projeto Web Semântica, conforme apresentado no capítulo seguinte.

3 WEB SEMÂNTICA

No âmbito do tratamento e da recuperação da informação as formas de organizar e de recuperar sempre estiveram relacionadas às tecnologias vigentes, de modo que a constante evolução das tecnologias e o avanço exponencial na quantidade de informação disponível, principalmente no meio digital, vêm conduzindo a um estágio em que os modelos clássicos de organização e recuperação de informações precisam ser (re)pensados sob diferentes perspectivas.

Em decorrência desta nova realidade informacional, o ambiente digital passou a requerer métodos mais adequados, principalmente no que tange a formas de representação e organização de recursos informacionais. Dentro desse contexto, situa-se o projeto da Web Semântica, um título genérico que representa uma série de pesquisas que têm como principal objetivo possibilitar um melhor aproveitamento das potencialidades do ambiente Web, a partir do desenvolvimento de meios de organização de conjuntos específicos de informação e instrumentos de representação da informação.

Nessa perspectiva, este capítulo apresenta uma discussão de literatura acerca dos conceitos e tecnologias inerentes ao projeto Web Semântica, buscando identificar sua fundamentação teórica a partir da apresentação dos componentes básicos necessários para sua concretização, contextualizando as ontologias como instrumentos de modelagem semântica de objetos e conceitos e descrevendo as principais linguagens e formatos computacionais padronizados pelo W3C.

3.1 Definições

Em outubro de 1994, no *Massachusetts Institute of Technology* (MIT), foi fundado o *World Wide Web Consortium* (W3C), um consórcio mundial liderado por Tim Berners-Lee que reúne empresas, instituições acadêmicas, profissionais e cientistas, com o intuito de padronizar novas tecnologias que possibilitem estender gradativamente as funcionalidades do ambiente Web. Assim, a partir do empenho de um grande contingente de pesquisadores filiados ao W3C, em meados de 1998 começaram a

formalizar-se pesquisas relacionadas ao desenvolvimento de uma nova geração da Web, com o objetivo de possibilitar a incorporação de ligações semânticas aos recursos informacionais, de modo que os computadores possam “compreendê-las”. “*Machine understandable information*”, com esta sucinta expressão Berners-Lee (1998) impulsionou os primeiros estudos em direção ao projeto da Web Semântica.

Segundo Bernes-Lee (1999), o primeiro passo para o desenvolvimento da Web Semântica é a inclusão de dados em um formato que os sistemas computacionais possam naturalmente compreender de forma direta ou indireta. Após a publicação desta definição, em seu livro *Weaving the Web*, a expressão “Web Semântica” começou a ser disseminada como um título genérico que representa uma série de pesquisas que têm como principal objetivo possibilitar um melhor aproveitamento das potencialidades do ambiente Web, onde por meio do uso intensivo de linguagens computacionais e instrumentos de metadados espera-se obter o acesso automatizado às informações de maneira mais precisa, utilizando-se para isso processamentos semânticos de dados e heurísticas automáticas.

Conforme afirmam Berners-Lee et al. (2001, p.2, tradução nossa): “A Web Semântica é uma extensão da Web atual, onde a informação possui um significado claro e bem definido, possibilitando uma melhor interação entre computadores e pessoas”. Deste modo, é evidente que o objetivo final da Web Semântica é atender as pessoas e não os computadores, mas para isso torna-se necessário construir instrumentos que forneçam sentido lógico e semântico para as máquinas. Assim, pode-se verificar que a Web Semântica é uma tentativa inversa de solução, comparando-se com as tradicionalmente desenvolvidas, onde a idéia é pensar nas máquinas para que estas possam servir aos humanos de maneira mais eficiente.

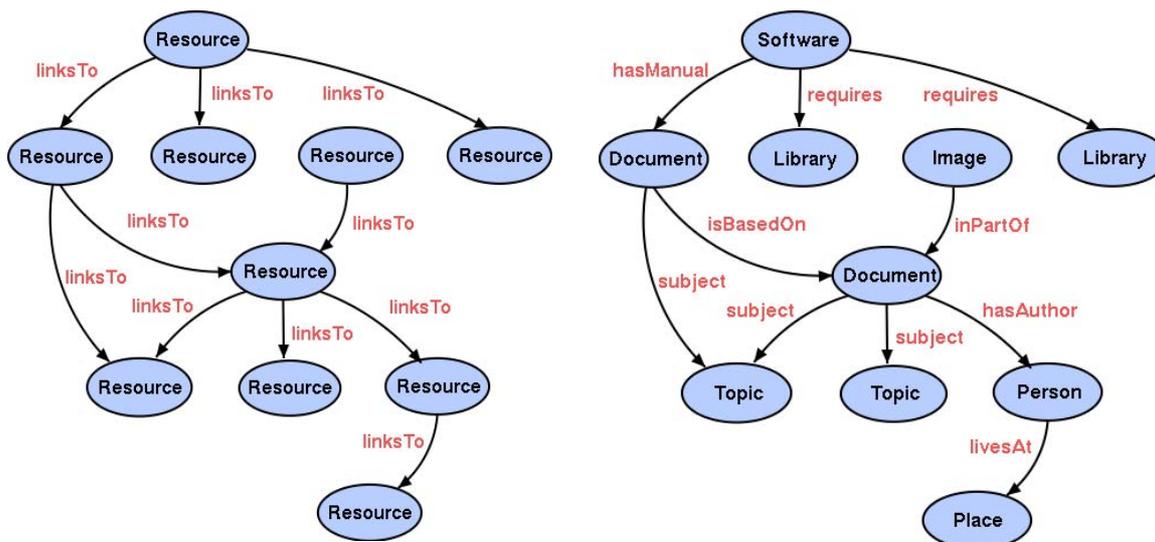
Para ilustrar as dificuldades de recuperação de informações na Web atual basta analisar os resultados de uma busca simples realizada por meio dos tradicionais “motores de busca” (*search engines*), como por exemplo, a busca de trabalhos científicos de autoria de uma pessoa específica. Nesta perspectiva, utilizando-se, por exemplo, “Arlindo Machado” como expressão de busca ter-se-ia como resultado todas as espécies de documentos contendo “Arlindo Machado” em alguma parte de seu conteúdo.

Caso esta mesma busca seja realizada utilizando-se apenas o sobrenome deste autor o problema se agravaria ainda mais, pois “Machado” pode igualmente se referir ao sobrenome de uma pessoa, com também a um instrumento cortante utilizado para rachar madeira e a Web atual não fornece condições que possibilitem aos tradicionais “motores de busca” distinguir entre os vários significados semânticos que um termo pode comportar, o que favorece a recuperação de uma grande quantidade de documentos irrelevantes ou não relacionados com a busca realizada, tornando muitas vezes inexecutável a tarefa de localizar uma informação específica no ambiente Web.

Segundo Koivunen e Miller (2001), um dos princípios básicos que constituem o projeto Web Semântica é o princípio de que “tudo” pode ser identificado por um URI, de modo que pessoas, lugares e elementos do mundo físico podem ser referenciados utilizando-se vários identificadores. Desta maneira, pode-se identificar um lugar, tal como a biblioteca universitária da UNESP de Marília, referindo-se ao URI de sua página Web, sendo possível também se referir a entidades físicas de modo indireto, como por exemplo, utilizando-se o URI da caixa de e-mail de uma pessoa.

Outra característica importante do projeto Web Semântica é que os *links* podem possuir diferentes tipos, possibilitando a definição de conceitos úteis para as máquinas, como por exemplo, que um recurso é uma versão de outro recurso ou que contém informações a respeito de uma determinada pessoa. A Web atual consiste de recursos e *links*, porém estes *links* são criados apenas para o uso de seres humanos, de modo que é relativamente simples para um ser humano identificar se um *link*, contido em um determinado recurso, referencia uma fatura, um romance ou um trabalho científico, porém tais informações não estão acessíveis para as máquinas, pois os *links* na Web atual não indicam quais são os tipos de relações existentes entre os recursos referenciados.

A figura 5, a seguir, ilustra como são referenciados os *links* na Web atual e alguns tipos de relacionamentos que se espera que a Web Semântica possibilite.



Web Atual

Web Semântica

FIGURA 5 – Estruturação dos recursos e links na Web atual e na Web Semântica
 Fonte: Miller, 2004, p.5-6.

Deve-se ressaltar que, de acordo com a Figura 5, disponível na página do W3C, os tipos de recursos não estão rotulados de maneira adequada, pois se dois recursos distintos estão ligados a um terceiro recurso a partir de um mesmo tipo de relacionamento, cada um deles deve ser identificado de maneira única, evitando-se assim ambigüidades. Deste modo seria mais adequado, por exemplo, utilizar rótulos como “*Library A*” e “*Library B*”, ao invés de rotular ambos os recursos como “*Library*”.

Outro fato que merece destaque é que na proposta original da *World Wide Web*, apresentada por Berners-lee em 1989, já estavam presentes tipos diferentes de relacionamentos entre recursos e links, conforme apresentado na Figura 1 (pág.22), porém tal característica não foi desenvolvida na Web atual, de modo que se espera que, com a concretização da Web Semântica, possa finalmente ser implementada.

Nesse contexto, um dos principais pontos de sustentação do projeto Web Semântica reside na utilização de ontologias⁶, de modo que se espera que, com o desenvolvimento de ontologias formais, seja possível descrever as informações semânticas dos recursos Web, possibilitando o compartilhamento e a manipulação de informações que possam ser interpretadas computacionalmente de maneira automática, a partir da utilização de regras lógicas.

⁶ A utilização do conceito de ontologia no âmbito do projeto Web Semântica é abordado na seção 3.3, a partir da página 54.

A figura 6 apresenta um esquema de como se espera realizar a recuperação de informações na Web Semântica, a partir de “motores de busca” mais eficientes, baseados em ontologias.

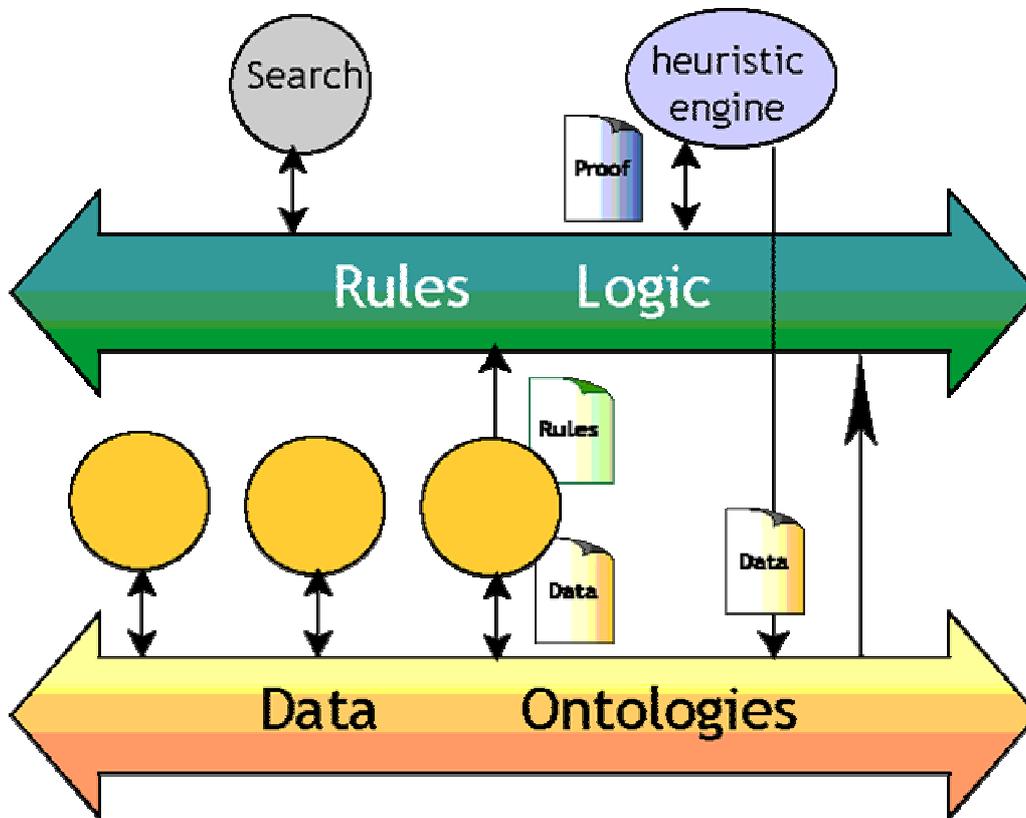


FIGURA 6 – Estrutura de recuperação de informações na Web Semântica
Fonte: Berners-Lee, 2000, p. 14.

Nesta perspectiva, para a concretização do projeto Web Semântica torna-se necessário padronizar a maneira pela qual os diversos tipos de softwares, utilizados no ambiente Web, manipulam as informações, assim como possibilitar meios que possam ser utilizados para descrever os aspectos semânticos de cada recurso, de modo que possa haver intercâmbio de informações de maneira padronizada, e que os recursos sejam descritos a partir de bases tecnológicas compatíveis. Deste modo, as novas aplicações desenvolvidas devem ser construídas a partir do padrão RDF e da estrutura formada pelos componentes que compõem o projeto Web Semântica, conforme pode ser observado na figura 7, a seguir.

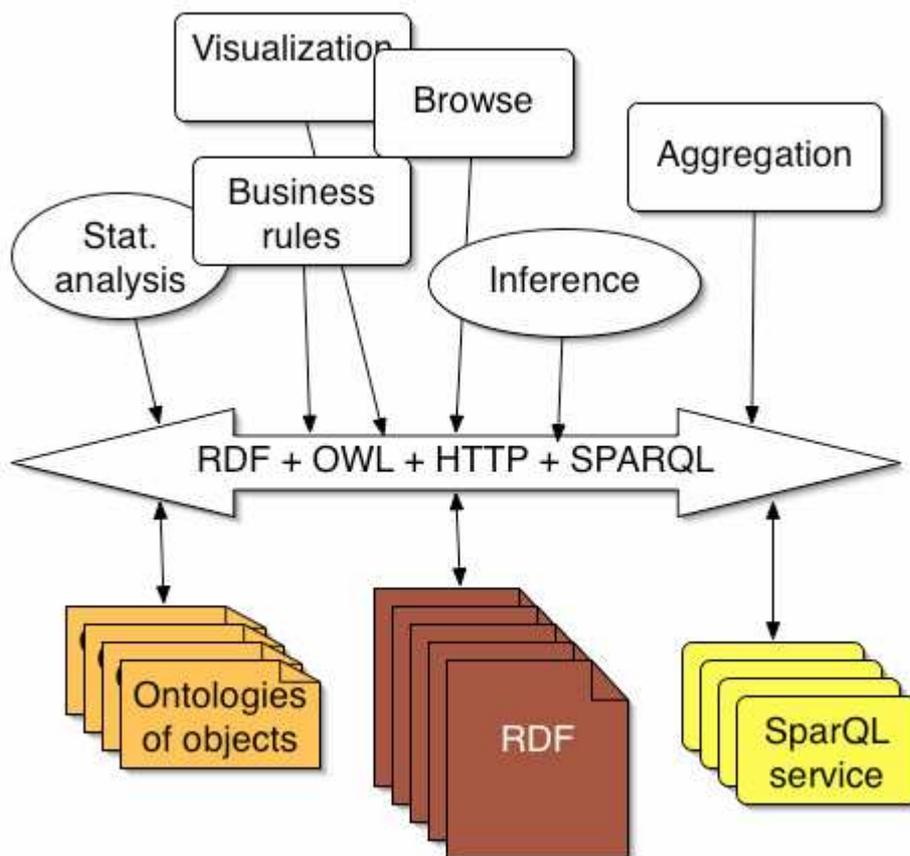


FIGURA 7 – Estrutura para o desenvolvimento de novas aplicações no âmbito da Web Semântica

Fonte: Berners-Lee, 2005, p. 17.

Nesse contexto também merece destaque o conceito de agentes computacionais. Segundo a definição de Souza e Alvarenga (2004), agentes computacionais são: “softwares que empregam técnicas de inteligência artificial com o objetivo de auxiliar o usuário na realização de determinada tarefa, agindo de forma autônoma e utilizando a metáfora de um assistente pessoal”.

Para Berner-Lee et al. (2001), os benefícios da Web Semântica poderão ser melhor explorados a partir do desenvolvimento de agentes computacionais que possibilitem coletar as informações advindas de fontes diversas, relacioná-las automaticamente e retorná-las de maneira mais organizada para os usuários, possibilitando inclusive a troca e o compartilhamento de informações entre agentes. A figura 8 ilustra como pode ser a utilização de agentes computacionais na recuperação de informações no âmbito da Web Semântica.

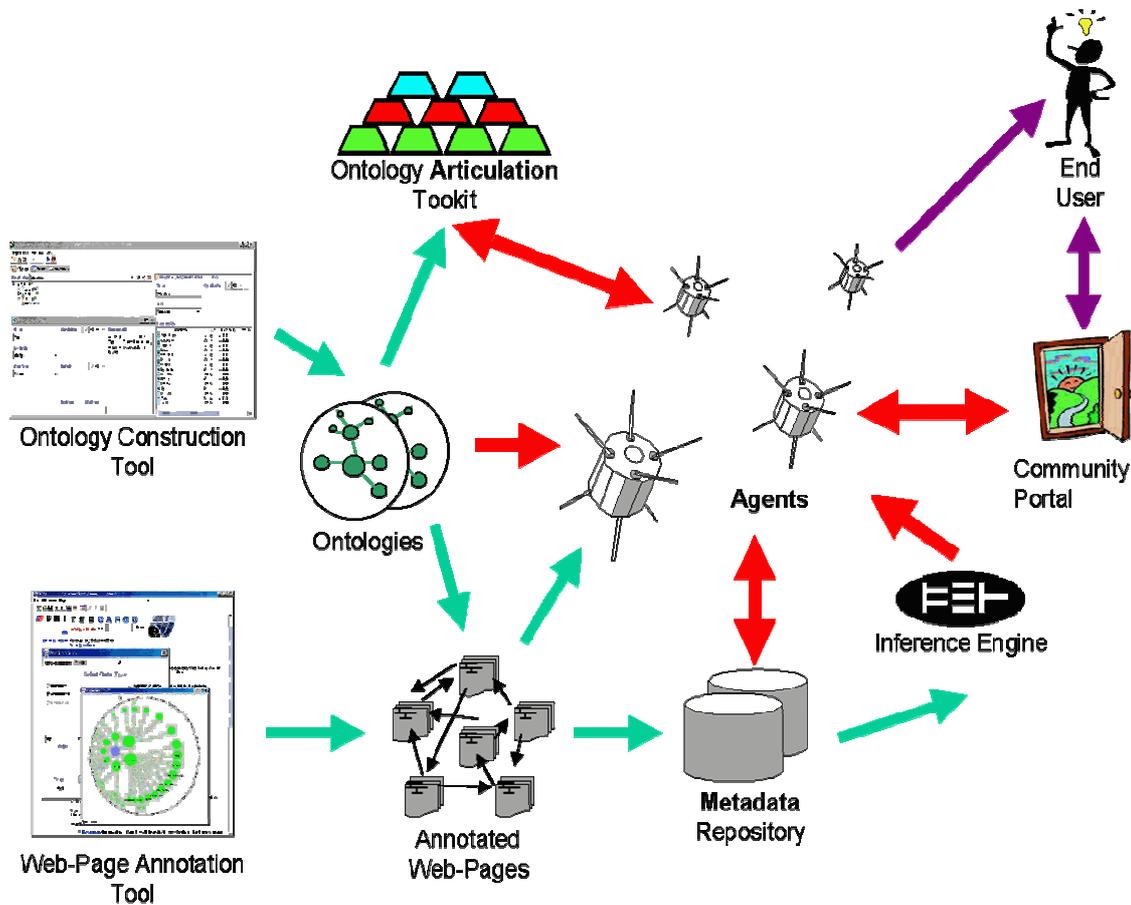


FIGURA 8 – A utilização de Agentes no âmbito da Web Semântica

Fonte: <http://www.semanticweb.org/about.html>

Acesso: 20/10/2005.

Assim, verifica-se que o projeto Web Semântica é composto por uma série de tecnologias inter-relacionadas, de modo que para a sua concretização torna-se necessária a criação de um ambiente adequado onde tais tecnologias estejam desenvolvidas, aplicadas e padronizadas de modo integral. Nesta perspectiva, com o intuito de facilitar o desenvolvimento e padronização das tecnologias relacionadas ao projeto Web Semântica, o W3C ilustrou os relacionamentos de tais tecnologias em uma arquitetura de camadas, conforme apresentado na seção seguinte.

3.2 Arquitetura da Web Semântica

O *World Wide Web Consortium* (W3C) desempenha um papel fundamental no desenvolvimento e padronização de novas tecnologias baseadas no ambiente Web, de modo que desde sua criação o W3C tem se empenhado em desenvolver e padronizar tecnologias diretamente relacionadas ao projeto Web Semântica.

Assim, em 2000, o W3C, tendo como seu maior expoente Berners-Lee, divulgou publicamente a primeira proposta de arquitetura da Web Semântica, conforme apresentado na figura 9, com base em uma série de camadas sobrepostas, onde cada camada ou tecnologia deveria obrigatoriamente ser complementar e compatível com as camadas inferiores, ao mesmo tempo em que não deveria depender das camadas superiores, possibilitando assim uma estrutura idealmente escalonável. Deste modo, tal arquitetura visava indicar os passos e as tecnologias necessários para a concretização do projeto Web Semântica, apresentando quais as tecnologias subjacentes e como estas se relacionam.

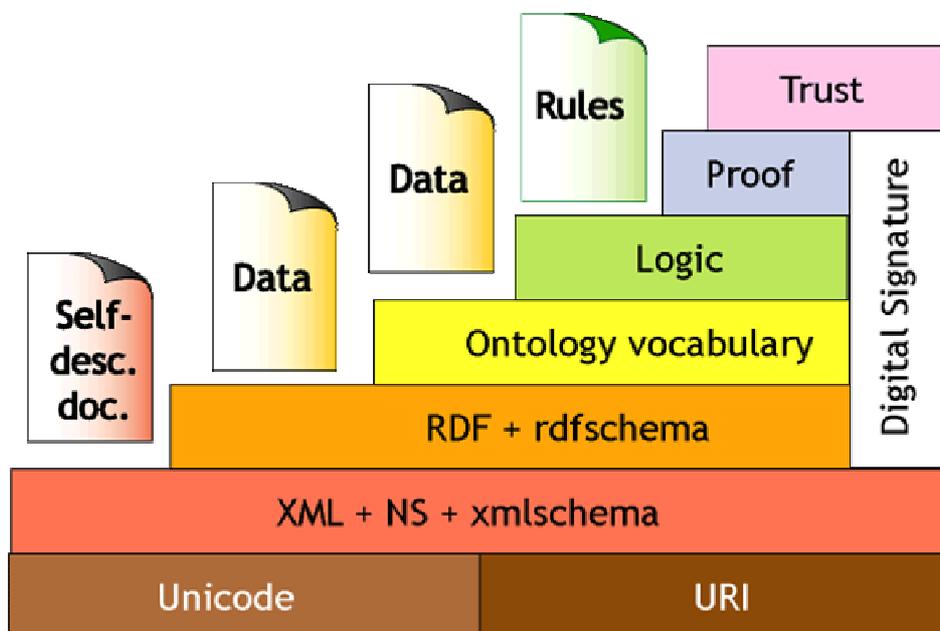


FIGURA 9 – Arquitetura da Web Semântica proposta em 2000
Fonte: Berners-Lee, 2000, p.10

Durante os anos que antecederam e sucederam a publicação desta proposta de arquitetura a grande maioria dos trabalhos relacionados com o projeto Web Semântica tinham como principal enfoque estudos sobre a linguagem XML e o seu relacionamento com a *Resource Description Framework* (RDF) (Miller, 1998; Lassila e Swick, 1999; Brickley e Guha, 2000; Brown et al., 2001). Tal fato justifica-se devido à falta de informações, nesta época, a respeito de como as camadas superiores seriam desenvolvidas, pois era necessário formular e padronizar as camadas iniciais. É por este motivo, também, que nesta primeira proposta foram apresentadas apenas recomendações de tecnologias até a camada “*RDF + rdfschema*”, de modo que a partir da camada “*Ontology vocabulary*” foram indicados apenas títulos genéricos de qual seria a função básica de cada camada, porém não apresentando tecnologias recomendadas para o seu desenvolvimento, como nas camadas anteriores.

Devido ao grande número de pesquisadores envolvidos no projeto Web Semântica, após sua divulgação inicial, não tardou para que as camadas iniciais fossem completamente desenvolvidas e padronizadas, dando origem, assim, a uma segunda proposta publicada pelo W3C em 2002, conforme pode ser visualizado na figura 10.

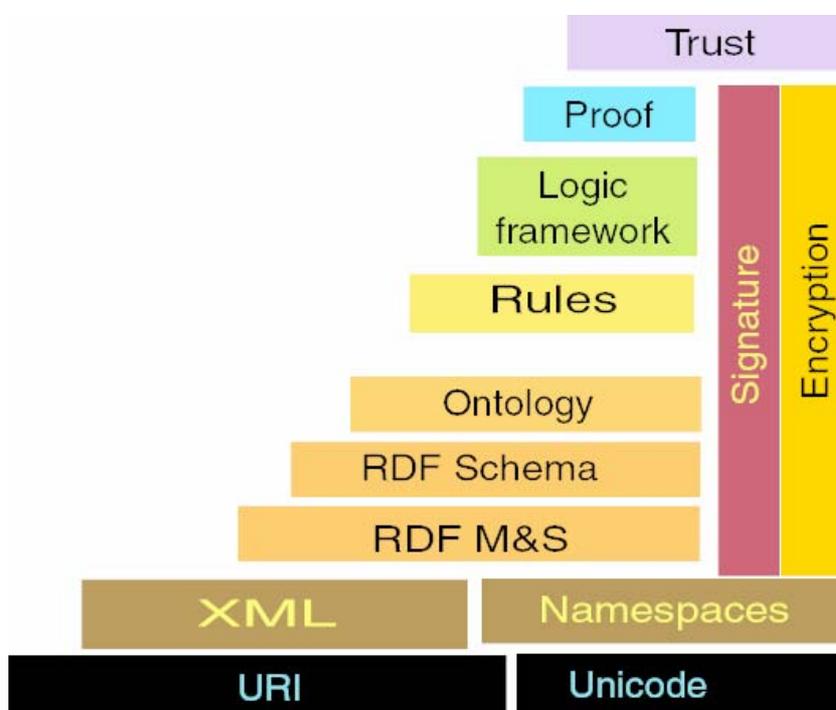


FIGURA 10 – Arquitetura da Web Semântica proposta em 2002
Fonte: Berners-Lee, 2002. p.24

A proposta de arquitetura publicada em 2002 não apresentou mudanças significativas em relação à arquitetura anterior, apenas o detalhamento de algumas tecnologias e o agrupamento de outras, como por exemplo, as especificações XML e xmlschema, pois, nesta época, a recomendação da linguagem XML já abarcava totalmente a especificação xmlschema. Já a camada RDF foi subdividida em uma camada contendo o Modelo e Sintaxe RDF e uma camada superior contendo especificamente o padrão RDF Schema. Foi acrescentada também uma indicação de criptografia, em paralelo à assinatura digital, com o intuito de garantir a confidencialidade das informações na Web Semântica.

Com as camadas iniciais totalmente desenvolvidas e padronizadas, após a publicação da proposta de arquitetura apresentada em 2002, as pesquisas relacionadas ao projeto Web Semântica concentraram-se principalmente na implementação de ferramentas e linguagens computacionais que possibilitassem o desenvolvimento de ontologias, conforme relata Ferneda (2003, p.118) ao mencionar que tais pesquisas tinham como enfoque principal o desenvolvimento de linguagens, bibliotecas, editores e sistemas que possibilitassem a integração de ontologias.

Deste modo, várias linguagens computacionais foram criadas especificamente para o desenvolvimento de ontologias, dentre elas: *Ontology eXange Language* (XOL), *Ontology Markup Language* (OML), *Ontology Inference Layer* (OIL) e *Darpa Agent Markup Language* (DAML).

Posteriormente, em 2004, como resultado do trabalho do *Web Ontology Working Group* - um grupo de trabalho do W3C que tem como principal objetivo desenvolver e padronizar especificações que possibilitem a utilização e o desenvolvimento de ontologias, foi publicada a recomendação do W3C para a utilização da *Ontology Web Language* (OWL) para o desenvolvimento de ontologias. A OWL foi criada a partir da junção das especificações das linguagens DAML e OIL, denominada inicialmente como DAML+OIL, constituindo assim um novo padrão de tecnologia recomendado pelo W3C. (BECHHOFFER et al., 2004)

Mesmo com a padronização da linguagem OWL para o desenvolvimento de ontologias, verificou-se que uma das principais dificuldades para concretização do projeto Web Semântica situava-se justamente na integração e compatibilidade da

camada de ontologias com as demais camadas da arquitetura, fato que culminou, recentemente, em uma nova proposta do W3C publicada em 2005.

Nessa proposta de arquitetura, a linguagem OWL já figura como recomendação para o desenvolvimento de ontologias e foram incorporadas tecnologias como “SparQL” e “DLP”, com o intuito de possibilitar a integração da camada de ontologias com as demais camadas e de facilitar a realização de consultas semânticas. Isso pode ser verificado na figura 11.

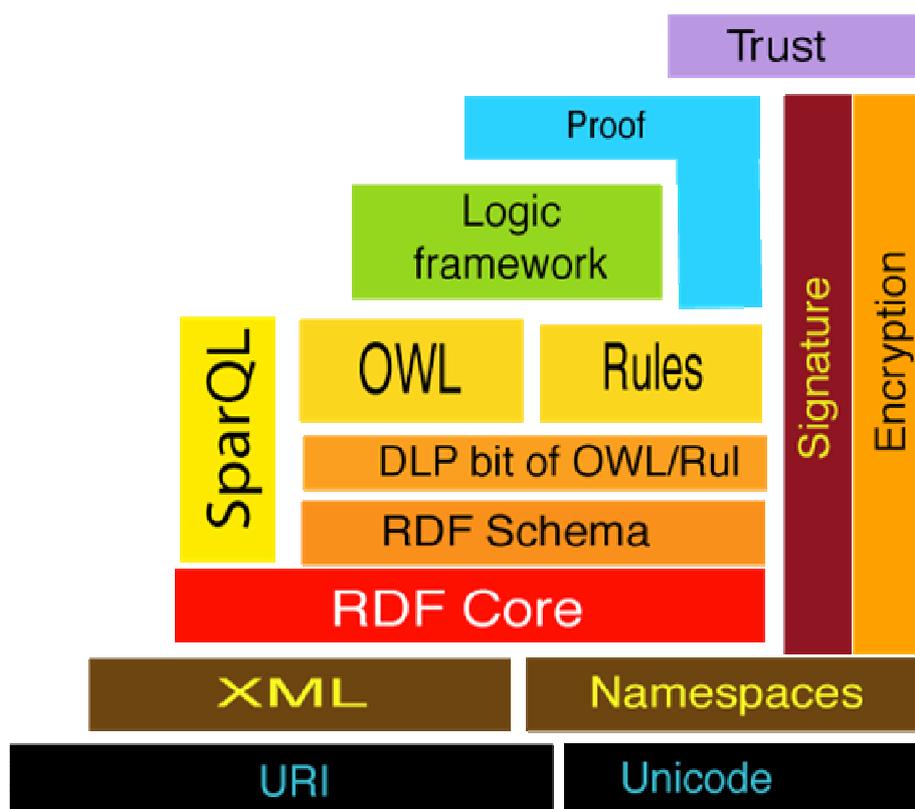


FIGURA 11 – Arquitetura da Web Semântica proposta em 2005
Fonte: Berners-Lee, 2005. p.17

Assim, baseando-se nesta proposta de arquitetura, pode-se descrever as principais tecnologias e camadas inerentes ao projeto Web Semântica, sucintamente, da seguinte maneira:

- **URI:** Conforme apresentado no Capítulo 2, tal componente consiste de um Identificador Único de Recursos que possibilita a definição e adoção, de maneira precisa, de nomes aos recursos e seus respectivos endereços na Internet.
- **UNICODE:** Esquema padronizado de codificação dos caracteres, que diminui consideravelmente a possibilidade de redundâncias dos dados, pois funciona independentemente da plataforma utilizada.
- **Signature:** Conjunto de tecnologias desenvolvidas com o intuito de substituir em ambiente computacional a função exercida pela assinatura formal de uma pessoa em um suporte físico. Segundo Pfützenreuter (2004), a assinatura digital garante a integridade dos dados e a comprovação da procedência dos recursos.
- **Encryption:** Consiste de um processo em que as informações são cifradas de modo que não possam ser interpretadas por qualquer pessoa ou sistema computacional, garantindo assim a confidencialidade das informações. Segundo Nakamura e Geus (2003, p.287), “[...] *encryption* é o processo de disfarçar a mensagem original, [...], de tal modo que sua substância é escondida em uma mensagem com texto cifrado”.
- **XML:** Conforme será apresentado detalhadamente na Seção 3.4.1, é uma linguagem computacional que possibilita a estruturação dos dados por meio da definição de elementos e atributos, e que permite a definição de regras sintáticas para a análise e validação dos recursos.
- **Namespace:** Coleção de nomes, identificados por um URI, que são utilizados em documentos XML para validar elementos e atributos, conforme será apresentado na Seção 3.4.1.
- **RDF Core:** Núcleo que compreende as especificações do modelo e a sintaxe da *Resource Description Framework* (Estrutura de Descrição de Recursos), possibilitando a descrição dos recursos por meio de suas propriedades e valores. Segundo Daum e Merten (2002), a RDF pode ser vista como uma tecnologia de capacitação para a modelagem semântica, sobre a qual podem ser criadas linguagens computacionais específicas.

- **RDF Schema:** Utilizada para a descrição do vocabulário RDF, possibilitando a definição de taxonomias de recursos em termos de uma hierarquia de classes. Segundo Brickley (2004), a RDF Schema é uma extensão semântica do código RDF, fornecendo mecanismos para descrever grupos de recursos relacionados e os relacionamentos existentes entre tais recursos.
- **SparQL:** Segundo recentes trabalhos apresentados por pesquisadores pertencentes ao W3C, (Prud'hommeaux e Seaborne, 2005; Clark, 2005), SparQL é uma linguagem computacional utilizada para realizar consultas a partir de estruturas RDF, favorecendo a recuperação de informações de maneira mais eficaz. Tal linguagem ainda não se encontra completamente padronizada, motivo pelo qual não é recomendada oficialmente pelo W3C, sendo denominada como uma tecnologia candidata à recomendação.
- **DLP:** A DLP é uma tecnologia candidata à recomendação e constitui a intersecção entre os dois principais paradigmas utilizados atualmente para desenvolver computacionalmente sistemas baseados em representação do conhecimento, Lógica Descritiva (OWL DL) e Programação Lógica (F-Logic), fornecendo uma estrutura extremamente flexível. (GROSOF et al. (2003) e VRANDECIC et al. (2005). Deste modo, a DLP ainda não é considerada atualmente como uma linguagem de representação do conhecimento, mas sim como uma “ponte” que possibilita a união entre os dois principais paradigmas utilizados.
- **OWL:** Linguagem computacional recomendada pelo W3C para o desenvolvimento de ontologias. Segundo McGuinness e Harmelen (2004), a linguagem OWL permite descrever formalmente, de modo mais eficiente, os aspectos semânticos dos termos utilizados e seus respectivos relacionamentos, possibilitando representações mais abrangentes das linguagens RDF e RDF Schema e favorecendo uma maior interoperabilidade.

- **Rules:** Permite a definição de regras lógicas relacionadas aos recursos informacionais. Segundo Daconta, Obrst e Smith (2003), esta camada possibilita uma espécie de “Introdução Lógica”, enquanto que a camada superior, Logic Framework, possibilita a incorporação de “Lógicas Avançadas”.
- **Logic Framework:** Camada para a definição de regras mais abrangentes para o tratamento das informações descritas nos níveis inferiores, possibilitando que agentes computacionais possam realizar inferências automáticas a partir das relações existentes entre os recursos informacionais, podendo inclusive inferir novas informações.
- **Proof:** Espera-se que esta camada possibilite a verificação/comprovação da coerência lógica dos recursos, de modo que os aspectos semânticos das informações estejam descritos de maneira consideravelmente adequada, atendendo a todos os requisitos das camadas inferiores.
- **Trust:** Camada de Confiança, a partir da qual se espera garantir que as informações estejam representadas de modo correto, possibilitando um certo grau de confiabilidade.

Como pode ser percebido, o projeto Web Semântica encontra-se em constante desenvolvimento, de modo que as tecnologias propostas ainda estão em fase de avaliação e de verificação de seus resultados. Nesta perspectiva, até mesmo a proposta de arquitetura publicada em 2005, há poucos meses, já tem recebido críticas e sugestões de alterações. Segundo Horrocks et al. (2005), algumas das linguagens apresentadas nesta arquitetura não são semanticamente compatíveis de forma direta, como a *Description Logic Programs* (DLP) e a linguagem de desenvolvimento de ontologias OWL, deste modo, tais pesquisadores propõem que estas tecnologias deveriam ser apresentadas paralelamente ao invés de sobrepostas, formando duas torres em uma parte da arquitetura, conforme figura 12, a seguir.

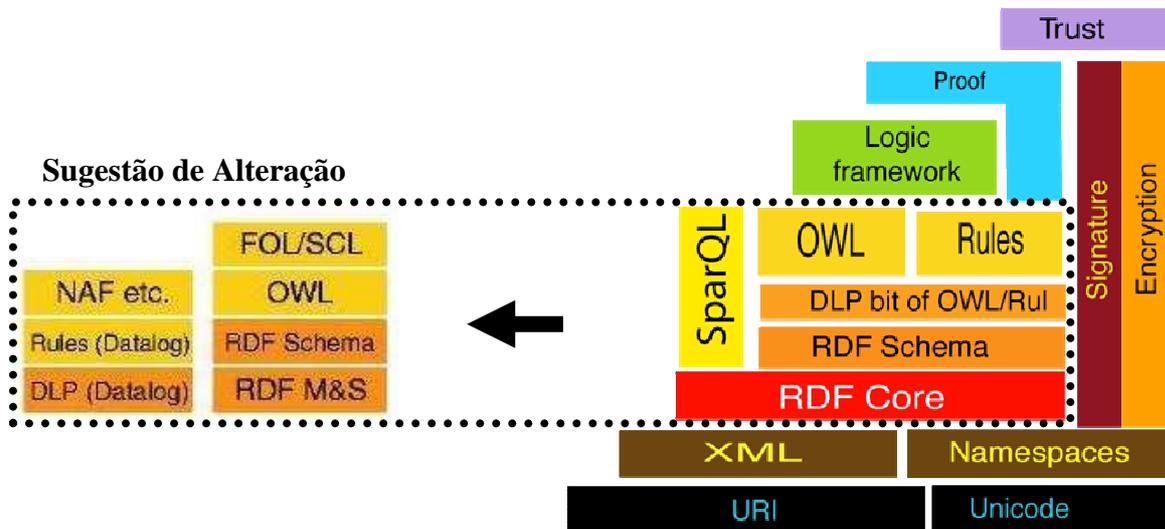


FIGURA 12 – Sugestão de alteração na Arquitetura proposta para Web Semântica
 Adaptado de: Berners-Lee, 2005, p.17 e Horrocks et al., 2005, p. 3.

Assim, observa-se que é muito provável que a arquitetura da Web Semântica ainda sofrerá modificações, para comprovar esta tendência basta verificar-se que as camadas “*Logic Framework*” e “*Prof*”, por exemplo, ainda não possuem sequer tecnologias recomendadas para suas implementações, pois apesar da padronização da linguagem OWL, como recomendação para o desenvolvimento de ontologias, atualmente ainda não está muito claro como esta camada irá relacionar-se com as demais camadas da arquitetura, informação esta imprescindível para o desenvolvimento das camadas superiores.

Nesse contexto, é importante ressaltar que apesar das novas tecnologias incorporadas à última proposta de arquitetura publicada pelo W3C e das recentes críticas e sugestões de alterações de tal arquitetura, é possível identificar aspectos que não devem ser alterados no projeto Web Semântica, pois mesmo considerando que novas tecnologias estarão sempre sendo desenvolvidas, os conceitos básicos que norteiam o desenvolvimento do projeto Web Semântica tendem a permanecer estáveis.

Deste modo, baseando-se nos conceitos básicos e nas funções das principais tecnologias apresentadas na figura 11, apresenta-se um esboço, figura 13, de um “Espectro Funcional”, omitindo detalhes técnicos para facilitar a compreensão das principais características e funções que se espera de cada uma das camadas da arquitetura da Web Semântica.

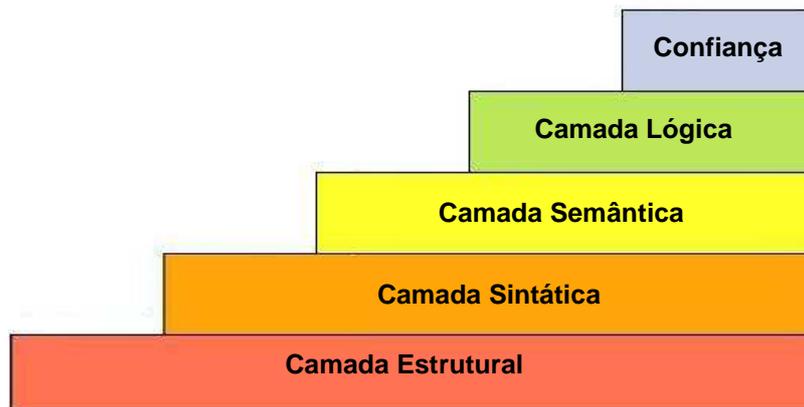


FIGURA 13 – “Espectro Funcional” da Web Semântica

Este “Espectro Funcional” utiliza-se de uma série de categorias para representar genericamente as funções das principais tecnologias inerentes ao projeto Web Semântica. Desta maneira, é possível descrever as funções de cada uma das categorias que representam as camadas desta arquitetura da seguinte forma:

- **Camada Estrutural:** Constitui o alicerce para todas as demais camadas, possibilitando a identificação dos recursos de forma única e padronizada e fornecendo meios seguros para representação, armazenamento e transmissão das informações, utilizando-se de mecanismos que forneçam garantias de integridade e confidencialidade aos dados.
- **Camada Sintática:** Fornece meios para a verificação da consistência dos recursos, por meio da definição e validação de regras sintáticas formalmente descritas, possibilitando a estruturação dos conteúdos associados a cada recurso.

- **Camada Semântica:** Permite a criação de vocabulários para a descrição dos aspectos semânticos dos recursos e a definição das relações existentes entre estes, a partir de especificações formais, explícitas e compartilhadas de conceitos.
- **Camada Lógica:** Define regras lógicas que possam ser verificadas computacionalmente, permitindo a realização de inferências automáticas e a verificação do nível de coerência lógica dos recursos.
- **Confiança:** Camada onde é realizada a comprovação de que os aspectos semânticos das informações estão descritos de modo consideravelmente adequado, atendendo a todos os requisitos das camadas anteriores e possibilitando um certo grau de confiança das informações.

Desta forma, é possível verificar que para a concretização do projeto Web Semântica é necessária a utilização intensiva de lógicas computacionais que possibilitem inferir novas informações, baseadas nos significados semânticos dos conteúdos das páginas Web. Dentro desta perspectiva, seria possível tecer uma extensa rede de conhecimento a partir do ambiente Web, porém ironicamente a sucinta definição “*Machine understandable information*”, aparentemente simplista, a qual deu origem às primeiras pesquisas relacionadas ao projeto Web Semântica, representa um dos maiores desafios a ser superado, pois ainda não existe um consenso de como se deve representar informações no ambiente Web de modo a possibilitar aos computadores “a compreensão” dos significados de tais informações.

Nesse contexto, destaca-se atualmente uma forte tendência no desenvolvimento de pesquisas relacionadas à representação de informações, referindo-se a instrumentos de modelagem cognitiva de objetos digitais por meio da utilização do termo ontologia, conforme será apresentado na seção seguinte.

3.3 Ontologias na Web Semântica

No âmbito da Web Semântica verifica-se atualmente uma forte demanda para o desenvolvimento de pesquisas relacionadas aos instrumentos de modelagem cognitiva de objetos digitais, por meio da utilização do termo ontologia.

Segundo Lima (1998), um termo corresponde a um conceito particular dentro de um campo conceitual, designando um conjunto de propriedades e relações com outros conceitos em um determinado contexto. Deste modo, devido ao fato do termo ontologia ser utilizado por inúmeras comunidades científicas, tal termo apresenta muitas variações de significado e interpretações distintas, de acordo com o propósito esperado e o enfoque de cada área do conhecimento.

Historicamente a palavra ontologia tem origem no grego *ontos* (ser) e *logos* (palavra), de modo que apesar do estudo do *ontos* originar-se nos estudos de Aristóteles e Platão, o uso do termo Ontologia para designar um ramo da Filosofia é muito mais recente, tendo sido introduzido na transição da Idade Média para a Idade Moderna, na escolástica, por volta dos séculos XVII e XVIII. Segundo Welty e Guarino (2001), o termo foi cunhado na área de Filosofia em 1613 por Rudolf Goclenius e aparentemente de forma independente por Jacob Lorhard.

Numa formulação bastante sucinta pode-se dizer que Ontologia no contexto filosófico é o ramo que estuda tudo aquilo que existe, ou utilizando-se a formulação clássica de Aristóteles, estuda o Ser enquanto Ser, do Ser concebido como tendo uma natureza comum que é inerente a todos e a cada um dos Seres. Deste modo, pode-se classificar uma ontologia como um sistema de categorias que explicam uma certa visão do mundo. Contudo, tal definição de Ontologia acarreta alguns problemas, mesmo dentro do campo da Filosofia, pois a partir de tal definição genérica este ramo da Filosofia também englobaria questionamentos como, por exemplo: “O que é uma obra de arte? Ou no que consiste um raciocínio válido?”. Entretanto, conforme afirma Teixeira (1999), os problemas relacionados com tais questionamentos constituem o objeto de estudo da estética e da lógica, respectivamente.

Assim, recorrendo-se a definição de Grossmann, citado por Teixeira (1999), Ontologia é o ramo da Filosofia que tem como objetivo responder às seguintes questões: “Quais são as categorias de mundo? E quais são as leis que regulam tais categorias?”. Deste modo, observa-se que na Filosofia o termo Ontologia é frequentemente contrastado com a Epistemologia, e nesse contexto quando se classifica uma Ontologia como um sistema de categorias que explicam determinada visão do mundo, tal sistema não depende de uma linguagem em particular, pois a Ontologia é sempre a mesma, independente da linguagem utilizada para descrevê-la.

Quanto à utilização do termo ontologia no âmbito da representação do conhecimento, sua origem remete aos estudos apresentados no início da década de 1990 na área de Ciência da Computação, mais especificamente na subárea de Inteligência Artificial (IA), em projetos voltados para a organização de bases de conhecimento.

Nesta perspectiva, Guarino e Giaretta (1995) apresentam no artigo intitulado *Ontologies and Knowledge Bases*, uma compilação de sete possíveis interpretações para o termo ontologia, limitando o escopo às acepções mais comumente encontradas, conforme apresentadas a seguir:

1. Ontologia como uma disciplina da Filosofia;
2. Ontologia como um sistema conceitual informal;
3. Ontologia como uma proposta semântica formal;
4. Ontologia como uma especificação de uma conceitualização;
5. Ontologia como uma representação de um sistema conceitual por meio de uma teoria lógica:
 - 5.1 Caracterizada por propriedades formais ou
 - 5.2 Caracterizada apenas para propósitos específicos;
6. Ontologia como um vocabulário usado por uma teoria lógica;
7. Ontologia como um meta-nível de especificação de uma teoria lógica.

De acordo com tais definições, com exceção da primeira que se refere ao sentido filosófico do termo, pode-se identificar duas grandes correntes teóricas:

- A) Que concebem ontologia como uma entidade conceitual semântica, formal ou informal, (definições 2 e 3);
- B) Que concebem como um objeto concreto em nível sintático, que tem seu desenvolvimento e sua utilização guiados por um propósito específico, (definições de 4 a 7).

Nesse contexto, a definição 4, que define uma Ontologia como uma especificação de uma conceitualização, é a mais freqüentemente identificada no âmbito da representação do conhecimento, tendo como seu maior expoente a definição de Gruber (1993, p.1, tradução nossa), que define uma ontologia como: “uma especificação explícita de uma conceitualização”, considerando que o termo “explícita” significa que um objeto de nível simbólico deve ser expresso formalmente e de maneira clara, e uma “conceitualização” será composta por objetos, conceitos e as relações existentes em um determinado domínio.

Apesar de tal definição ser a mais referenciada na literatura relacionada às ontologias, no contexto da representação do conhecimento Guarino e Giaretta (1995) apontam problemas quanto à noção de conceitualização utilizada por Gruber, devido ao fato de considerar uma conceitualização como um conjunto de relações extensíveis (*extensional relations*) descrevendo um estado particular das coisas, enquanto que comumente a noção que se tem de conceitualização é de algo intencional, como uma grade que deve ser preenchida com os vários estados das coisas.

Assim, Guarino (1998, p.5, tradução nossa), preocupando-se em estender a definição apresentada por Gruber(1993), define ontologia como: “uma teoria lógica para relacionar o significado pretendido de um vocabulário formal, seu comprometimento com uma conceitualização particular do mundo”.

Deste modo, torna-se conveniente estabelecer uma definição formal de conceitualização, que pode ser apresentada como uma tripla ordenada $C = \langle D, E, \mathfrak{R} \rangle$,

onde “D” representa um domínio, “E” o conjunto máximo de estados relacionais desse domínio e “ \mathfrak{R} ” o conjunto de relações conceituais sob o espaço de domínio “ $\langle D,E \rangle$ ”. De acordo com tais considerações, uma ontologia é uma teoria lógica cujo modelo restringe uma conceitualização particular, sem especificar exatamente qual, ou em outras palavras, pode-se definir como uma caracterização axiomática do significado de um vocabulário lógico, a qual tem o compromisso apenas com a consistência em um determinado domínio, e não com a completude.

Nessa perspectiva, Guarino (1998, tradução nossa), ainda apresenta uma interessante distinção quanto à utilização do termo ontologia para designar instrumentos de representação do conhecimento, e o sentido que tal termo abarca no âmbito filosófico:

No sentido do filosófico, nós podemos referir a uma ontologia como um sistema particular de categorias que consideram uma certa visão do mundo. Como tal, este sistema não depende de uma linguagem particular: ontologia de Aristóteles é sempre a mesma, independentemente da linguagem utilizada para descrevê-la. Por outro lado, em seu uso mais predominante na IA, uma ontologia é referida como um artefato de engenharia, constituído por um vocabulário específico utilizado para descrever uma certa realidade e um conjunto de pressupostos explícitos relacionados com o significado pretendido para as palavras do vocabulário.

Dentro deste complexo cenário, com o intuito de tentar diminuir as dificuldades de comunicação entre áreas, Guarino e Giaretta (1995) propuseram uma diferenciação entre a ontologia estudada pela área de Ciência da Computação e a Ontologia filosófica de Aristóteles, segundo a qual “uma ontologia” (com o artigo indefinido e inicial minúscula), diz respeito a um determinado objeto em particular, enquanto “Ontologia” (sem o artigo indefinido e com a inicial maiúscula), refere-se à disciplina filosófica que lida com a natureza e a organização da realidade.

Desta maneira, a partir das definições apresentadas, é possível verificar que as ontologias no contexto da Ciência da Computação, e no âmbito da Web Semântica, são utilizadas com o intuito de criar modelos formais para a representação do conhecimento, dentro de um determinado domínio restrito, não possuindo “propriedades transcendentais” ou “formas substanciais” e tendo como meta maior o registro de informações por meio de linguagens que possam ser processadas computacionalmente, favorecendo a realização de inferências automáticas.

Assim, com o intuito de esclarecer a diferença entre ontologia e base de conhecimentos, pode-se considerar uma ontologia para o domínio acadêmico, proposta para fins estritamente didáticos, limitando a abrangência de tal ontologia a poucos conceitos, a saber: pessoa, professor, aluno, curso, disciplina, avaliação, trabalho e prova. Deste modo, é possível descrever graficamente as relações existentes entre tais conceitos, conforme apresentado na figura 14, a seguir.

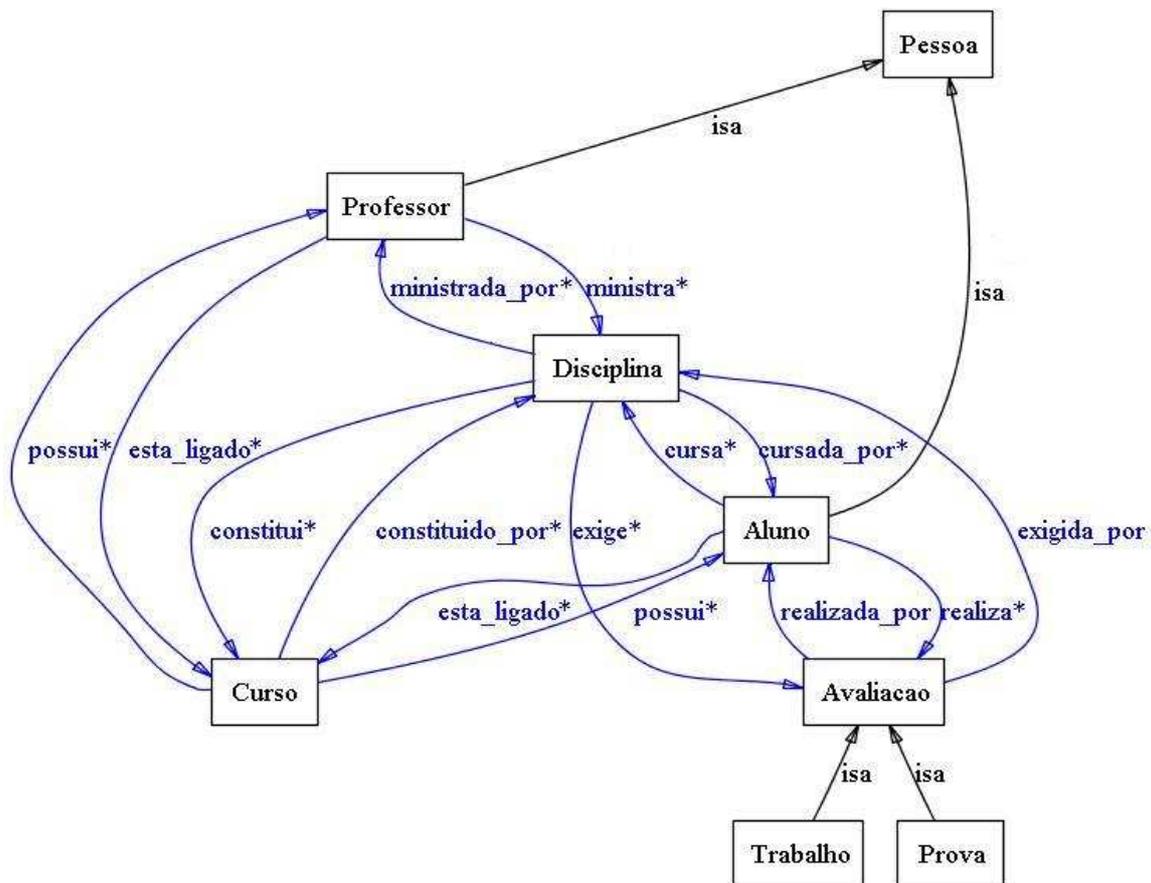


FIGURA 14 – Diagrama de uma ontologia de um domínio acadêmico

Nesta perspectiva, pode-se utilizar a estrutura fornecida pelas ontologias para a construção de bases de conhecimento, descrevendo um conjunto de conceitos e as relações existentes entre estes em um domínio específico, possibilitando o armazenamento de informações referentes a um determinado contexto. Deste modo, caso o contexto seja modificado, automaticamente a base de conhecimentos também será modificada, em contra partida a ontologia permanecerá inalterada, desde que o domínio permaneça inalterado.

A partir da ontologia apresentada na figura 14, embora sua abrangência seja limitada, é possível construir uma base de conhecimentos, descrevendo a realidade de um domínio acadêmico, armazenando informações contextualizadas a respeito de um determinado curso ou programa. Deste modo, verifica-se que a partir de uma mesma ontologia é possível construir bases de conhecimentos distintas, pois as realidades de cada curso ou programa são diferentes de acordo com o contexto ao qual estão inseridos, porém a ontologia permanece inalterada, desde que o domínio não seja modificado.

Desta maneira, verifica-se que a utilização de ontologias no âmbito da Web Semântica favorece o compartilhamento da mesma estrutura de informações entre pessoas e softwares, pois possibilita a descrição formal das relações existentes entre os objetos em um formato que as máquinas possam identificar, permitindo inclusive o reuso de conhecimentos dentro de um determinado domínio, pois torna possível, por exemplo, associar uma ontologia a uma página Web, definindo o significado de cada uma das informações existentes e possibilitando a integração e reutilização de ontologias entre diversos domínios. A partir desta perspectiva, uma página poderia ser relacionada automaticamente com outras, utilizando regras de inferência, e possibilitando inferir novas informações.

Assim, verifica-se que no contexto do projeto Web Semântica a utilização do termo ontologia refere-se ao desenvolvimento de instrumentos de representação do conhecimento definidos em uma linguagem formal e processável por máquina que possibilitem a descrição dos aspectos semânticos dos conteúdos informacionais, explicitando seus relacionamentos de modo detalhado a partir de restrições lógicas que possam ser processadas de forma automatizada, possibilitando inclusive relacionamentos baseados na Lógica de Segunda Ordem.

Deste modo, observa-se que as ontologias representam uma evolução no modo como os recursos informacionais são disponibilizados no ambiente Web, favorecendo melhorias significativas nos processos de recuperação de informações. Conforme afirma Mucheroni et al. (2004, p.3), “tais recursos passam a ter um papel fundamental no que diz respeito a processos de pesquisa e recuperação de dados de uma forma mais direta e objetiva”.

Nessa perspectiva, com o intuito de favorecer a concretização do projeto Web Semântica, o W3C tem se empenhado em desenvolver e padronizar novas linguagens computacionais para a representação de recursos informacionais, conforme apresentado na seção seguinte.

3.4 Linguagens de representação de recursos informacionais

Uma das principais características dos estudos relacionados ao projeto Web Semântica refere-se à preocupação em desenvolver linguagens computacionais que possibilitem estruturar os recursos informacionais de maneira adequada e descrever os aspectos semânticos inerentes a tais recursos.

Nesta perspectiva, logo após a padronização da linguagem computacional XML, os engenheiros de softwares descobriram que não era suficiente apenas descrever os recursos informacionais sintaticamente para que os sistemas colaborassem, verificando que tão importante quanto a sintaxe seria o desenvolvimento de tecnologias que permitissem descrever o significado das informações.

Dentro deste contexto, este capítulo apresenta algumas considerações acerca das principais linguagens computacionais padronizadas pelo W3C para o desenvolvimento da Web Semântica: XML, RDF e OWL.

3.4.1 XML

A recomendação XML foi divulgada em 1998 pelo W3C, baseando-se também no padrão SGML, assim como a linguagem HTML. Porém, a XML foi criada não como uma linguagem de uso especial, mas sim, como metalinguagem genérica, constituindo uma tecnologia básica capaz de possibilitar o desenvolvimento de outras linguagens computacionais. Segundo documento do W3C, publicado por Bray, Paoli e Sperberg-McQueen (1998, tradução nossa), os objetivos iniciais que pautaram o desenvolvimento da linguagem XML foram:

1. XML deve ser utilizada de forma direta e objetiva na Internet.
2. XML deve suportar uma ampla gama de aplicativos.
3. XML deve ser compatível com SGML.
4. Deve ser fácil desenvolver programas que processem documentos XML.
5. O número de recursos adicionais na XML deve ser mantido em um nível mínimo, idealmente zero.

6. Os documentos XML precisam ser legíveis e relativamente claros.
7. O projeto XML deve ser preparado rapidamente.
8. O design XML deve ser formal e conciso.
9. Os documentos XML devem ser fáceis de serem criados.
10. A concisão na marcação XML é de importância mínima.

Um dos fatores que dificultaram inicialmente a disseminação da XML foi uma concepção equivocada de que esta se apresentava como uma linguagem sucessora da HTML, pois embora a XML resolva alguns dos problemas da HTML, ela foi desenvolvida com um propósito diferente: enquanto a linguagem HTML tem como função principal formatar e exibir o conteúdo de um documento, a linguagem XML possui a função específica de apenas estruturar as informações, não se preocupando com a maneira como estas serão exibidas.

Assim, a XML permite a criação de marcações definidas pelo próprio usuário, diferentemente da HTML, podendo ser considerada como um sistema gramatical para construção de linguagens de marcação personalizadas. Conforme relata Almeida (2002),

[...] XML tem uma importante característica adicional: permite ao autor do documento a definição de suas próprias marcas. Esta característica confere à linguagem XML “habilidades” semânticas, que possibilitam melhorias significativas em processos de recuperação e disseminação da informação.

Segundo Daum e Merten (2002, p27), “[...] o conjunto de caracteres da XML é UNICODE, permitindo que a XML contenha a maioria dos caracteres internacionais”. As codificações padrões utilizadas em XML são UTF-8 e UTF-16, salientando que o código ASCII é um subconjunto do UTF-8.

Para entender melhor a utilização de tais codificações deve-se lembrar que os computadores fundamentalmente lidam com números para representação dos dados, de modo que gravam letras e outros caracteres na memória designando um valor numérico para cada um deles. Deste modo, antes do UNICODE ser inventado, havia centenas de sistemas diferentes de codificação, nenhum destes, porém, representava um número suficiente de caracteres que possibilitassem abarcar todos os símbolos lingüísticos necessários para constituir um sistema de codificação mundial, podendo inclusive ser conflitantes entre si. Em outras palavras, os sistemas de codificação poderiam utilizar o

mesmo número para dois caracteres diferentes ou utilizar números diferentes para representar o mesmo caracter, tal fato favorecia a possibilidade de redundância e inconsistência dos dados, pois os mesmos poderiam ser corrompidos sempre que houvesse a necessidade de troca de dados entre codificadores ou plataformas diferentes. Segundo (Araujo, 2003 p.7),

UNICODE é um esquema de padronização de codificação de caracteres onde um número representará um caracter. O UNICODE vem a substituir centenas de codificadores utilizados por exemplo pela União Européia, onde a língua inglesa requer vários codificadores para representar todas as letras, pontuação e símbolos técnicos.

Assim, como o conjunto de caracteres da XML é UNICODE, um documento XML pode ser escrito em qualquer editor de textos, independentemente da plataforma utilizada, pois será representado de uma maneira “universal”.

Nesse contexto, é inegável o fato de que a possibilidade do próprio desenvolvedor criar os “elementos” que serão utilizados para construir o documento, através de “marcações” personalizadas, propicia uma maior flexibilidade e liberdade na hora da implementação, porém em contra partida torna-se necessário definir um conjunto de regras que devem ser seguidas no momento do desenvolvimento e do processamento do documento, indicando como o mesmo deve ser interpretado pelos softwares. Os arquivos que contém um conjunto de regras para a formação de um documento XML, ao qual se associam, independentemente do padrão utilizado para sua criação, são denominados esquemas.

Segundo Castro (2001), a especificação XML é simples, mas o que a torna poderosa são as tecnologias que ela possibilita utilizar, tais tecnologias também são desenvolvidas a partir de padrões apresentados pelo W3C, constituindo o que se denomina como padrões companheiros. Atualmente existem dois padrões para a construção de esquemas em XML, o primeiro destes é denominado *Document Type Definition* (DTD) e o segundo *XML Schema Definition* (XSD), também conhecido simplesmente como *XML Schema*. A proposta do XSD é mais recente que a do DTD, o XSD foi aprovado como padrão pelo W3C em 16 de março de 2001, e foi desenvolvido justamente com o intuito de suprir as necessidades e limitações que foram identificadas no padrão DTD. Devido a tal fato, o padrão mais conhecido e recomendado atualmente para a construção de esquemas em XML é o XSD.

Deste modo, durante o processo de análise de um documento XML o processador XML, um módulo de software incorporado aos navegadores de Internet, verifica se o documento possui um conjunto de regras e definições associadas a ele, que conduzam a sua interpretação de modo correto, e também se o documento está de acordo com tais regras. Quando um documento atende a esses requisitos diz-se que é um documento XML válido, sendo possível organizá-lo segundo uma estrutura de árvore e representá-lo via XML no ambiente Web. A figura 15, a seguir, ilustra tal processo.

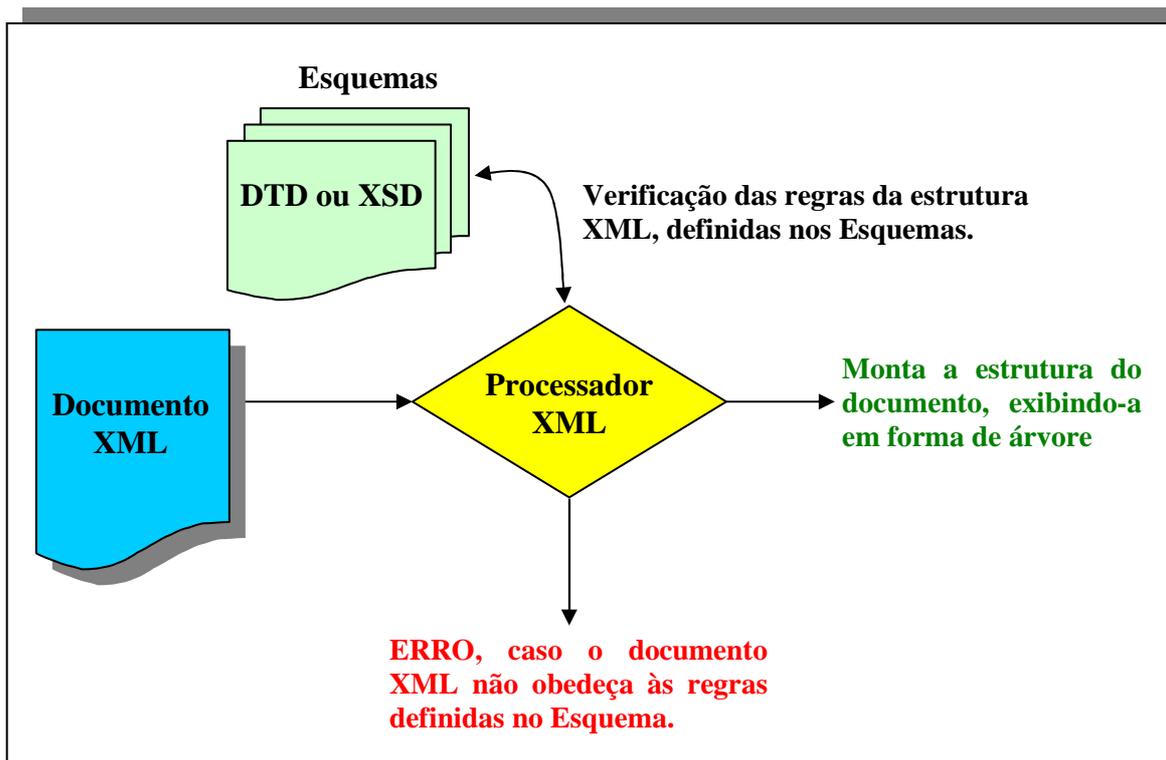


FIGURA 15 – Processo de Análise de um Documento XML
Adaptado de: Tesch Jr., 2002, p.4.

Segundo Almeida (2002, p.11), um documento XML pode ser considerado como “uma representação textual de dados”. Nesse contexto, o componente básico de um código XML é denominado “elemento”, que pode ser entendido como uma entidade utilizada para representar tanto a estrutura quanto os dados contidos em um documento.

Um “elemento” em XML possui uma finalidade diferente de uma marcação, *tag*, utilizada em HTML. Para exemplificar tal afirmação pode-se analisar a marcação em HTML, a qual indica apenas que um texto deve ser apresentado em negrito, em XML, porém, um “elemento” sempre armazena o conteúdo de uma determinada entidade representada no contexto do documento.

Em um código XML um “elemento” é sempre delimitado por duas marcações, de modo que expressões como <estudante> e </estudante> são denominadas “marcação inicial” e “marcação final”, respectivamente. A estrutura textual localizada entre tais marcações é denominada “conteúdo”, assim, um “elemento” compreende o conjunto de todos os dados existentes entre as marcações inicial e final, inclusive as mesmas, conforme ilustrado na figura 16, a seguir.

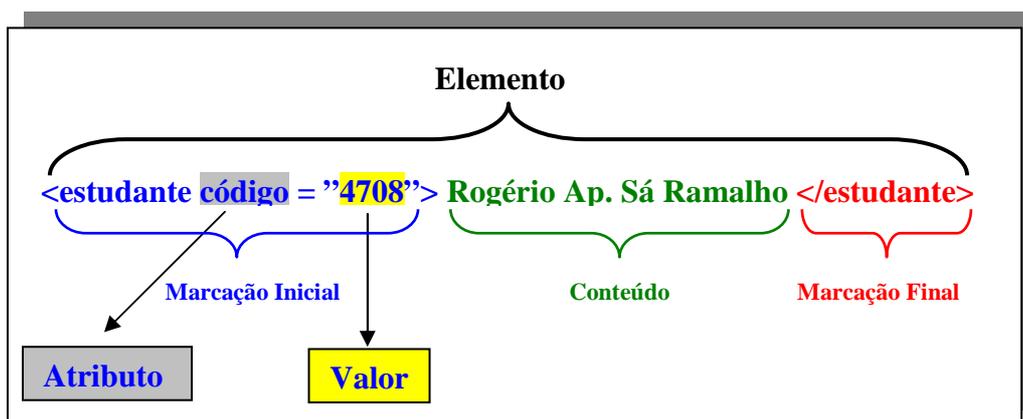


FIGURA 16 – Estrutura de um “elemento simples” em XML

A figura 16 apresenta também outro termo empregado na nomenclatura da linguagem XML denominado como “atributo”, o qual é descrito dentro da marcação inicial de um “elemento”. Os “atributos” são fontes de informação adicionais sobre um “elemento” e sempre possuem um “valor” que deve ser delimitado por aspas. Segundo Almeida (2002, p.12), “o termo ‘atributo’ é utilizado no contexto do XML para especificar propriedades ou características do elemento”.

Os “elementos” podem ser declarados em XML como sendo do tipo simples ou complexo, a diferença básica entre estes é que os “elementos simples” não contêm outros “elementos”, enquanto que os “elementos complexos” contêm. Nessa perspectiva, é utilizado o termo “subelemento” para descrever a relação entre um “elemento” e os “subelementos” que o compõem.

A figura 17, a seguir, representa a codificação de um documento XML válido, indicando a distinção entre um “elemento simples” e um “elemento complexo” e facilitando uma melhor compreensão da estrutura de um documento XML.

```

<?xml version="1.0"?>
<estudante codigo="4708"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="esquema_estudante.xsd"
>
.....:→ Elemento Simples
<nome>Rogério Ap. Sá Ramalho</nome> "nome".
<email>ramalho.rogerio@uol.com.br</email>
<telefone>14 3496-1234</telefone>
.....:
<endereço>
.....:
  <rua>Rua Mandaguaris nº 1498</rua>
  <cep>17600-000</cep>
  <cidade>Tupã</cidade>
  <estado>SP</estado>
  </endereço>
.....:
<grupo_de_pesquisa>
.....:
  <grupo>Novas Tecnologias em Informação</grupo>
  <universidade>Unesp</universidade>
  <campus>Marília</campus>
  </grupo_de_pesquisa>
  <disciplina>
.....:
    <nome>Bibliotecas Digitais</nome>
    <professor>Silvana Vidotti</professor>
    <creditos>6</creditos>
  </disciplina>
  <disciplina>
.....:
    <nome>Catalogação</nome>
    <professor>Plácida L. V. A. Costa Santos</professor>
    <creditos>6</creditos>
  </disciplina>
</estudante>

```

Elemento Complexo "endereço", composto pelos "subelementos": rua, cep, cidade e estado.

FIGURA 17 – Codificação de um documento em XML

Como é possível observar na figura 17, o “elemento” delimitado pelas marcações <estudante> e </estudante> abarca todo o conteúdo do documento, sendo o primeiro “elemento” definido na estrutura XML e por isso é denominado “elemento raiz”. Assim, pode-se considerar todos os demais “elementos” de um documento XML como “subelementos” do “elemento raiz”.

A seguir, é apresentado na figura 18 um possível esquema XSD para a validação do documento XML descrito anteriormente na figura 17.

```

<?xml version='1.0'?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="estudante" type="TipoEstudante"/>
  <xsd:complexType name="TipoEstudante">
    <xsd:sequence>
      <xsd:element name="nome" type="xsd:string"/>
      <xsd:element name="email" type="xsd:string"/>
      <xsd:element name="telefone" type="TipoTelefone"/>
      <xsd:element name="endereço" type="TipoEndereço"/>
      <xsd:element name="grupo_de_pesquisa" type="TipoGrupodepesquisa"
        minOccurs="0" maxOccurs="unbounded"/>
      <xsd:element name="disciplina" type="TipoDisciplina"
        minOccurs="0" maxOccurs="unbounded"/>
    </xsd:sequence>
    <xsd:attribute name="codigo" type="xsd:string" use="required"/>
  </xsd:complexType>
  <xsd:simpleType name="TipoTelefone">
    <xsd:restriction base="xsd:string">
      <xsd:pattern value="\d{2} \d{4}-\d{4}"/>
    </xsd:restriction>
  </xsd:simpleType>
  <xsd:complexType name="TipoEndereço">
    <xsd:sequence>
      <xsd:element name="rua" type="xsd:string"/>
      <xsd:element name="cep" type="TipoCep"/>
      <xsd:element name="cidade" type="xsd:string"/>
      <xsd:element name="estado" type="TipoEstado"/>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:simpleType name="TipoCep">
    <xsd:restriction base="xsd:string">
      <xsd:pattern value="\d{5}-\d{3}"/>
    </xsd:restriction>
  </xsd:simpleType>
  <xsd:simpleType name="TipoEstado">
    <xsd:restriction base="xsd:string">
      <xsd:pattern value="[A-Z]{2}"/>
    </xsd:restriction>
  </xsd:simpleType>
  <xsd:complexType name="TipoGrupodepesquisa">
    <xsd:sequence>
      <xsd:element name="grupo" type="xsd:string"/>
      <xsd:element name="universidade" type="xsd:string"/>
      <xsd:element name="campus" type="xsd:string"/>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="TipoDisciplina">
    <xsd:sequence>
      <xsd:element name="nome" type="xsd:string"/>
      <xsd:element name="professor" type="xsd:string" maxOccurs="2"/>
      <xsd:element name="creditos" type="xsd:decimal"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:schema>

```

FIGURA 18 – Codificação XSD do documento apresentado na figura 17

Nesse contexto, também é possível representar graficamente as estruturas dos esquemas. Assim, considerando-se o esquema descrito anteriormente pode-se obter um diagrama conforme apresentado na figura 19, a seguir.

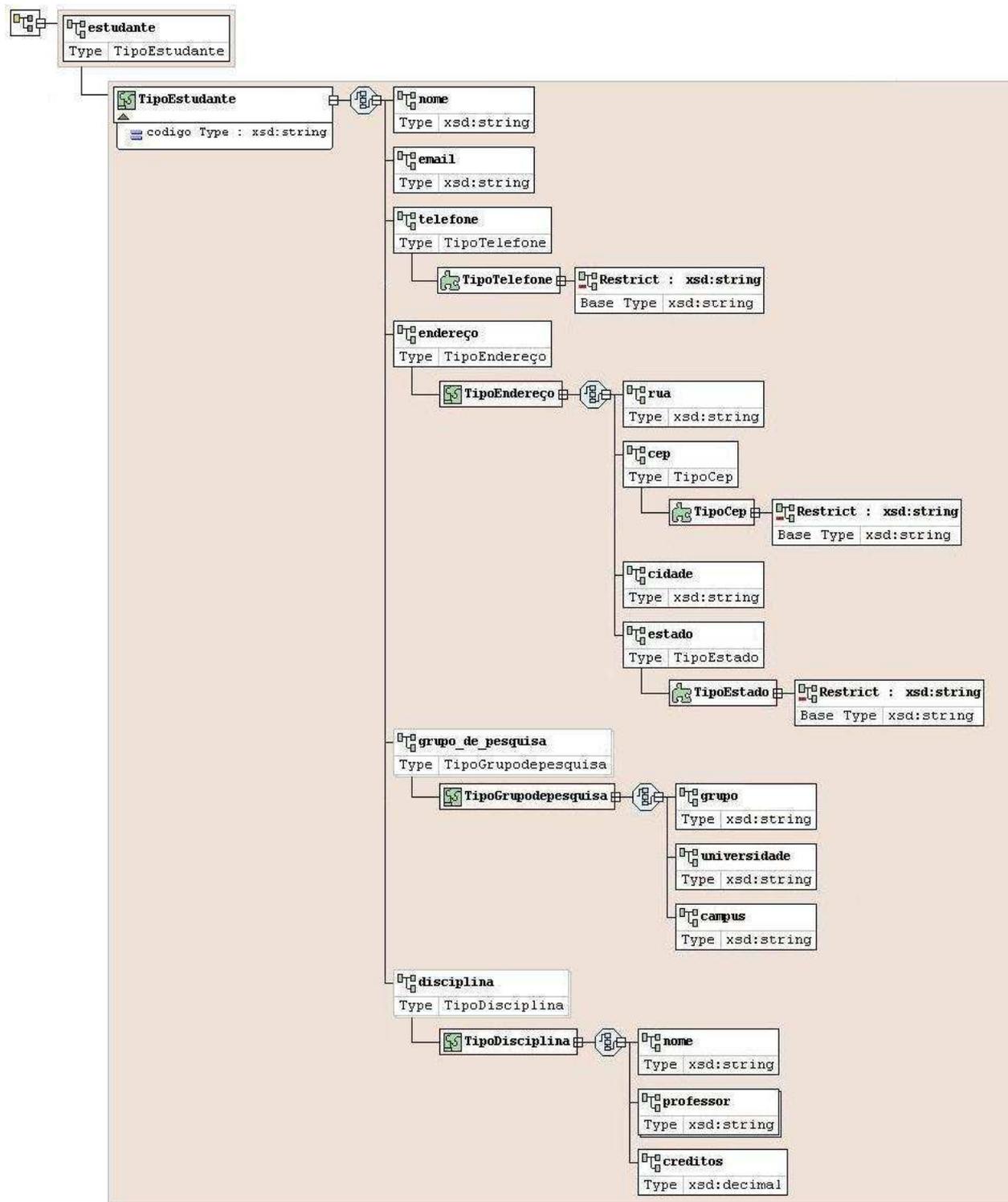


FIGURA 19 – Diagrama do Esquema apresentado na figura 17

Segundo Daum e Merten (2002), a liberdade que a linguagem XML propicia ao desenvolvedor, permitindo a criação de suas próprias marcações e de “elementos” de acordo com suas necessidades e interesses, constitui uma das principais vantagens da XML, mas também pode gerar alguns problemas. Isso porque os “elementos” podem ser definidos livremente pelos próprios desenvolvedores, e é provável que uma mesma marcação seja utilizada em diferentes contextos, com o intuito de atender finalidades distintas.

Deste modo, quando existe a necessidade de mesclar documentos desenvolvidos em diferentes contextos, ou mesmo de realizar interações entre estes, é possível haver colisões de “elementos”, gerando redundâncias e inconsistência de dados.

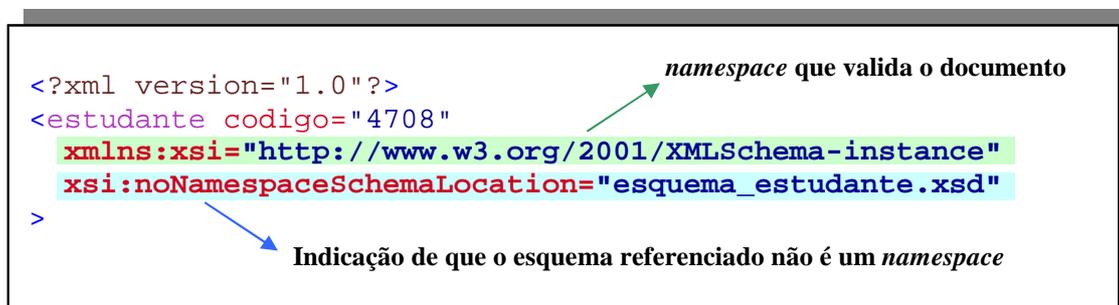
Para ilustrar tal fato, pode-se verificar a codificação do documento XML apresentada na figura 6 e seu respectivo esquema descrito na figura 7, onde o “elemento” delimitado pelas marcações <disciplina> e </disciplina> representa os dados referentes às disciplinas cursadas por um estudante, de modo que é bem provável que as mesmas marcações, <disciplina> e </disciplina>, também possam ser utilizadas em outros documentos XML para representar “elementos” distintos, como por exemplo para definir o comportamento de um estudante do ensino básico.

Assim, com o intuito de evitar tais problemas criou-se o conceito de *namespace* em XML. Segundo documento do W3C, publicado por Bray et al. (2004), um *namespace* é uma coleção de nomes que são utilizados em um documento XML para validar elementos e atributos, e que são identificados por um URI.

Segundo Daum e Merten (2002), os *namespaces* devem ser identificados a partir de um nome de domínio registrado, pertencente ao autor, e uma expressão de caminho arbitrária. Assim, deve-se ressaltar a importância da utilização de nomes de domínios registrados e não nomes de fantasia, pois somente os nomes de domínios registrados são globalmente exclusivos. De modo que, o nome do caminho é utilizado para diferenciar entre vários *namespaces* definidos em um domínio, possibilitando que qualquer um possa criar quantos *namespaces* desejar, bastando para isso possuir um nome de domínio registrado.

Segundo Tesch Jr. (2002), o DTD, padrão mais antigo para desenvolvimento de esquemas XML, não suporta a utilização de *namespaces*, sendo assim necessária a utilização do padrão XSD, ou XML *Schema*, o qual oferece suporte completo para tal tecnologia. Nesse contexto, cabe ressaltar que apesar dos *namespaces* serem identificados por URIs, tal fato se deve apenas para disponibilizar um nome único para o *namespace*, associando-o a um determinado domínio e conseqüentemente a um contexto específico.

Em um documento XML é possível tanto declarar um *namespace* padrão para todo o documento como também definir múltiplos *namespaces*, de modo que as instâncias de um documento XML possam utilizar “elementos” e atributos de vários *namespaces*, definidos nos esquemas correspondentes. A figura 20, a seguir, mostra um trecho da codificação do documento apresentada na figura 17, destacando a declaração de um *namespace* em sua 3ª linha e também ressalta a declaração “*xsi:noNamespaceSchemaLocation*”, a qual possui a finalidade de apenas indicar a localização do esquema do documento, informando que o mesmo não define um *namespace*.



```
<?xml version="1.0"?>
<estudante codigo="4708"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="esquema_estudante.xsd"
>
```

FIGURA 20 – Trecho da Codificação de um documento em XML

Assim, conforme afirmam Daum e Merten (2002), a XML apresenta-se como uma tecnologia básica capaz de fornecer subsídios para o desenvolvimento de várias linguagens de usos especiais, para os mais diversificados fins, bastando para isso a definição de um conjunto de “elementos” e regras de sintaxe. De modo que diversos padrões já foram estabelecidos baseando-se em XML, entre eles se pode citar: *Dublin Core*, *PICS – Platform for Internet Content Selection* e o *Resource Description Framework (RDF)*.

Nessa perspectiva, a partir da análise da literatura apresentada, pode-se verificar que a linguagem computacional XML possibilita novas perspectivas de estruturação de conteúdos, constituindo uma evolução importante com relação à forma de representação de recursos informacionais no ambiente Web.

No entanto, conforme afirmam diversos autores, (Decker et al., 2000; Daum e Merten, 2002), a principal limitação da XML deve-se ao fato da mesma concentrar-se na análise sintática dos conteúdos, não fornecendo meios padronizados para a definição de estruturas semânticas que descrevam formalmente os significados subjacentes aos conteúdos representados, pois apesar das especificações XML possibilitarem a representação dos dados a partir de estruturas, nada informam a respeito da semântica de tais estruturas, o que torna complexa a tarefa de reconstrução dos significados inerentes a um determinado documento.

3.4.2 RDF

Segundo definições apresentadas nos documentos do W3C, (Brickley, 2004; Klyne et al., 2004; Manola e Miller, 2004), RDF é uma linguagem de propósito geral para representação de informações contidas nos recursos Web. Nesse contexto, Daum e Merten (2002, p. 124) afirmam que “[...] RDF pode ser vista como uma tecnologia de capacitação para modelagem semântica, como uma ‘linguagem montadora’ genérica, sobre a qual podem ser criadas linguagens específicas do domínio e da tarefa”.

O modelo e a especificação da sintaxe RDF foram propostos em fevereiro de 1999 pelo W3C, com o intuito de possibilitar uma maior interoperabilidade no ambiente Web, oferecendo um padrão aberto para a descrição de recursos. Deste modo, o padrão RDF possibilita uma ampla gama de aplicações, permitindo que sejam feitas declarações a respeito de praticamente qualquer tipo de objeto, desde que este possa ser identificado a partir de um URI.

O RDF permite descrever declarações a respeito de recursos, não exigindo modificações nos mesmos, de modo que uma declaração RDF é uma entidade separada do recurso ao qual ela se refere, podendo inclusive constituir outro recurso. Nesta perspectiva, pode haver muitas declarações RDF distribuídas pela Web referindo-se ao mesmo recurso, descrevendo diferentes propriedades, ou mesmo contextualizando-o a partir de domínios diferentes.

Segundo a recomendação do W3C, publicada por Klyne et al.(2004), os princípios fundamentais do padrão RDF baseiam-se na tripla “*subject, predicate e object*”, “sujeito, predicado e objeto”. De modo que “Sujeito” é o recurso ao qual uma sentença está se referindo, “Predicado” descreve uma característica, propriedade, ou relacionamento utilizado para descrever algo sobre este recurso e “Objeto” é o valor de uma determinada característica do recurso referenciado, podendo inclusive ser outro recurso.

Assim, em RDF toda sentença é formada por um Sujeito que está relacionado a um Predicado que possui um valor indicado a partir de um Objeto, sendo esta sentença denominada como Declaração.

Nesta perspectiva, pode-se apresentar graficamente uma declaração RDF a partir de um grafo rotulado direcionado, também denominado como “diagramas de nós e arcos”, conforme figura 21. Tais grafos representam os recursos como elipses, os valores literais das propriedades como retângulos e os predicados utilizando arcos direcionados do recurso (sujeito) para o valor (objeto).

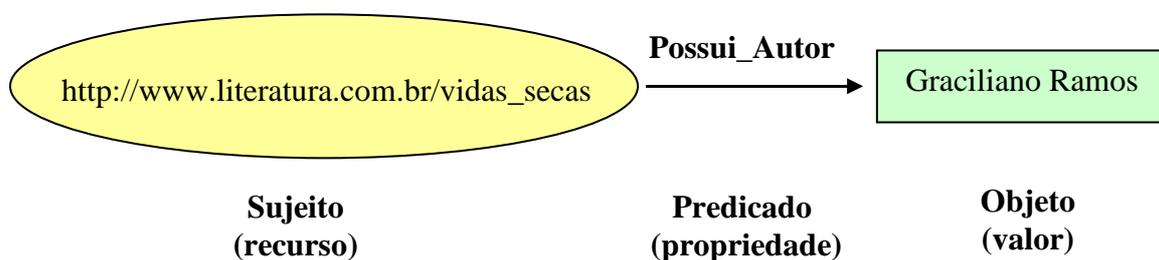


FIGURA 21 – Grafo de uma declaração RDF

Nesse contexto, a figura 21 apresenta o sujeito referenciado a partir do URI “http://www.literatura.com.br/vidas_secas” correspondente a um recurso Web, o predicado “Possui_Autor” indicando uma propriedade deste mesmo recurso e o objeto “Graciliano Ramos” que consiste em uma sentença literal que representa o valor específico da propriedade.

Em RDF os recursos e os predicados devem ser identificados utilizando-se URIs, possibilitando uma maneira global e única de nomear itens. Deste modo, considerando-se o exemplo anterior, o predicado “Possui_Autor” também deve ser descrito de maneira formal em algum local que possa ser referenciado por um URI, assim como o sujeito,

para que os computadores possam analisá-lo e “compreender” o seu significado. Quanto aos objetos que representam os valores dos predicados, estes podem constituir apenas sentenças literais, como o objeto “Graciliano Ramos” apresentado na figura 21, porém quando estes referenciam outro recurso também devem ser identificados a partir de um URI.

Assim, pode-se verificar que em uma declaração RDF tanto o sujeito quanto o valor das propriedades do sujeito podem ser considerados como objetos, ligados por um certo tipo de relacionamento. Deste modo, outra maneira que pode ser utilizada para descrever uma declaração RDF, considerando-se a tripla Sujeito “s”, Predicado “P” e Objeto “o”, é utilizando-se um enunciado Lógico no qual o predicado binário “P” relaciona o objeto “s” ao objeto “o”, originando a sentença lógica “P(s,o)”.

Nesta perspectiva, é possível estender o exemplo anterior representando o objeto “Graciliano Ramos”, a partir do URI do site oficial deste autor, possibilitando inclusive considerá-lo como sujeito de uma outra declaração RDF, relacionando-o com outros predicados e valores, originando uma declaração RDF composta, conforme apresentado na figura 22. Assim, verifica-se que o padrão RDF permite o desenvolvimento de declarações recursivas, declarações sobre declarações.

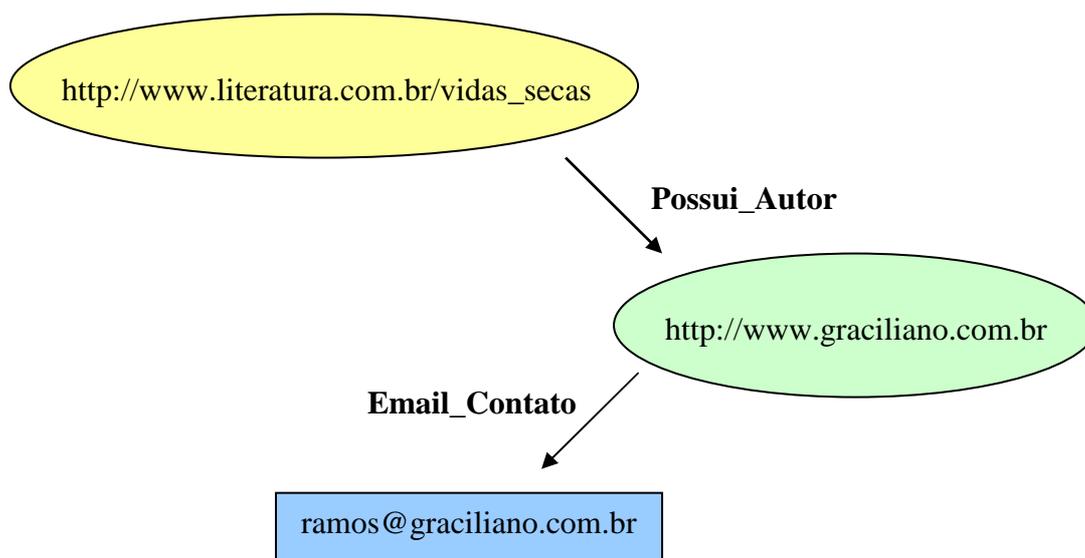


FIGURA 22 – Grafo de uma declaração RDF Composta

Nesse contexto, verifica-se que os grafos apresentam-se como excelentes instrumentos para transmitir informações entre seres humanos, porém no âmbito da Web Semântica torna-se necessário representar as informações em um formato que possa ser processado por máquinas. Deste modo, é possível representar as declarações RDF por meio da linguagem XML.

Uma declaração RDF pode ser representada em linguagem XML a partir de um elemento com a etiqueta `rdf:RDF`, onde o conteúdo desse elemento é identificado utilizando-se a etiqueta `rdf:Description`. Nesta perspectiva cada descrição refere-se a um recurso, que pode ser identificado utilizando-se atributos dos seguintes tipos:

- **about** - Faz a referência a um recurso existente.
- **ID** – Cria um novo recurso.
- **“Sem nome”** – Cria um atributo anônimo

Deste modo, é possível descrever o exemplo apresentado na figura 22, página anterior, utilizando-se a sintaxe XML, da seguinte forma:

```
<rdf:RDF>
  <rdf:Description about="http://www.literatura.com.br/vidas_secass">
    <a:Possui_Autor>
      <rdf:Description about="http://www.graciliano.com.br">
        <e:Email_Contato>
          ramos@graciliano.com.br
        </e:E_mail_Contato>
      </rdf:Description>
    </a:Possui_Autor>
  </rdf:Description>
</rdf:RDF>
```

FIGURA 23 – Representação de uma declaração RDF utilizando a Sintaxe XML

Assim, verifica-se que o padrão RDF fornece mecanismos que possibilitam descrever recursos e seus relacionamentos de modo independente de qualquer implementação e sintaxe, a partir de uma semântica simplificada que pode ser representada utilizando-se a linguagem XML, porém o RDF possibilita apenas a descrição de recursos individuais, ou coleções de recursos individuais, limitando-se ao nível de instância.

Nesta perspectiva, com o intuito de complementar o padrão RDF e fornecer a este um maior nível de abstração, possibilitando a definição de conceitos primitivos e tipos de objetos, foi desenvolvido o RDF-Schema (RDFS), uma linguagem de descrição de vocabulários que objetiva descrever propriedades e classes para os recursos RDF. Conforme afirmam Daum e Merten (2002, p. 139):

[...] RDFS nos permite definir uma taxonomia de recursos em termos de classes, superclasses e subclasses de recursos. Para cada uma dessas classes, podemos definir quais os tipos de declarações podem ser feitas sobre as instâncias dessas classes, ou seja, quais tipos de propriedades podem ser associadas.

O RDFS possibilita que as comunidades possam desenvolver vocabulários particulares, de maneira independente, de acordo com seus domínios de aplicação. De modo que o RDFS não fornece classes e propriedades propriamente ditas, mas sim uma estrutura a partir da qual é possível descrevê-las.

Assim, o RDFS possibilita definir uma terminologia para representar conceitos como recurso, classe (tipos de recursos), subclasse e propriedade (atributos de classes), que podem ser utilizados para expressar taxonomias de conceitos e suas relações. Segundo Campos et al. (2005, p. 67):

[...] o RDFS vem apoiar não só a utilização de vocabulários diversos para diferentes domínios, como também dar um sentido consensual a alguns elementos básicos que podem ser explorados pelos agentes de software que estejam de acordo com esta semântica, para interpretar de forma inequívoca as afirmativas feitas sobre os recursos por eles manipulados.

Dentro deste contexto, com o intuito de exemplificar a utilização do RDFS, a figura 24, apresenta um trecho de código baseado no padrão RDFS, o qual indica que “professor” e “aluno” são SubClasses da classe “pessoa”.

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

<rdf:Class rdf:ID="pessoa" />

<rdfs:Class rdf:ID="professor">
  <rdfs:subClassOf rdf:resource="#pessoa"/>
</rdfs:Class>

<rdfs:Class rdf:ID="aluno">
  <rdfs:subClassOf rdf:resource="#pessoa"/>
</rdfs:Class>

</rdf:RDF>
```

FIGURA 24 – Trecho de código baseado em RDFS

Deste modo, verifica-se que o padrão RDFS fornece um conjunto de primitivas que possibilitam a modelagem de ontologias simples, porém deve-se ressaltar que o RDFS não foi concebido com o propósito de ser uma linguagem para construção de ontologias e, desta forma, apresenta limitações, pois os conectivos lógicos de negação, disjunção e conjunção não existem em RDFS, limitando a sua expressividade, assim para atingir melhores níveis de expressividade é necessária a utilização de linguagens criadas especificamente para o desenvolvimento de ontologias. Nesse contexto, a linguagem mais utilizada atualmente e recomendada para o desenvolvimento de ontologias é denominada OWL.

3.4.3 OWL

A OWL é uma linguagem computacional utilizada para o desenvolvimento de ontologias, a qual se originou a partir da junção das especificações das linguagens DAML e OIL, sendo denominada inicialmente como DAML+OIL. Assim, desde fevereiro de 2004 a OWL é recomendada pelo W3C como linguagem padrão para o desenvolvimento de ontologias.

Segundo recomendação do W3C, publicada por McGuinness e Harmelen (2004), a OWL é indicada para ser utilizada em situações onde as informações contidas em documentos necessitem ser processadas de forma automatizada, e não apenas apresentadas para seres humanos, pois a OWL permite descrever formalmente o significado dos termos utilizados em um documento e seus respectivos relacionamentos, possibilitando representações mais abrangentes que as linguagens RDF e RDF Schema, favorecendo uma maior interoperabilidade.

Nesse contexto, pode-se considerar OWL como um padrão que na realidade abarca três tipos de linguagens com diferentes níveis de expressividade, conforme apresentado a seguir:

- **OWL Lite** – Projetada para permitir uma fácil implementação, fornecendo aos usuários um subconjunto funcional baseado em classificações hierárquicas e restrições simples, possibilitando inclusive a migração de sistemas baseados em tesauros e taxonomias para o formato de ontologias.
- **OWL DL** – Projetada para suportar implementações baseadas em Lógica Descritiva, fornecendo um subconjunto que possua propriedades desejáveis em sistemas que necessitem ontologias com um maior nível de detalhamento e restrições.
- **OWL Full** – Projetada para possibilitar o máximo de expressividade enquanto mantém completude computacional, de modo que, diferentemente da OWL DL, pode violar restrições da Lógica Descritiva com o objetivo de ser compatível com o maior número possível de bancos de dados e sistemas de representação do conhecimento.

Cabe ressaltar que a OWL Lite possibilita apenas a definição de cardinalidades binárias, contendo os valores “0” ou “1”. Outro esclarecimento importante é que utilizando a OWL DL é possível considerar as mesmas construções que a OWL Full oferece, porém a OWL DL não permite que uma classe possa ser considerada um indivíduo ou uma propriedade e também não permite que uma propriedade possa ser considerada como um indivíduo ou uma classe. Já a OWL Full permite essas construções, pois considera uma classe como um conjunto de indivíduos, assim como um próprio indivíduo, de forma simultânea.

Nesta perspectiva, conforme apresentado no manual de OWL publicado por Smith et al. (2004) e disponibilizado na página do W3C, pode-se dividir a estrutura de um documento OWL baseando-se nos seguintes elementos básicos:

- *Namespaces*
- Cabeçalhos
- Classes
- Indivíduos
- Propriedades
- Restrições

Os *namespaces*, conforme apresentado anteriormente, são definidos a partir de declarações XML, permitindo identificar sem ambigüidades a localização dos vocabulários correspondentes ao conjunto de conceitos utilizados na ontologia. Os *namespaces* são declarados entre etiquetas do tipo *rdf:RDF*, conforme apresentado na figura 25, fornecida no manual de OWL da página do W3C.

```
1 <rdf:RDF
2   xmlns      = "http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#"
3   xmlns:vin  = "http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#"
4   xml:base   = "http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#"
5   xmlns:food = "http://www.w3.org/TR/2004/REC-owl-guide-20040210/food#"
6   xmlns:owl  = "http://www.w3.org/2002/07/owl#"
7   xmlns:rdf  = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
8   xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
9   xmlns:xsd  = "http://www.w3.org/2001/XMLSchema#">
```

FIGURA 25 – Trecho de código OWL indicando os *Namespaces* utilizados
Fonte: Smith et al., 2004.

Neste exemplo, a linha 1 contém apenas a etiqueta *rdf:RDF*, indicando que as próximas declarações apontam os *namespace* da ontologia. A linha 2 indica o *namespace* padrão da ontologia, de modo que qualquer termo utilizado dentro desta ontologia que não contenha nenhum prefixo se refere à própria ontologia; na linha 3 é

apresentado um nome, “vin⁷” para a ontologia. A linha 4 identifica a URI base para esta ontologia e a linha 5 referencia uma outra ontologia, a de comida (*food*), da qual alguns conceitos foram incorporados. As demais linhas: 6, 7, 8 e 9 indicam a localização dos vocabulários suportados pelas linguagens: OWL, RDF, RDFS e XSD, respectivamente, pois é necessário indicar não apenas o vocabulário convencional OWL, como também, das primitivas definidas nas camadas inferiores, tipos de dados nativos de RDF, RDFS e do XML Schema.

Após a definição dos *namespaces* é comum incluir uma coleção de sentenças comumente denominadas como cabeçalhos, utilizadas para registrar comentários, controlar a versão da ontologia e a inclusão de conceitos e propriedades de outras ontologias. Tais informações são agrupadas sob a etiqueta *owl:Ontology*, conforme apresentado na figura 26.

```
<owl:Ontology rdf:about="">
  <rdfs:comment>
    Exemplo de um comentário em uma ontologia escrita em OWL
  </rdfs:comment>

  <owl:priorVersion
    rdf:resource="http://www.w3.org/TR/2003/PR-owl-guide-20031215/wine"/>
  <owl:imports
    rdf:resource="http://www.w3.org/TR/2004/REC-owl-guide-20040210/food"/>
  <rdfs:label>
    Ontologia de Vinhos
  </rdfs:label>
</owl:Ontology>
```

FIGURA 26 – Trecho de código OWL referente ao cabeçalho de uma ontologia

⁷ O trecho de código apresentado na figura 26 compõe parte de um exemplo de ontologia sobre tipos de vinhos, muito conhecida pela comunidade de pesquisadores de Web Semântica, a qual pode ser encontrada na íntegra em <http://www.daml.org/ontologies/76>

O elemento *owl:Ontology* normalmente é utilizado para descrever informações úteis sobre a ontologia, como a versão e comentários adicionais que auxiliem na sua identificação. O atributo *rdf:about* fornece o nome ou referência para a ontologia, de modo que quando nenhum valor é especificado a este atributo, caso padrão, o nome da ontologia é considerado como sendo a URI do elemento *owl:ontology*, que normalmente se refere ao próprio documento que contém a ontologia. O atributo *rdfs:comment* fornece a possibilidade de incluir comentários sobre a ontologia, *owl:priorVersion* fornece indicativos para sistemas de controle de versão e o atributo *owl:imports* fornece um mecanismo de inclusão, permitindo apenas um único argumento, indicado a partir do atributo *rdf:resource*.

Nesse contexto, um dos principais elementos que constituem uma ontologia são as classes. Segundo Smith et al. (2004), uma classe representa um conjunto ou coleção de indivíduos (objetos, pessoas, coisas) que possuem um grupo de características comuns, as quais os distinguem dos demais.

Em OWL classes são utilizadas para descrever conceitos básicos de um domínio, que vão servir como raízes de várias taxonomias. Na classificação de Aristóteles foi introduzida a idéia de *genus supremo*, quando se definiu a classe substância, da qual todas as outras classes originam-se. Assim, de forma análoga, em OWL existe a conceituação convencionalmente abstrata de *Thing*, “Coisa”, de modo que todos os elementos em uma ontologia OWL pertencem à classe genérica *owl:Thing*, possibilitando que exista sempre uma única raiz comum a qualquer taxonomia.

Nesta perspectiva, para construir taxonomias é necessário definir uma hierarquia de classes, a partir de relacionamentos de generalização (tipo - de), definindo assim subclasses de indivíduos. Em OWL para a definição de classes utiliza-se a marcação *owl:Class*, e a marcação *rdfs:subClassOf* é utilizada para a definição de subclasses, conforme apresentado na figura 27, a seguir.

```

<owl:Class rdf:ID="Pessoa"/>

<owl:Class rdf:ID="Professor">
  <rdfs:label>Professor</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Pessoa" />
</owl:Class>

```

FIGURA 27 – Declaração de Classes e Subclasses em OWL

Nesse contexto, deve-se ressaltar que o trecho de código, apresentado na figura 27, indica apenas que as classes Pessoa e Professor existem, sendo esta última uma subclasse da classe Pessoa, pois formalmente nada mais está descrito sobre estas classes, lembrando-se que o fato de utilizar substantivos comuns em português como “Professor”, por exemplo, nada significa para a ontologia, de modo que mesmo que seja atribuída a sentença “XYK12” para identificar esta classe, a ontologia permaneceria com o mesmo nível de expressividade computacional, pois para atribuir semântica a uma classe, de modo que as máquinas possam interpretar, é necessário explicitar todas as suas características e relacioná-la com outras classes e propriedades.

Assim, a partir da definição de classes pode-se definir indivíduos, os quais constituem objetos do mundo, que pertencem a determinadas classes, e são relacionados com outros indivíduos a partir de propriedades. Em OWL um indivíduo é adicionado declarando-o como membro de uma determinada classe. Para exemplificar tal afirmação pode-se considerar a seguinte sentença: *<Professor rdf:ID= "Rogério Ramalho" />*

As propriedades servem para descrever fatos em geral, de modo que podem referir-se a todos os membros pertencentes a uma determinada classe, descrevendo afirmações do tipo: *“todos os professores lecionam aulas”*, como também se referir a um indivíduo específico de uma classe, como por exemplo, *“Rogério Ramalho é aluno da UNESP”*.

Em OWL as propriedades são relacionamentos binários e podem ser classificadas a partir dos seguintes tipos distintos:

- **Propriedades do tipo Objeto** – Definem um relacionamento entre duas classes de objetos, conforme exemplo a seguir.

```
<owl:ObjectProperty rdf:ID="é_aluno">
  rdfs:domain rdf:resource="#Rogério Ramalho"/>
  rdfs:domain rdf:resource="#Unesp"/>
</owl:ObjectProperty>
```

FIGURA 28 – Declaração de uma propriedade do tipo Objeto

- **Propriedades do tipo *Datatype*** – Definem um relacionamento entre instâncias de classes e literais expressos em RDF e *datatypes* do XML Schema. As referências dos *datatypes* devem ser apresentadas utilizando-se o URI que contém os *datatypes* recomendados pelo W3C para utilização em OWL⁸, conforme apresentado na tabela 2, a seguir.

TABELA 2 – *Datatypes* definidos pelo W3C para utilização em OWL

xsd:string	xsd:normalizedString	xsd:boolean	
xsd:decimal	xsd:float	xsd:double	
xsd:integer	xsd:nonNegativeInteger	xsd:positiveInteger	
xsd:nonPositiveInteger	xsd:negativeInteger		
xsd:long	xsd:int	xsd:short	xsd:byte
xsd:unsignedLong	xsd:unsignedInt	xsd:unsignedShort	xsd:unsignedByte
xsd:hexBinary	xsd:base64Binary		
xsd:dateTime	xsd:time	xsd:date	xsd:gYearMonth
xsd:gYear	xsd:gMonthDay	xsd:gDay	xsd:gMonth
xsd:anyURI	xsd:token	xsd:language	
xsd:NMTOKEN	xsd:Name	xsd:NCName	

Fonte: Smith et al., 2004.

⁸ A URI que contém os *datatypes* recomendados pelo W3C para utilização em OWL é: <http://www.w3.org/2001/XMLSchema>

Assim, para exemplificar a declaração de uma propriedade do tipo *datatype* pode-se verificar a figura 29, que relaciona todos os indivíduos da classe pessoa com um ano de nascimento.

```
<owl:Class rdf:ID"Pessoa" />

<owl:DatatypeProperty rdf:ID"anoNascimento">
  rdfs:domain rdf:resource="#Pessoa" />
  rdf:range rdf:resource="&xsd:positiveInteger"/>
</owl:DatatypeProperty>
```

FIGURA 29 – Declaração de uma propriedade do tipo *Datatype*

As propriedades do tipo *datatype* também podem ser adicionadas a indivíduos específicos, como por exemplo, afirmando que uma instância chamada João, da classe pessoa, tem o valor 1970 registrado como seu ano de nascimento. Conforme apresentado na figura 30.

```
<owl:Class rdf:ID"Pessoa" />

<Pessoa rdf:ID="João">
  <anoNascimento rdf:datatype="&xsd:positiveInteger">
    1970
  </anoNascimento>
</Pessoa>
```

FIGURA 30 – Declaração de uma propriedade do tipo *Datatype* representando a instância de um objeto

Outros elementos que assumem um importante papel no desenvolvimento de ontologias são as restrições, *restrictions*. Conforme o próprio nome sugere, uma restrição é utilizada para definir limites para indivíduos pertencentes a uma determinada classe. Em OWL as restrições são definidas utilizando-se o elemento *owl:Restriction*, e podem ser divididas basicamente em três tipos:

- **Restrições que utilizam quantificadores** – Restringem que todos (\forall) ou ao menos um (\exists) elemento de uma classe especificada possui uma determinada propriedade e deve ser membro da classe referida por ela, utilizando-se os atributos *owl:allValuesFrom* e *owl:someValuesFrom*, respectivamente.
- **Restrições de cardinalidade** – Restringem a quantidade mínima, máxima ou exata da cardinalidade de uma determinada propriedade, a partir dos atributos *owl:minCardinality*, *owl:maxCardinality* e *owl:cardinality*, respectivamente.
- **Restrições do tipo hasValue(“tem valor de”)** – Permite especificar classes baseadas na existência de valores de propriedades particulares. Utiliza o atributo *owl:hasValue* para informar uma propriedade de todos os indivíduos que forem membros de uma determinada classe.

Nesse contexto, utilizando-se tais definições da linguagem OWL, um número cada vez maior de pesquisadores tem se empenhado em desenvolver ontologias. Contudo, observa-se que atualmente a construção de ontologias ainda é uma tarefa mais artesanal do que científica (Jones et al., 1998), devido principalmente à falta de metodologias unificadas que atendam diferentes abordagens de utilização em domínios distintos.

Assim, verifica-se que devido principalmente à fragmentação do conhecimento, as variáveis necessárias para o desenvolvimento de ontologias são muito diversificadas. Nessa perspectiva, é possível identificar a possibilidade de contribuições da área de Ciência da Informação, a partir de estudos interdisciplinares que permitam otimizar a estruturação e organização das informações contidas em ambiente digitais, favorecendo assim, o desenvolvimento do projeto Web Semântica, conforme será apresentado no próximo capítulo.

4 CONTEXTUALIZANDO A WEB SEMÂNTICA NO ÂMBITO DA CIÊNCIA DA INFORMAÇÃO

Neste capítulo busca-se contextualizar, a partir da análise apresentada nos capítulos anteriores, o projeto Web Semântica de acordo com o embasamento teórico da área de Ciência da Informação, verificando em que medida esta área do conhecimento pode contribuir para a concretização deste projeto e avaliando também os reflexos destas novas abordagens tecnológicas dentro de seu *corpus* teórico.

A Ciência da Informação tem como um de seus objetivos estudar/desenvolver métodos e técnicas que favoreçam a otimização dos processos de armazenamento, organização e recuperação de informações, levando em consideração aspectos científicos e profissionais que abarcam dimensões sociais e humanas que transcendem os aspectos tecnológicos.

Segundo Saracevic (1996), a recuperação de informação pode ser considerada, de certo modo, como a vertente tecnológica da Ciência da Informação. Nesta perspectiva, verifica-se uma tendência de aproximação entre as áreas de Ciência da Informação e Ciência da Computação, ambas empenhadas no estudo e desenvolvimento de meios de organização de conjuntos específicos de informação e instrumentos de representação.

Nesse contexto, apesar de relacionadas, é possível identificar um distanciamento teórico entre tais Ciências. Ferneda (2003, p.1-2) ao comentar sobre as relações existentes entre a Ciência da Informação e a Ciência da Computação, ressalta que: “[...] a informação, objeto de comum interesse de ambas as ciências, é paradoxalmente o que mais as distancia”.

Tal afirmação sustenta-se no fato de que ambas as Ciências utilizam diferentes conceitos ao se referir ao termo informação. Segundo Le Coadic (1996, p.5), para a área de Ciência da Informação, a informação “[...] comporta um elemento de sentido. É um significado transmitido a um ser consciente por meio de uma mensagem inscrita em um suporte espaço-temporal: impresso, sinal-elétrico, onda sonora, etc.”. Em contra partida, para a área de Ciência da Computação o conceito de informação tradicionalmente restringe-se aos conceitos relacionados com a Teoria Matemática da Informação, mais especificamente com os fundamentos estabelecidos por Shannon e Weaver (1949), presentes no artigo intitulado *The Mathematical Theory of Communication*, onde é

apresentada uma teoria matemática para a transmissão de mensagens e troca de sinais, a qual não se preocupa com a semântica dos dados, porém adequada para a construção de sistemas computacionais, onde a informação possa ser quantificada, processada e transmitida por máquinas. Conforme afirma Shannon (1948, p.3): “*the theory of computing machines*”.

A identificação de tal enfoque na área de Ciência da Computação pode ser comprovada a partir de uma análise em sua literatura e dos resultados práticos identificados no seu campo de atuação. No contexto nacional também é possível identificar tal abordagem nas diretrizes curriculares de cursos na área de Computação e Informática disponibilizadas pelo Ministério da Educação (MEC), elaboradas pela Comissão de Especialistas de Ensino em Computação e Informática (CEEInf, 1999), assim como no currículo de referência para cursos de graduação em Computação, proposto pela Sociedade Brasileira de Computação (SBC, 1999, p.14), no qual figura como parte do currículo básico para cursos de computação, o estudo de:

Princípios da teoria da informação: codificação da informação e sua medida, entropia de código. Transmissão da informação e modelagem do sistema de transmissão, maximização do fluxo de informação por um canal. Processamento digital de sinais, análise espectral. Transmissão analógica e digital.

Segundo Capurro (2003), embora a área de Ciência da Informação também tenha sido influenciada em seu campo teórico pela assim chamada “*information theory*” de Shannon e Weaver, a mesma não se limita a esta visão puramente fisicista, abarcando também outros aspectos: contextuais, semânticos, sociais e culturais.

Contextualizando tais definições, verifica-se que a Ciência da Computação tradicionalmente preocupa-se com o desenvolvimento de sistemas computacionais que possibilitem “manusear” informações, sem necessariamente preocupar-se com os aspectos semânticos subjacentes, enquanto que a Ciência da Informação preocupa-se com a natureza das informações, assim como sua comunicação e uso pelos humanos.

Deste modo, utilizando-se como exemplo um processo de recuperação de informação no ambiente Web, pode-se identificar que no contexto da área de Ciência da Informação é evidente a presença de componentes semânticos em tal processo, pois, para

os usuários as informações recuperadas possuem um significado semântico implícito. Porém, levando-se em consideração o enfoque da área de Ciência da Computação, ainda utilizando o mesmo exemplo, observa-se que os tradicionais “motores de busca”, *search engines*, baseiam-se exclusivamente na recuperação de dados, não levando em consideração as semânticas contidas nas páginas da Web, recuperando apenas seqüências de caracteres que satisfaçam determinadas condições de busca.

Assim, verifica-se que um dos principais “*slogans*” do projeto Web Semântica baseia-se na possibilidade de permitir a classificação de recursos informacionais disponíveis no ambiente Web, “rotulando-os” a partir de categorias que possam ser “interpretadas” automaticamente pelos computadores. Analisando tal definição sob o prisma da área de Ciência da Informação pode-se fazer uma analogia à definição apresentada por Lancaster (2004, p. 21):

No campo do armazenamento e recuperação de informação a *classificação* de documentos refere-se à formação de classes de itens com base no conteúdo temático. Tesouros, cabeçalhos de assuntos e esquemas de classificação bibliográfica são essencialmente listas de *rótulos* com os quais se identificam e, por ventura, se organizam essas classes.

Nesse contexto, os recursos informacionais disponíveis no ambiente Web só poderão ser classificados e categorizados a ponto de permitir a realização de inferências automáticas, como esperam os idealizadores e defensores do projeto Web Semântica, se os computadores forem capazes de captar as informações descritivas e temáticas referentes aos conhecimentos representados nas páginas Web. Assim, verifica-se o motivo do projeto Web Semântica ter despertado uma forte demanda de desenvolvimento de modelos/instrumentos de representação do conhecimento no âmbito computacional.

Para Campos (2004), no contexto da área de Ciência da Computação, espera-se que tais instrumentos auxiliem a implementação de estruturas computáveis, que possibilitem aos computadores a realização de tarefas mais sofisticadas de forma automatizada, enquanto que no contexto da área de Ciência da Informação, os modelos de representação do conhecimento são utilizados há muito tempo na elaboração de linguagens documentárias verbais e notacionais visando à recuperação de informação e à organização dos conteúdos informacionais dos documentos.

Nesta perspectiva, observa-se que muitos dos conceitos disseminados a partir do desenvolvimento do projeto Web Semântica como “absoluta novidade” constituem na realidade uma nova “roupagem tecnológica” para métodos e técnicas que já são utilizados há décadas na área de Ciência da Informação.

Porém, é evidente que as novas tecnologias informacionais usufruindo das facilidades e dos benefícios que os ambientes digitais podem proporcionar, apontam perspectivas renovadoras, muitas delas inimagináveis antes do desenvolvimento/popularização do ambiente Web.

Assim, pode-se destacar o desenvolvimento de ontologias como um dos temas relacionados ao projeto Web Semântica que vem despertando um grande interesse de profissionais, nas mais variadas áreas do conhecimento, empenhados no desenvolvimento desta nova e instigante categoria de instrumentos de representação do conhecimento, conforme será apresentado a partir da próxima seção.

4.1 Ontologias como instrumento de representação

Considerando que “representar” pode ser entendido como: “ato de apresentar algo por meio de”, verifica-se que nas últimas décadas as novas tecnologias da informação e comunicação têm impulsionado uma série de pesquisas relacionadas ao estudo e desenvolvimento de instrumentos de representação, focadas principalmente no ambiente digital.

Assim, no bojo de tal desenvolvimento tecnológico surgiu a utilização do termo ontologia, conforme apresentado na seção 3.3 do capítulo anterior, designando uma nova categoria de instrumentos de representação no âmbito computacional.

Conforme afirma Alvarenga (2003), com o uso intensivo das novas tecnologias constata-se na perspectiva da Ciência da Informação uma turbulência principalmente no que tange à representação do conhecimento, assim como nos processos de armazenamento e recuperação de informações, áreas intensamente relacionadas à cognição humana.

Nesse contexto, alguns pesquisadores consideram que a noção de representação de conhecimento pode ser melhor compreendida a partir de definições ligadas aos papéis que pode desempenhar. Davis (1993, tradução nossa), afirma que a representação do conhecimento pode ser definida basicamente pelas cinco funções que ela exerce:

1. Substituição da realidade: um substituto.
2. Um conjunto de compromissos ontológicos.
3. Materialização da concepção de um raciocínio, um conjunto de inferências que ela apóia e recomenda.
4. Um meio de computação eficiente.
5. Um meio de expressão que permite aos humanos trocar conhecimento.

Segundo Campos (2004, p.24), baseando-se em tais definições, no âmbito da gestão de recursos informacionais pode-se considerar que:

Uma representação de conhecimento é um meio de computação pragmaticamente eficiente. Na realidade, esta questão aborda a utilidade prática da representação. Se ela torna coisas possíveis, mas não facilmente computáveis, a representação pode, então, não ser de muita valia para o problema em mãos.

Para Campos (2001), no contexto da área de Ciência da Computação, as ontologias são utilizadas com o intuito de desvendar o significado das coisas do mundo, procurando descrever suas naturezas, referindo-se, na realidade, a artefatos de engenharia formados por um vocabulário específico que é usado para descrever uma certa realidade e um conjunto de afirmações explícitas sobre o significado das palavras do vocabulário, podendo ser consideradas como modelos de representação do conhecimento.

Os modelos de representação do conhecimento são construídos a partir da definição de conceitos e os relacionamentos existentes entre eles, por esse motivo também são conhecidos como modelos conceituais. Os mesmos podem ser elaborados por meio de uma perspectiva indutiva, partindo de conceitos individuais de um determinado domínio para se alcançar os conceitos gerais, ou a partir de uma perspectiva dedutiva, realizando abstrações e pensando primeiramente no domínio e nos conceitos gerais para numa segunda etapa se alcançar os conceitos individuais e suas relações.

Campos (2004) apresenta como exemplo de modelo de representação que segue a perspectiva dedutiva a teoria de classificação facetada do domínio da Ciência da Informação, e considera as teorias de ontologia formal do domínio da Ciência da Computação como mecanismos para construção de modelos conceituais que seguem a perspectiva indutiva.

Nesse contexto, cabe ressaltar que no domínio da área de Ciência da Computação também é possível construir ontologias utilizando-se uma perspectiva dedutiva, sendo que a partir de uma ontologia mais genérica de um domínio é possível desenvolver uma ontologia mais específica e restritiva, dentro de um subconjunto deste mesmo domínio.

No âmbito da área de Ciência da Informação, uma comparação frequentemente encontrada na literatura diz respeito as familiaridades existentes entre tesauros e ontologias. Segundo Moreira, Alvarenga e Oliveira (2004, p. 23): “Uma ontologia como vista pela ciência da computação é um sistema de conceitos, da mesma forma que os tesauros, e como tal pertence ao nível epistemológico e não ontológico”. Em contra partida, Alvarenga (2003) afirma que no ambiente digital o processo de representação extrapola a instância da epistemologia atingindo acervos ontológicos, ou utilizando-se dos próprios objetos digitais para a construção de pontos de acesso condizentes, pois no meio digital a representação não se limita a uma nova informação condensada que substitui um determinado conteúdo, mas apresenta-se, muitas vezes, como parte constituinte do próprio documento, não limitando as representações ao momento de organização ou acesso das informações secundárias, a partir do conhecimento sobre os seres ou objetos, possibilitando constituir excertos dos próprios objetos virtuais ou representações primárias como no caso das imagens digitais.

Poli (1998), no artigo intitulado *Framing ontology*, também apresenta considerações acerca da distinção entre o aspecto ontológico e epistemológico. Segundo o autor, uma ontologia é uma teoria de objetos, a qual abarca todo tipo de objeto, concreto ou abstrato, existente e não existente, dependente ou independente, de modo que quaisquer que sejam os objetos que se esteja considerando, a ontologia é a teoria deles, em contra partida a epistemologia é a teoria dos diferentes tipos de conhecimento e das formas como estes são utilizados. Neste mesmo artigo o autor reconhece que se baseando em perspectivas ontológicas e epistemológicas torna-se difícil a separação entre tais teorias, pois as mesmas indicam caminhos complexos e mutuamente complementares.

Sendo assim, destaca-se a importância de análises comparativas entre as ontologias e os instrumentos de representação de informações, utilizados tradicionalmente no âmbito da área de Ciência da Informação, a partir dos propósitos práticos que norteiam o desenvolvimento de cada um desses instrumentos e dos reflexos de suas utilizações nos processos de organização e recuperação de informações.

No contexto da área de Ciência da Informação pode-se destacar a utilização de Linguagens Documentárias (LDs), construídas para a indexação, armazenamento e recuperação de informações a partir da “tradução” dos conteúdos dos documentos, descritos em linguagem natural, para um vocabulário controlado. (CINTRA et al., 2002)

Segundo definição da organização norte-americana *National Information Standards Organization*, presente no documento que propõe as linhas gerais para a construção, formatação e manutenção de vocabulários controlados monolíngües (ANSI/NISO Z39-19-2005), um vocabulário controlado é uma lista finita de termos que tem seus respectivos significados explicitados com o intuito de evitar redundâncias e ambigüidades, utilizados para representar informações de maneira padronizada.

Ainda segundo o mesmo documento, os vocabulários controlados são estruturados para possibilitar diferentes tipos de relacionamentos entre termos, determinando desde níveis de relacionamentos simples até estruturas mais complexas, conforme exibido na figura 31.

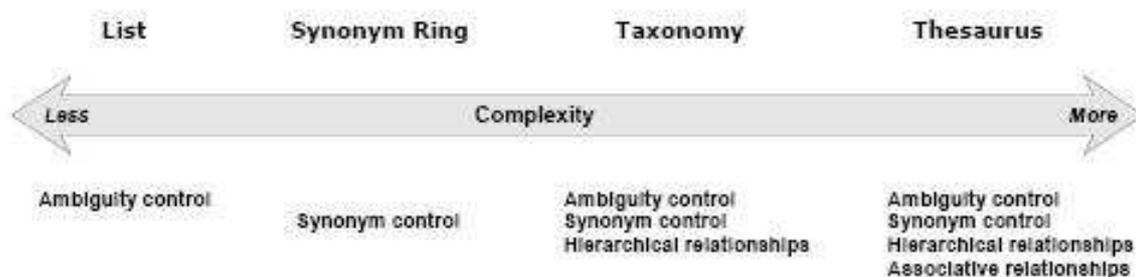


FIGURA 31 – Níveis de complexidade estrutural dos Vocabulários Controlados
Fonte ANSI/NISO Z39-19-2005, p. 17

Assim, de acordo com o enfoque desta pesquisa, pode-se destacar as taxonomias e os tesouros como importantes instrumentos de representação utilizados tradicionalmente no âmbito da área de Ciência da Informação.

Segundo a norma ANSI/NISO Z39-19-2005 (p.9, tradução nossa), uma taxonomia é: “Uma coleção de termos de um vocabulário controlado organizada em uma estrutura hierárquica”. As taxonomias permitem classificar informações em uma estrutura de árvore, por meio de relacionamentos de generalização (“pai-filho”, “tipo-de”), não possibilitando atribuir características ou propriedades aos termos nem expressar outros tipos de relacionamentos.

Quanto aos tesouros a norma ANSI/NISO Z39-19-2005 define como um vocabulário controlado organizado segundo uma ordem conhecida e estruturada com o intuito de disponibilizar claramente os relacionamentos de equivalência, associação, hierárquicos e homônimos existentes entre termos.

Os tesouros também comportam características de taxonomias com um conjunto de relacionamentos semânticos, visando garantir que os conceitos e seus relacionamentos sejam descritos de maneira consistente em um sistema de classificação e recuperação de informações.

Atualmente verifica-se uma forte tendência de pesquisas desenvolvidas com o objetivo de avaliar a influência das novas tecnologias informacionais no âmbito da gestão de recursos informacionais, como por exemplo, identificando os reflexos de tais tecnologias em instrumentos tradicionalmente utilizados na organização e recuperação de informações, como os tesouros.

Nesta perspectiva Arano (2005) apresenta uma compilação de alguns dos reflexos do ambiente digital na estrutura dos tesouros, pontuando quatro itens:

- O enriquecimento da funcionalidade da estrutura dos tesouros a partir da hipertextualidade, possibilitando o estabelecimento de hiperlinks entre os elementos estruturais e as diferentes partes do tesouro.
- A Redução dos custos de atualização e manutenção. Devido à crescente informatização dos processos de construção de tesouros e o progressivo abandono do suporte papel para a publicação destas ferramentas, viabilizando uma redução dos custos.

- A integração do usuário nos processos de criação, gestão e otimização dos tesouros, por meio de testes de viabilidade e uso de técnicas ajustadas ao usuário. Isso permite elaborar ferramentas que levem em conta os requerimentos dos usuários, e descartar sua construção como simples estruturas teóricas.
- A possibilidade de aplicar medidas de reutilização e interoperabilidade no momento de planejar e construir os tesouros. Possibilitando assim o aproveitamento e enriquecimento da informação conceitual e lingüística que é gerada para outros recursos.

Deste modo, com o aumento exponencial na quantidade de recursos informacionais disponíveis, e devido a necessidade de organiza-los e relaciona-los de maneira adequada, verifica-se uma crescente demanda de desenvolvimento de instrumentos que possibilitem descrever níveis de relacionamentos mais avançados e que possam ser processados de forma automatizada por computadores, os quais vêm sendo denominados como ontologias.

Assim, torna-se conveniente recorrer novamente à definição de Gruber (1993, p.1, tradução nossa), apresentada na seção 3.3, que define uma ontologia como: “uma especificação explícita de uma conceitualização”, na qual um objeto de nível simbólico será expresso a partir de uma linguagem formal, tendo como conceituação uma visão abstrata e simplificada do mundo a ser representado.

Nessa perspectiva, alguns autores considerando tal definição têm apresentado esquemas denominados como “espectros ontológicos” com o intuito de classificar as diversas variações que o termo ontologia pode assumir, de acordo com o seu nível de expressividade, abarcando desde instrumentos que possibilitam descrever um “baixo nível semântico” até instrumentos que possibilitam expressar arbitrariamente “relações semânticas mais complexas”. Conforme pode ser observado no gráfico apresentado por Lassila e McGuinness (2001, p.4).

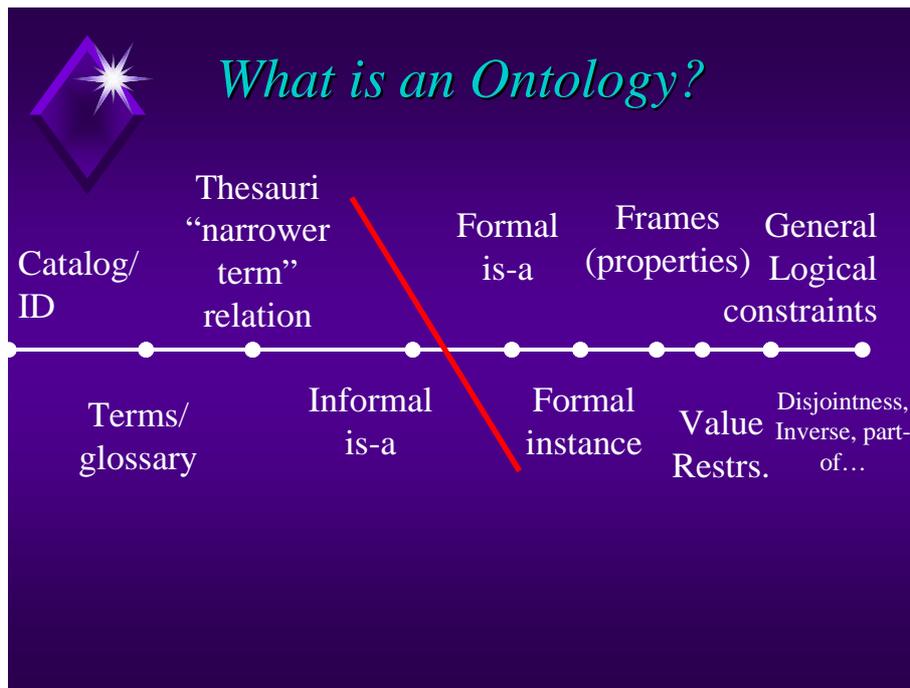


FIGURA 32 – An Ontology Spectrum
Fonte: Lassila e McGuinness (2001, p.4)

Como pode ser observado, essa classificação abarca desde listas de índices informais de termos, com definições em linguagem natural desestruturada, até ontologias com o máximo de expressividade, de modo que todos estes instrumentos de classificação objetivam estabelecer um vocabulário compartilhado que permita a troca de informações, porém, a partir de diferentes níveis de complexidade, formalismo e expressividade.

Lassila e McGuinness (2001), definem sucintamente os instrumentos mencionados no gráfico apresentado na figura 32. Com base em tais definições pode-se descrevê-los da seguinte maneira:

- **Catalog** – É uma lista finita de termos, e seus respectivos significados, que devem ser utilizados para representar informações em um contexto específico, de maneira padronizada.
- **Glossary** – É uma lista de termos com seus respectivos significados em linguagem natural, apresentados de maneira similar a um dicionário.
- **Thesauri** – É um vocabulário controlado que também fornece relacionamentos entre os termos.

- ***Informal Hierarchies*** – Hierarquias que utilizam relacionamentos informais, de modo que conceitos podem ser incorporados a uma categoria, mesmo que formalmente não faça parte da mesma. Um exemplo típico seria a incorporação dos conceitos “passagem aérea” e “hotel” na categoria “Viagem”, apesar de tais conceitos não representarem “tipos de viagens”. Tais hierarquias são comumente encontradas em alguns tipos de mecanismos de busca utilizados no ambiente Web, baseados em categorias temáticas.
- ***Formal Hierarchies*** – Hierarquias que incluem instâncias de um domínio, de modo que os relacionamentos são respeitados integralmente e descritos a partir de taxonomias.
- ***Frames*** – Modelos que incluem classes e propriedades, de modo que as propriedades não possuem escopo global, aplicando-se apenas nas classes para as quais foram definidas, possibilitando contextualizar informações em um domínio específico.
- ***Value Restrictions*** – Tipo de ontologias que definem restrições para os valores assumidos pelas propriedades de suas classes.
- ***Logic Constraints*** – Tipo de ontologias que possibilitam a definição de restrições lógicas, favorecendo a realização de inferências automatizadas.

Na figura 32, é possível verificar que na elaboração de tal classificação, os autores traçaram uma barra no centro da figura separando os instrumentos comumente utilizados por humanos daqueles descritos em linguagens formais, desenvolvidos para serem utilizados em ambientes computacionais. Segundo Gruber (1993) tal distinção é evidente na IA, que possui como foco as ontologias formais, onde a sintaxe e a semântica são descritas em linguagens formais, expressas em sentenças equivalentes à lógica de primeira ordem, possibilitando a definição de vocabulários baseados na lógica computacional.

Nessa perspectiva, compartilhando a mesma visão de ontologia apresentada por Lassila e McGuinness, pode-se destacar também o gráfico apresentado por Daconta, Obrst e Smith (2003, p.157), o qual de forma análoga ao anteriormente apresentado também foi denominado pelos referidos autores como “O espectro ontológico”, conforme pode ser verificado na figura 33:

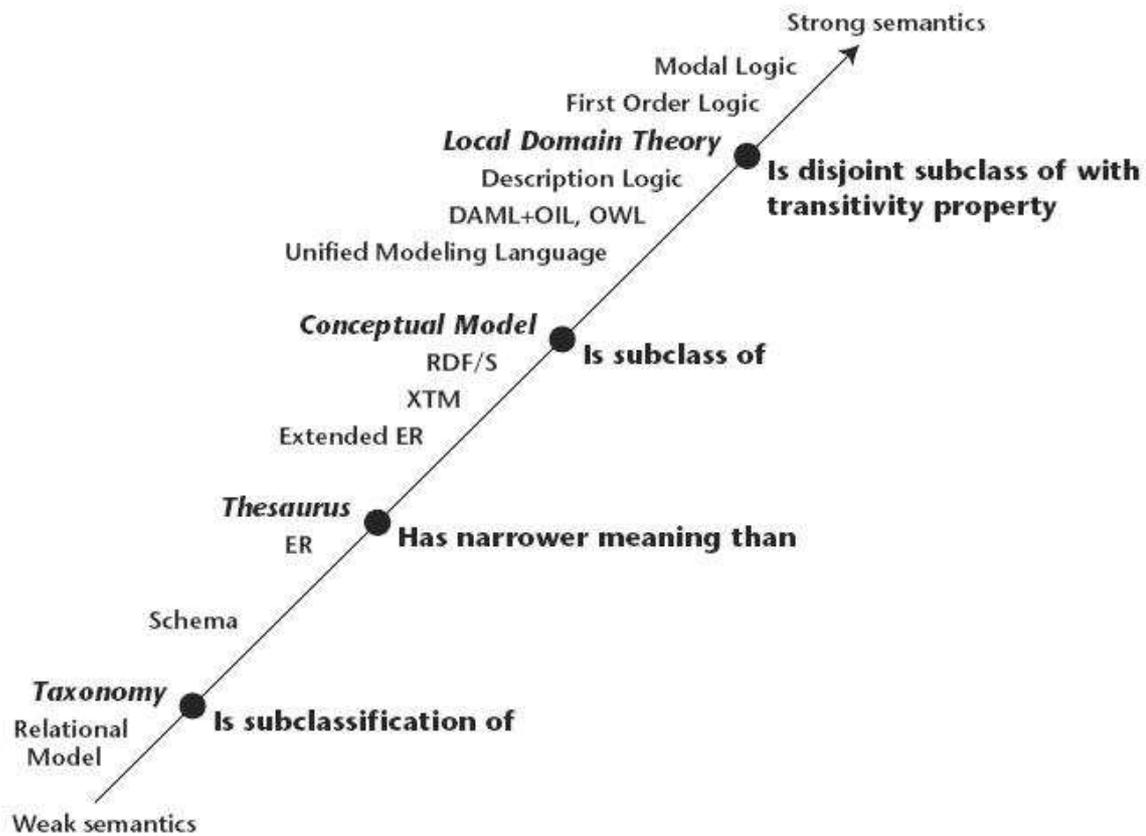


FIGURA 33 – The ontology spectrum: Weak to strong semantics
 Fonte: Daconta, Obrst e Smith (2003, p.157)

Nesse contexto, cabe ressaltar que os autores de tais definições (Lassila; McGuinness; Daconta; Obrst e Smith) são pesquisadores reconhecidamente atuantes do W3C, deste modo pode-se considerar que tais definições refletem a visão de ontologia compartilhada pela maioria dos pesquisadores deste consórcio e conseqüentemente da área de Ciência da Computação.

Assim, outras definições generalistas têm sido disseminadas na literatura de ontologias, como por exemplo, a definição apresentada por Jasper e Uschols, (1999, p.2, tradução nossa):

Uma ontologia pode assumir vários formatos, mas necessariamente deve incluir um vocabulário de termos e alguma especificação de seu significado. Esta deve abranger definições e uma indicação de como os conceitos estão inter-relacionados, o que resulta na estruturação do domínio e restringe possíveis interpretações de seus termos.

Nessa perspectiva, verifica-se que a partir de tais definições, instrumentos tradicionalmente utilizados no contexto da área de Ciência da Informação, já conceitualmente sedimentados, podem ser considerados como tipos potenciais de ontologias. Tal interpretação tem impulsionando uma série de pesquisas no âmbito da gestão de recursos informacionais, desenvolvidas com o intuito de identificar convergências e divergências existentes entre as ontologias e os instrumentos convencionais de representação.

Segundo Qin e Paling (2001) as ontologias são instrumentos adequados para as definições de um vocabulário de representação e coincidem com os tesouros e com as classificações em sua capacidade para representar o conteúdo de um documento por meio da abstração das relações entre conceitos, porém destacam que as ontologias apresentam um nível mais alto de concepção e de descrição do vocabulário, caracterizando-se por um desenvolvimento semântico mais profundo para relações do tipo classe/subclasse e relações cruzadas, devido à utilização da lógica descritiva. Tais autores ressaltam também a reusabilidade das ontologias e a possibilidade de trabalhar com sistemas heterogêneos.

Na mesma linha de pensamento, Ding e Foo (2002) afirmam que as diferenças fundamentais entre ontologias e os instrumentos convencionais de representação situam-se em nível de abstração, nas relações entre conceitos, na capacidade de ser compreensível para as máquinas e, o mais importante, na expressividade que podem proporcionar.

Do mesmo modo, para Sánchez e Martinez (2002) o que diferencia as ontologias dos tesouros é o fato das ontologias possibilitarem uma maior variedade de relações entre conceitos, tais relações são criadas a partir do modelo conceitual existente no domínio que se está formalizando, deste modo, as ontologias permitem fazer deduções mais complexas sobre a árvore taxonômica, favorecendo a sua utilização em sistemas documentais, o que possibilita diversas possibilidades de recuperação de informação.

Para Arano (2005) a diferença entre ontologias e tesouros baseia-se principalmente no nível de abstração como se constitui a organização conceitual, de modo que as ontologias permitem descrever formalmente relações semânticas complexas, favorecendo um maior nível de expressividade e possibilitando tanto a usuários humanos como a programas computacionais interpretar tais relações, enquanto que os tesouros são ferramentas com menor expressividade semântica, desenvolvidas para serem interpretadas apenas por usuários humanos.

Em contra partida, Moreira, Alvarenga e Oliveira (2004, p.23-24) afirmam que esta visão, de que a distinção entre tesouros e ontologias estaria no fato das ontologias permitirem uma maior variedade de relações, não procede e advém da falta de entendimento do que é um termo e o que é uma relação segundo a teoria dos tesouros, pois as relações observadas no domínio são representadas nos tesouros da mesma forma que qualquer outro conceito, enquanto que nas ontologias as relações são representadas de forma distinta das propriedades (classes) e a elas podem ser atribuídas restrições e propriedades estruturais, que podem ser utilizadas na realização de inferências. No entanto, as referidas autoras também apresentam semelhanças entre tesouros e ontologias, sendo ambos instrumentos constituídos a partir de um sistema de conceitos, e destacam que a diferença entre tais instrumentos pode ser identificada principalmente em termos de linguagem, de nível de formalização e de propósitos.

Segundo Garcia Jiménez (2004), ambos os formatos de representação do conhecimento têm suas conseqüências teóricas e práticas, de modo que não é fácil afirmar que os tesouros podem ser entendidos como tipos de ontologias, não apenas por suas diferentes trajetórias históricas, como também por suas diferentes vinculações operativas e teóricas, no entanto, é possível identificar similaridades entre os instrumentos.

Nesse contexto, com o intuito de facilitar a distinção entre ontologias e os instrumentos convencionais de representação, utilizados no âmbito da área de Ciência da Informação, pode-se definir uma ontologia como: Um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica e formal, a partir da descrição dos aspectos semânticos de conteúdos informacionais, possibilitando a realização de inferências automáticas por programas computacionais. Destacando assim o seu propósito e as novas possibilidades oferecidas no contexto da recuperação de informações.

Desta maneira, verifica-se que apesar das ontologias abordarem níveis de relacionamentos e aspectos lógicos, explícitos, mais abrangentes que os instrumentos convencionais de representação, frente ao estado incipiente das ontologias, há de se destacar a maior tradição dos tesouros e sistemas de classificação. Nessa perspectiva, é evidente a necessidade de estudos que apontem as contribuições da área de Ciência da Informação para o desenvolvimento de ontologias, a partir de seus aportes teóricos, pois se observa atualmente uma forte tendência de aproveitamento dos tradicionais instrumentos de representação como bases para o desenvolvimento de ontologias, conforme será apresentado na seção seguinte.

4.2 A Web Semântica sob o prisma da Ciência da Informação

Assim, a partir da análise apresentada, pode-se afirmar que o objetivo principal do projeto Web Semântica é possibilitar a classificação de recursos informacionais disponíveis no ambiente Web a partir de categorias que possam ser “interpretadas” automaticamente por computadores, possibilitando a realização de tarefas mais sofisticadas de forma automatizada.

Desta maneira, verifica-se que os estudos relacionados ao projeto Web Semântica possuem grande familiaridade com os instrumentos desenvolvidos no âmbito da área de Ciência da Informação, de modo que os aportes teóricos identificados nesta área do conhecimento podem contribuir substancialmente para o seu desenvolvimento.

Nessa perspectiva, destacam-se os estudos desenvolvidos no *Semantic Web Advanced Development for Europe*⁹ (SWAD-E), que têm demonstrado um forte interesse de pesquisadores ligados ao W3C nos instrumentos de representação tradicionalmente utilizados no âmbito da área de Ciência da Informação, conforme pode ser identificado na *SWAD-E Thesaurus Activity*¹⁰, uma iniciativa que tem como objetivo desenvolver tecnologias que possibilitem expressar, de maneira formal, a estrutura básica e o conteúdo de tesouros e vocabulários controlados, possibilitando a sua utilização de forma automatizada em *Knowledge Organization Systems* (KOS), Sistemas de Organização do Conhecimento.

Segundo Binding e Tudhope (2004), é possível identificar nos últimos anos uma crescente demanda por pesquisas relacionadas a KOS, devido principalmente ao rápido crescimento de comunidades científicas empenhadas no desenvolvimento de projetos relacionados com: Web Semântica, *Semantic Grid* e Bibliotecas Digitais. Deste modo, várias iniciativas têm sido propostas em recentes encontros do *Networked Knowledge Organization Systems/Services*¹¹ (NKOS) com o intuito de atualizar os padrões de tesouros internacionais para que considerem esses desenvolvimentos on-line.

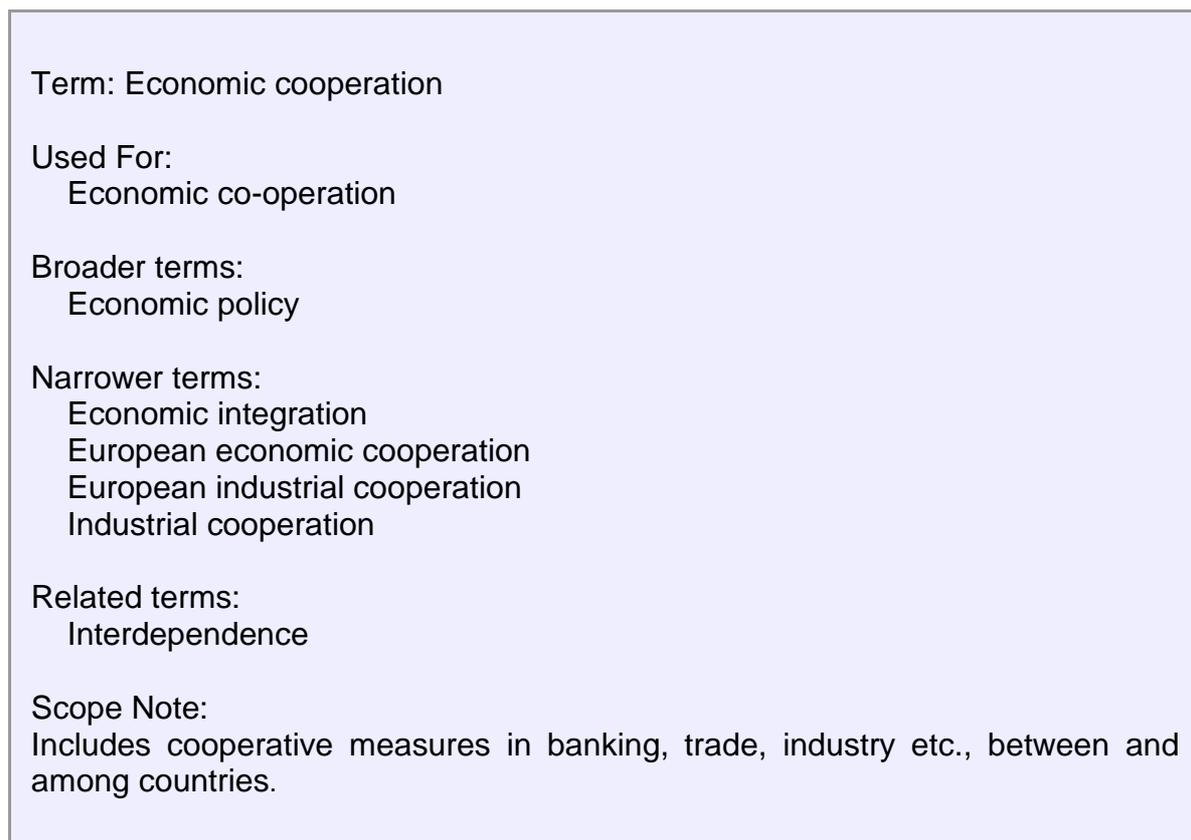
⁹ <http://www.w3.org/2001/sw/Europe>

¹⁰ <http://www.w3.org/2001/sw/Europe/reports/thes>

¹¹ <http://nkos.slis.kent.edu/>

Nesse contexto, o W3C tem se empenhado no desenvolvimento de padrões que dêem suporte ao uso de KOS a partir da estrutura da Web Semântica, podendo-se destacar o *Simple Knowledge Organisation System* (SKOS). Segundo recentes trabalhos apresentados por pesquisadores ligados ao W3C (Miles e Brickley, 2005), o SKOS representa desde estruturas simples utilizadas para expressar sistemas de organização do conhecimento, até estruturas mais complexas e poderosas, fornecendo um modelo para expressar a estrutura básica e o conteúdo de tesouros, esquemas de classificação, lista de cabeçalho de assunto, taxonomias e também outros esquemas conceituais.

O SKOS é um modelo, ainda em fase de desenvolvimento, que compreende um conjunto de propriedades descritas em RDF, a partir de classes RDFS, que podem ser utilizadas para expressar o conteúdo e a estrutura de um esquema de conceito como um gráfico RDF. A figura 34 apresenta parte de um tesouro, descrito no *SKOS Core Guide*, por Miles e Brickley (2005).



Term: Economic cooperation

Used For:
Economic co-operation

Broader terms:
Economic policy

Narrower terms:
Economic integration
European economic cooperation
European industrial cooperation
Industrial cooperation

Related terms:
Interdependence

Scope Note:
Includes cooperative measures in banking, trade, industry etc., between and among countries.

FIGURA 34 – An extract from the UKAT
Fonte: Miles e Brickley (2005).

O UK Archival Thesaurus¹² (UKAT) é um tesouro de assunto que tem sido desenvolvido no setor de arquivo do Reino Unido, um vocabulário controlado que arquivos podem utilizar para organizar suas coleções e catálogos.

Assim, é possível representar as informações contidas na figura 34, a partir do vocabulário SKOS, permitindo inclusive expressar um gráfico RDF que indique os relacionamentos entre os termos, a partir de descrições formais, conforme apresentado a seguir na figura 35.

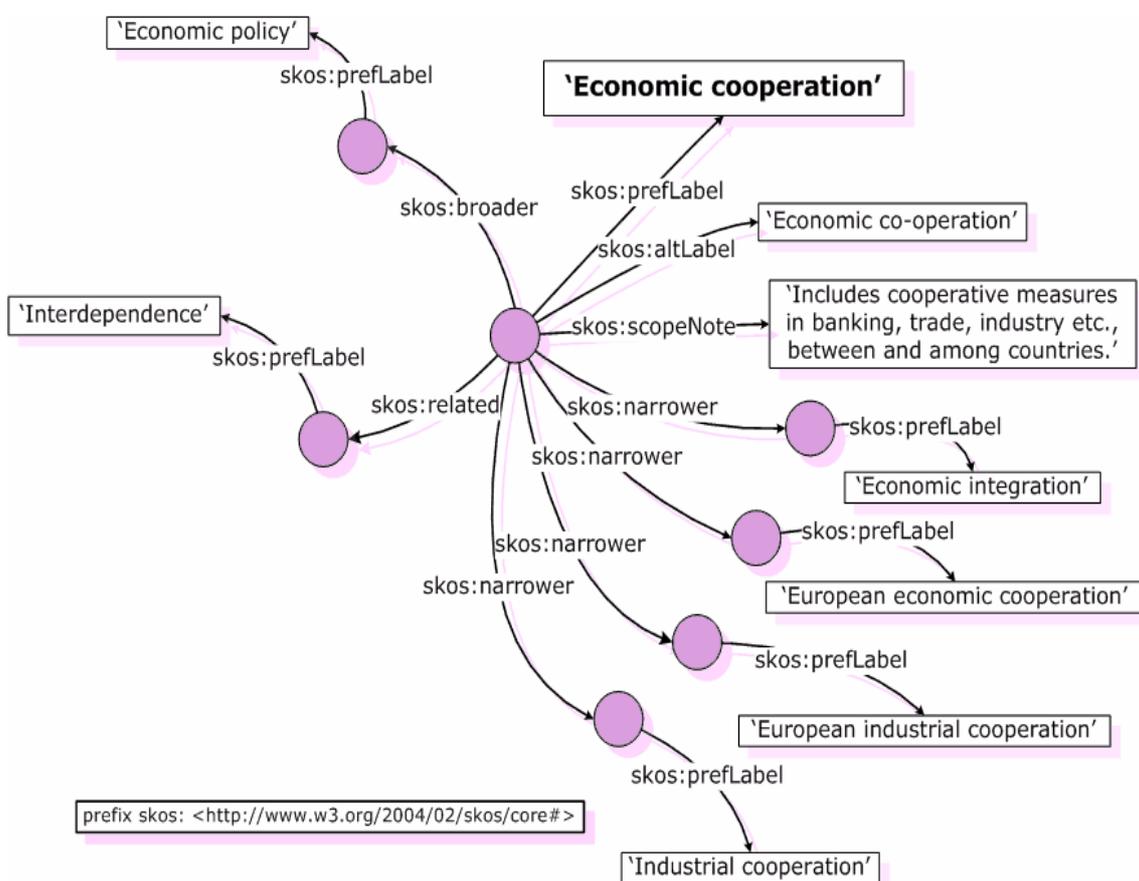


FIGURA 35 – Relacionamentos da figura 34 apresentados a partir do SKOS
Fonte: Miles e Brickley (2005).

¹² <http://www.ukat.org.uk/>

Nesse contexto, ressalta-se que a área de Ciência da Informação pode contribuir significativamente para o desenvolvimento da Web Semântica, pois os instrumentos de representação utilizados nessa área foram desenvolvidos a partir de aportes teóricos, e podem servir como substratos para o desenvolvimento de novos tipos de instrumentos de representação da informação no ambiente digital.

Nessa perspectiva, baseando-se no “Espectro Funcional”, figura 13, e a partir do enfoque da área de Ciência da Informação, pode-se descrever as camadas do projeto Web Semântica da seguinte maneira:

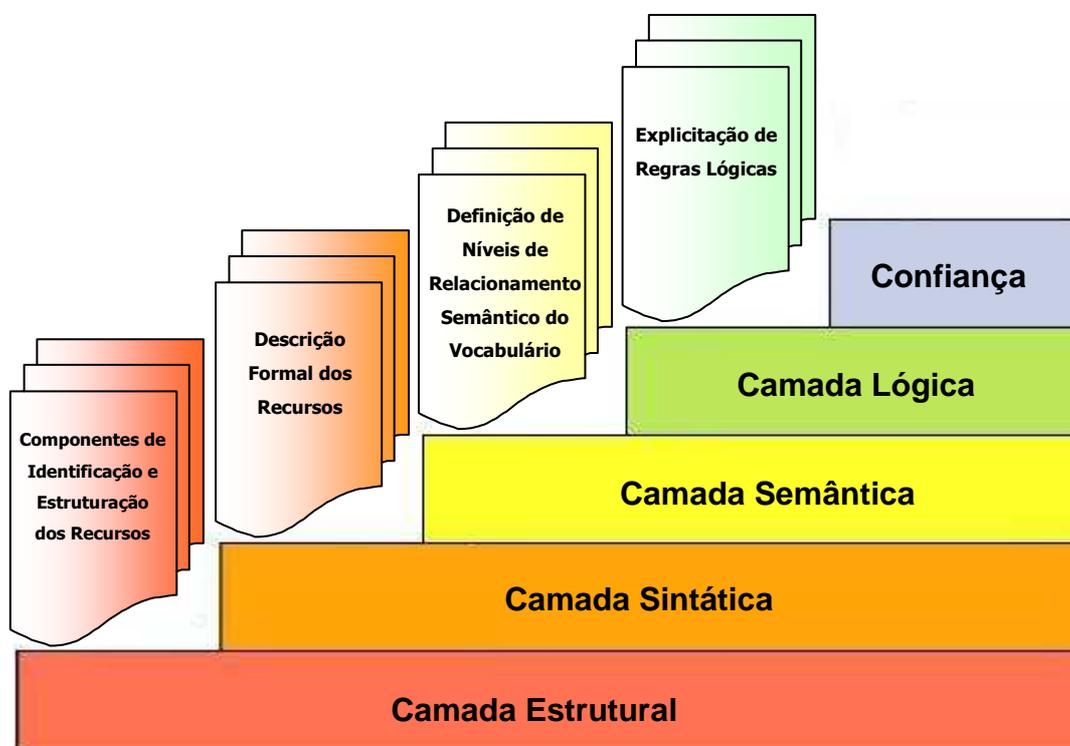


FIGURA 36 – Espectro Funcional das camadas da Web Semântica

Assim, conforme apresentado na Seção 3.2, para a concretização do projeto Web Semântica torna-se necessária primeiramente a identificação dos recursos a partir da “Camada Estrutural”, a qual permite identificar cada recurso de forma única e padronizada e possibilita meios seguros de transmissão e armazenamento das informações. A “Camada Sintática” possibilita a descrição dos recursos, por meio da

definição e validação de regras sintáticas formalmente descritas, possibilitando a estruturação dos recursos informacionais. Nesse contexto, pode-se destacar o forte embasamento da área de Ciência da Informação no que se refere a práticas de catalogação e indexação. Tal embasamento teórico pode ser de grande valia para o desenvolvimento do projeto Web Semântica.

Quanto à “Camada Semântica”, é nesta camada que se espera que sejam desenvolvidos vocabulários que permitam descrever os aspectos semânticos inerentes aos recursos informacionais e sistemas de conceitos que definam formalmente as relações existentes, de modo que a partir de tais vocabulários possam ser definidas, na “Camada Lógica”, as regras a serem interpretadas computacionalmente, possibilitando a realização de inferências automáticas e a verificação do nível de coerência lógica dos recursos. Assim, cabe ressaltar que a partir dos instrumentos de representação utilizados tradicionalmente no âmbito da área de Ciência da Informação (tesauros, lista de cabeçalhos de assunto, taxonomias, etc), é possível o desenvolvimento de ontologias, permitindo a representação formal dos relacionamentos existentes entre os termos e conceitos.

Segundo Garcia Jiménez (2004, p.90, tradução nossa): “As relações entre ontologias e tesauros parecem demonstrar uma tendência evidente: a intenção por parte de diversos especialistas em elaborar determinadas ontologias a partir de um tesouro”.

Deste modo, observa-se que para a concretização do projeto Web Semântica é necessária a utilização intensiva de lógicas computacionais, para que assim possa ser realizada, na “Camada de Confiança”, a comprovação de que os aspectos semânticos dos recursos estão descritos de modo consideravelmente adequado, atendendo a todos os requisitos das camadas anteriores e possibilitando um certo grau de confiança das informações.

Nesse contexto, é possível verificar que muitos dos instrumentos convencionais de representação podem ser aproveitados no desenvolvimento do projeto Web Semântica, do mesmo modo que muitas tecnologias desenvolvidas a partir do desenvolvimento do projeto Web Semântica não limitam seu escopo de aplicabilidade ao ambiente Web, como, por exemplo, a linguagem computacional XML e mais recentemente o desenvolvimento de ontologias que têm despertado o interesse de inúmeros

pesquisadores da área de Ciência da Informação, conforme pode ser observado nos trabalhos de Qin e Paling (2001), Ding e Foo (2002), Soergel (2002), Almeida e Bax (2003), Ferneda (2003), Golbeck et al. (2003), Pincemin (2003), Arano e Codina (2004), Campos (2004), Garcia Gimenez (2004), Pérez Agüera (2004), Souza e Alvarenga (2004), Moreira e Oliveira (2005), Ramalho, Vidotti e Fujita (2005), entre outros.

Conforme afirmam Souza e Alvarenga (2004, p. 139), no artigo intitulado: *A Web Semântica e suas contribuições para a ciência da informação*: “Tudo indica que os padrões que estão sendo desenhados para esta nova Web também sejam adotados na arquitetura de bibliotecas digitais e de novos sistemas de informação”.

Nessa perspectiva, é possível identificar atualmente alguns projetos que têm como principal objetivo o desenvolvimento de novas tecnologias no âmbito de bibliotecas digitais utilizando-se das tecnologias relacionadas ao projeto Web Semântica, entre os quais pode-se destacar o JeromeDL e o MarcOnt.

Conforme Kruk et al. (2005) apresentam no artigo sugestivamente intitulado como *JeromeDL - Reconnecting Digital Libraries and the Semantic Web*, o projeto JeromeDL consiste de uma biblioteca digital de código aberto baseada nas principais tecnologias presentes no projeto Web Semântica, permitindo a descrição de recursos a partir da linguagem computacional RDF e a realização de buscas semânticas baseadas em ontologias, possibilitando uma melhora considerável na precisão das buscas e um maior nível de interoperabilidade.

Quanto ao projeto MarcOnt, segundo Synak e Kruk (2005), o principal objetivo deste projeto é criar uma ontologia capaz de tornar-se um padrão de representação de informações para bibliotecas digitais, possibilitando a descrição dos aspectos semânticos dos conteúdos e favorecendo a integração de bibliotecas. Assim, está em fase de desenvolvimento e avaliação a ontologia MarcOnt, desenvolvida utilizando-se a linguagem OWL, de modo que se espera que tal ontologia seja compatível com o formato MARC 21, permitindo que as descrições semânticas possam ser convertidas para outros formatos, possibilitando grande interoperabilidade e o reaproveitamento das bases de conhecimento, por meio da incorporação de outras ontologias que sigam os mesmos critérios.

O formato MARC, *Machine Readable Cataloging*, foi desenvolvido com o intuito de suprir as necessidades de informatização de catálogos bibliográficos, possibilitando a catalogação e o intercâmbio de registro bibliográficos em um formato que pudesse ser legível por máquinas, favorecendo a redução de custos e minimizando os esforços necessários para se compartilhar informações. A partir da década de 1970 o formato MARC deu origem a diversos outros formatos em nível global, ocasionando dificuldades no intercâmbio de informações, de modo que a partir da harmonização dos formatos nacionais, o USMARC que foi desenvolvido em 1960 pela *Library of Congress – LC* (Biblioteca do Congresso Norte Americano), e o CAN/MARC desenvolvido em 1973 pela Biblioteca Nacional do Canadá, originou-se o padrão MARC 21 em 1998.

Cabe ressaltar que foi a partir da disseminação da linguagem computacional XML, conforme apresentado na seção 3.4.1, que se tornou possível o desenvolvimento da mais recente versão do formato MARC 21 em XML. Segundo Siqueira (2003, p. 82): “A grande diferença entre o formato tradicional do MARC 21 e a sua versão em XML está na estrutura utilizada para organizar os dados bibliográficos e catalográficos”.

Deste modo, pode-se considerar que as tecnologias subjacentes ao projeto Web Semântica e os instrumentos de representação de informações desenvolvidos no âmbito da área de Ciência da Informação possuem como objetivo comum propiciar meios mais adequados de representar e organizar conteúdos informacionais, possibilitando responder de maneira mais eficiente às buscas realizadas diretamente pelos usuários finais.

Contudo, deve-se ressaltar que muitas vezes devido às pressões para o desenvolvimento a curto prazo de novas tecnologias, buscando atender a demandas de mercado, as metodologias propostas a partir da área de Ciência da Computação podem não abarcar as devidas preocupações quanto às possíveis conseqüências que tais tecnologias possam acarretar, subestimando uma das principais preocupações da área de Ciência da Informação, quanto a indexação de informações.

Conforme relata Fujita (2003, p.180):

Uma das principais preocupações dos pesquisadores em indexação é a rápida evolução das técnicas de recuperação automática, acarretando o aumento da responsabilidade do indexador na determinação do assunto do documento. Novas formas de recuperação exigem maior aprofundamento teórico do indexador para que se evite o risco de uma prática descompromissada com a representação do contexto do documento e do sistema de recuperação de informação.

Nesta perspectiva, verifica-se que os profissionais da informação têm grandes responsabilidades dentro deste contexto, desempenhando um importante papel de agentes sociais, de modo que um de seus principais desafios é favorecer a diminuição das desigualdades no acesso à informação, a partir da utilização das novas tecnologias de acordo com princípios éticos que respeitem as especificidades, subjetividades e os valores de cada indivíduo ou comunidade.

Segundo Fernández-Molina e Guimarães (2002), quando tratamos das questões éticas que envolvem os profissionais da informação e as novas tecnologias, em geral levamos os conceitos e aplicações dentro de um nível restrito, deixando desta forma de gerar análises mais abrangentes sobre seu uso em uma esfera global. Assim, é evidente que o projeto Web Semântica, a partir da criação de categorias para a classificação dos recursos informacionais disponíveis no ambiente Web, trará no bojo de seu desenvolvimento novos dilemas éticos em uma escala global.

Buchanan (1999) afirma que os profissionais da informação têm a obrigação moral de responder aos novos dilemas éticos emergentes e de se esforçar para balancear as limitações práticas e tecnológicas de fornecimento e uso da informação.

Nessa perspectiva, conforme afirma Garcia-Gimenez (2004), deve-se lembrar que as novas tecnologias informacionais relacionadas à organização e recuperação de informações nasceram imersas nos ambientes computacionais, de modo que tal imbricação e dependência tecnológica exigem de seus criadores determinados conhecimentos e habilidades próprias em um contexto técnico e qualificado, para que possam representar o conhecimento e permitir sua posterior recuperação a partir de métodos que superem os tradicionais, melhorando sua eficácia.

Deste modo, torna-se evidente a necessidade de uma maior familiarização dos profissionais da informação com as novas tecnologias, para que as mesmas também possam ser desenvolvidas a partir de princípios éticos sociais e não baseadas única e exclusivamente em conhecimentos e processos puramente técnicos, pois só assim os profissionais da informação estarão realmente assumindo e desempenhando seu verdadeiro papel como agentes sociais.

5 CONSIDERAÇÕES FINAIS

É inegável o fato de que a ambiente o Web constitui-se como uma das mais ricas fontes de informações da atualidade, apresentando-se também como um ambiente interativo que possibilita a troca de informações em escala global. Tal fato, que à primeira vista apresenta ser o seu maior apelo, é ao mesmo tempo um de seus fatores críticos, pois devido ao aumento exponencial na quantidade de recursos informacionais disponíveis e a maneira como tais recursos são representados e estruturados no ambiente Web, torna-se a cada dia mais complexa a tarefa de localizar informações específicas.

Nesse contexto, a partir da análise da literatura apresentada, pode-se constatar que o projeto Web Semântica constitui uma evolução no modo como as informações são organizadas no ambiente Web, projetado com o intuito de possibilitar a incorporação de aspectos semânticos aos dados, favorecendo a contextualização das informações de forma automatizada, de acordo com o “contexto” no qual os dados estão inseridos e os critérios da busca realizados.

Observou-se que considerando os pressupostos da área de Ciência da Computação, é possível justificar a origem da expressão “Web Semântica”, devido ao fato que os “motores de busca” e as páginas Web tradicionais não levam em consideração os aspectos semânticos inerentes aos recursos informacionais. Fato que não ocorre caso um processo de recuperação de informação no ambiente Web seja analisado sob o prisma de um usuário que inegavelmente irá utilizar-se de aspectos semânticos para localizar as informações que procura.

Quanto à utilização do termo ontologia no âmbito da representação do conhecimento para denominar instrumentos de modelagem cognitiva de conceitos, objetos e seus respectivos relacionamentos no meio digital, que para alguns pode parecer imprópria, pode ser justificada a partir de uma análise mais profunda, de modo que uma ontologia formal possui, em sua essência, aspectos comuns à ontologia filosófica, do mesmo modo que a Lógica Matemática está relacionada com a Lógica Filosófica, apesar de serem utilizadas em contextos diferentes e buscando-se objetivos distintos. Nesta perspectiva, é notório o fato de que o termo ontologia adquire significados distintos de acordo com o domínio de utilização, sendo necessária uma contextualização adequada e cuidadosa, principalmente em áreas interdisciplinares por natureza, como é caso da área de Ciência da Informação.

Assim, no âmbito do projeto Web Semântica pode-se definir uma ontologia como um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica e formal, a partir da descrição dos aspectos semânticos de conteúdos informacionais, possibilitando a realização de inferências automáticas por programas computacionais.

Deste modo, verificou-se que os conceitos e tecnologias subjacentes ao projeto Web Semântica podem ser considerados como uma renovação ou desdobramento dos tradicionais métodos representação, organização e recuperação de informações, apontando a possibilidade de contribuições da área de Ciência de Informação para o desenvolvimento do projeto Web Semântica, devido ao seu embasamento teórico referente a formas de representação e as práticas profissionais identificadas em seu campo de atuação.

Nessa perspectiva, devido ao fato do projeto Web Semântica ter despertado o interesse de profissionais das mais variadas áreas do conhecimento, entre as quais destaca-se a área de Ciência da Informação, novos conceitos vêm sendo incorporados ou reformulados na área de Ciência da Computação, constituindo um novo paradigma computacional, segundo o qual os aspectos semânticos estão intrínsecos ao conceito de informação.

Portanto, torna-se evidente o caráter interdisciplinar que delinea o *corpus* teórico do projeto Web Semântica, englobando essencialmente áreas como a Ciência da Informação e a Ciência da Computação, entre outras, apresentando-se como um campo fértil para pesquisas, inclusive para a área de Ciência da Informação, a qual há longa data estuda conceitos inerentes ao projeto Web Semântica. Sendo necessários esforços interdisciplinares que possibilitem o desenvolvimento de soluções multidisciplinares, respeitando as especificidades de cada área do conhecimento e tendo como objetivo comum auxiliar na evolução do conhecimento humano de forma integral.

Verificou-se também que a partir dos estudos relacionados ao projeto Web Semântica, torna-se possível uma aproximação do objeto de estudo da Ciência da Informação com a área da Ciência da Computação, pois a partir do momento que os pesquisadores da área de Ciência da Computação empenham-se em desenvolver mecanismos que possibilitem descrever os aspectos semânticos inerentes aos recursos

informacionais, pode-se considerar que estão trabalhando com o mesmo conceito de informação utilizado pelos pesquisadores da área de Ciência da Informação. Barreto (2002, p.73) já sinalizava para tal fato, quando em seu artigo “A condição da Informação” utilizou a expressão “tecnologista da informação” ao se referir a Tim Berners-Lee, pesquisador diretamente ligado à área de Ciência da Computação e justamente o grande precursor do projeto Web Semântica.

Nesse contexto, observa-se que a concretização do projeto Web Semântica propiciará inúmeras vantagens aos usuários, de modo que se espera que tal projeto possibilite a cada indivíduo, instituição ou comunidade criar as suas próprias ontologias, favorecendo a contextualização das informações e uma maior liberdade no desenvolvimento de categorias mais adequadas para realidades específicas, frente aos atuais métodos de classificação, muitas vezes “impostos” e que nem sempre atendem às necessidades dos usuários.

Assim, é possível identificar atualmente inúmeras propostas de metodologias para a construção de ontologias, contudo verifica-se que tais metodologias, em sua maioria, não levam em consideração instrumentos e métodos já empregados com sucesso na área de Ciência da Informação, como por exemplo, as próprias técnicas de indexação e formatos de representação. Deste modo, ressalta-se a possibilidade de desenvolvimento de metodologias para construção de ontologias que reflitam os métodos tradicionais de representação e organização de recursos informacionais, já sedimentados na área de Ciência da Informação, possibilitando novas perspectivas do fazer profissional da área de Ciência da Informação e contribuindo substancialmente no desenvolvimento do projeto Web Semântica.

Sendo assim, verifica-se que os profissionais da informação podem assumir um importante papel no desenvolvimento de ontologias, porém, para isso torna-se necessária uma maior formação e especialização nos diversos aspectos relacionados às novas tecnologias, a partir da formação de equipes interdisciplinares que possibilitem otimizar os processos de transformação de tesouros em ontologias, pois ao mesmo tempo em que os profissionais da informação possuem muitas habilidades e embasamento teórico que podem ser aproveitados no desenvolvimento de novas abordagens tecnológicas de representação de informações, verifica-se também que tecnologias oriundas da área de

Ciência da Computação podem possibilitar melhorias significativas nos processos de gestão de recursos informacionais, contribuindo para um processo de retro-alimentação.

Deste modo, é possível considerar que o principal objetivo do projeto Web Semântica pode ser identificado pelo próprio *slogan* do W3C “*Leading the Web to Its Full Potential*”, Conduzir a Web para o Seu Potencial Máximo, entretanto, verifica-se poucas possibilidades de concretização de tal projeto de modo integral, ao menos a curto prazo, constituindo-se assim como um ideal abstrato muito mais do que uma possibilidade real a ser concretizada. Tal afirmação baseia-se no fato de que a concretização de uma Web Semântica global depende de uma série outros fatores sociais, econômicos, políticos e culturais, que vão além do desenvolvimento de novas tecnologias e padrões de representação. Conforme afirma Codina (2003), atualmente a expressão Web Semântica representa o rótulo de uma aspiração, um projeto de como seria idealmente o ambiente Web.

Nesta perspectiva, é inegável o fato de que as novas tecnologias têm provocado avanços consideráveis e verdadeiras revoluções em vários setores da sociedade, sendo evidente que o projeto Web Semântica traz no bojo de suas inovações tecnológicas uma série de avanços que podem possibilitar melhorias significativas nos processos de organização e recuperação de informações em ambiente digital. No entanto, observa-se também que tais tecnologias criam/agravam uma série de problemas sociais, econômicos, políticos e culturais, que em sua maioria ainda não foram enfrentados, ou muitas vezes sequer identificados, tanto no campo teórico quanto prático. Constituindo assim, um dos grandes desafios a ser superado por uma sociedade que almeja ostentar o título de “Sociedade da Informação”.

5.1 Pesquisas Futuras

A presente pesquisa caracterizou-se como uma discussão de literatura de caráter interdisciplinar, buscando identificar os principais conceitos inerentes ao projeto Web Semântica e estabelecer relacionamentos conceituais que permitissem uma maior sustentação teórica aos conceitos identificados, principalmente com relação ao referencial teórico da área de Ciência da Informação.

Nesse contexto, apresenta-se como perspectivas de continuidade desta pesquisa, estudos que possibilitem o desenvolvimento de protótipos de aplicações que demonstrem de modo prático a eficácia e aplicabilidade das tecnologias relacionadas ao projeto Web Semântica, baseando-se não apenas em aspectos técnicos, mas principalmente nos princípios teóricos e metodológicos identificados na área de Ciência da Informação.

Ressalta-se também a necessidade de elaboração de metodologias para o desenvolvimento de ontologias que reflitam os procedimentos tradicionalmente utilizados na área de Ciência da Informação, permitindo assim um melhor aproveitamento das habilidades dos profissionais da informação quanto à organização de recursos informacionais, disponibilizados em ambientes digitais, e favorecendo o desenvolvimento de aplicações que possibilitem otimizar as buscas realizadas diretamente pelos usuários finais.

REFERÊNCIAS

- ALMEIDA, M. B. Uma introdução ao XML, sua utilização na internet e alguns conceitos complementares. *Ciência da Informação*, Brasília, v. 31, n. 2, p. 5-13, maio/ago. 2002.
- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definição, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, Brasília, v. 32, n. 3, p. 7-20, set/dez. 2003.
- ALVARENGA, L. A teoria do conceito revisada em conexão com ontologias e metadados no contexto das Bibliotecas tradicionais e digitais. *DataGramaZero*, Rio de Janeiro, v. 2, n. 6, dez. 2001. Disponível em: <http://www.dgz.org.br/dez01/Art_05.htm> Acesso em: 15 set. 2005.
- ALVARENGA, L. Representação do conhecimento na perspectiva da Ciência da Informação em tempo e espaço digitais. *Encontros Bibli*, Florianópolis, n. 15, 1º. sem. 2003.
- ANSI Z39-19-2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda: NISO Press, 2005.
- ARANO, S. Los tesauros y las ontologías en la Biblioteconomía y la Documentación. *Hipertext.net*, n. 3, 2005. Disponível em: <<http://www.hipertext.net/web/pag260.htm>>. Acesso em: 5 jan. 2006.
- ARANO, S.; CODINA, L. La estructura conceptual de los tesauros en el entorno digital: nuevas posibilidades en la representación de la información y el control terminológico. In: *JORNADES CATALANES D'INFORMACIÓ I DOCUMENTACIÓ*, 9. Barcelona, 2004.
- ARAÚJO, G. B. *Sistemas de arquivos Windows e Unix*. 2003. Monografia (Especialização em Ciência da Computação) – Universidade Federal de Viçosa, Viçosa, 2003.
- BARRETO, A. de A. A condição da informação. *São Paulo em Perspectiva*, v.16, n.3, p.67-74, jul/set., 2002.
- BECHHOFFER, S. et al. OWL Web Ontology Language reference. *W3C recommendation*, 10 Fev. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-owl-ref-20040210>> Acesso em: 15 set. 2005.
- BEIGUELMAN, G. *O Livro Depois do Livro*. São Paulo: Peirópolis, 2003.

- BERNERS-LEE, T. *Semantic Web road map*. 1998. Disponível em: <<http://www.w3.org/DesignIssues/Semantic.html>>. Acesso em: 18 set. 2005.
- BERNERS-LEE, T. ;HENDER, J. ;LASSILA, O. The semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, New York, may. 2001.
- BERNERS-LEE, T. *Information Management: a proposal*. CERN, Genebra, mar. 1989. Disponível em: <<http://www.w3.org/History/1989/proposal.html>>. Acesso em: 20 Set. 2005.
- BERNERS-LEE, T. *Semantic Web - XML2000*. 2000. Disponível em: <<http://www.w3.org/2000/Talks/1206-xml2k-tbl>>. Acesso em: 18 set. 2005.
- BERNERS-LEE, T. *Semantic Web Concepts*. 2005. Disponível em: Disponível em: <<http://www.w3.org/2005/Talks/0517-boit-tbl>>. Acesso em: 18 out. 2005.
- BERNERS-LEE, T. *The Semantic Web*. 2002. Disponível em: <<http://www.w3.org/2002/Talks/04-sweb-sloan/Overview.html>>. Acesso em: 20 set. 2005.
- BERNERS-LEE, T. *Weaving the Web*. San Francisco: Harper, 1999.
- BERNERS-LEE, T.; MASINTER, L.; MCCAHILL, M. Uniform Resource Locators (URL). *RFC Sourcebook*, n.1738, mar. 1994. Disponível em: <<http://www.networksorcery.com/enp/rfc/rfc1738.txt>>. Acesso em: 20 Set. 2005.
- BINDING, C.; TUDHOPE, D. KOS at your Service: Programmatic Access to Knowledge Organisation Systems. *Journal of Digital Information*, Texas, v. 4, no. 4, 2004.
- BOSAK, J. *XML, Java, and the future of the Web*. 1997. Disponível em: <<http://www.ibiblio.org/pub/sun-info/standards/xml/why/xmlapps.htm>>. Acesso em: 18 set. 2005.
- BRAY, T. et al. Namespaces in XML 1.1. W3C Recommendation, 4 Feb. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-xml-names11-20040204>>. Acesso em: 15 set. 2005.
- BRAY, T.; PAOLI, J.; SPERBERG-MCQUEEN, C. M.; Extensible Markup Language (XML) 1.0. *W3C Recommendation*, 10 Feb. 1998. Disponível em: <<http://www.w3.org/TR/1998/REC-xml-19980210.pdf>>. Acesso em: 16 set. 2005.
- BRICKLEY, D. et al. RDF vocabulary description language 1.0: RDF schema. . *W3C recommendation*, 10 Fev. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-schema-20040210>>. Acesso em: 16 set. 2005.

- BRICKLEY, D. GUHA, R. V. Resource Description Framework (RDF) schema specification 1.0. *W3C candidate recommendation*, 27 Mar. 2000. Disponível em: <<http://www.w3.org/TR/2000/CR-rdf-schema-20000327>>. Acesso em: 15 set. 2005.
- BROWN, A. et al. XML Schema: formal description. *W3C Working Draft*, 25 Set. 2001. Disponível em: <<http://www.w3.org/TR/2001/WD-xmlschema-formal-20010925>>. Acesso em: 16 set. 2005.
- BUCHANAN, E.A. An overview of information ethics issues in a world-wide context. *Ethics and Information Technology*, Berlin, v.1, no. 3, p. 193-201, sep. 1999.
- BURNHAM, T. F. Tecnologias de informação e educação a distância: tecendo redes, interagindo com e-meios e ampliando espaços. In: _____; MATTOS, M. L. P. (Orgs). *Tecnologias de informação e educação a distância: tecendo redes, interagindo com e-meios e ampliando espaços*. Salvador: Edufba, 2004, 293 p.
- BUSH, V. *As we may think: the Atlantic monthly*. Boston. 1945. Disponível em: <<http://www.theatlantic.com/doc/194507/bush>>. Acesso em: 26 set 2005.
- CAMPOS, M. L. A. *A Organização de unidades de conhecimento em hiperdocumentos: o modelo conceitual como espaço comunicacional para a realização da autoria*. 2001. 98 p. Tese (Doutorado em Ciência da Informação) – Escola de Comunicações da Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001.
- CAMPOS, M. L. de A. *Modelização de domínios de conhecimento: uma investigação de princípios fundamentais*. Ciência da informação, Brasília, v. 33, n. 1, p. 22-32, jan./abr. 2004.
- CAMPOS, M. L. M. et al. Web semântica e a gestão de conteúdos informacionais. In: MARCONDES, C. H.; KURAMOTO, H.; TOUTAIN, L. B.; SAYÃO, L. (Orgs.). *Bibliotecas Digitais: Saberes e Práticas*. Salvador: EDUFBA; Brasília: IBICT, 2005. p. 55-75.
- CAPURRO, R. Epistemologia e ciência da informação. IN: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 5., 2003. Belo Horizonte. *Anais...* Belo Horizonte: UFMG, 2003.
- CASTRO, E. *XML para World Wide Web*. Rio de Janeiro: Campus, 2001.
- CEEInf - Comissão de Especialistas de Ensino de Computação e Informática. *Diretrizes Curriculares de cursos da área de Computação e Informática*. Brasília: CEEInf/MEC, 1999. Disponível em: <<http://www.inf.ufrgs.br/mec/ceeinf/diretrizes.html>>. Acesso em: 10 Jan 2006.
- CINTRA, A. M. M. et al. *Para entender as linguagens documentárias*. São Paulo: Polis, 2002. 152 p.

- CLARK, K. G. SPARQL Protocol for RDF. *W3C Working Draft*, 14 set.2005. Disponível em: <<http://www.w3.org/TR/2005/WD-rdf-sparql-protocol-20050914>>. Acesso em: 18 set. 2005.
- CODINA, L. La web semántica : una visión crítica. *El profesional de la información*, Barcelona, v. 12, n. 2, p. 149-152. 2003.
- DACONTA, M. C.; OBRST, L. J.; SMITH, K. T. *The Semantic Web: a guide to the Future of XML, Web Services, and Knowledge Management*. Indianápolis: Wiley Publishing, 2003. 279 p.
- DAUM, B.; MERTEN, U. *Arquitetura de sistemas com XML*. Rio de Janeiro: Campus, 2002.
- DAVIS, R. et al. What is a knowledge representation? *AI Magazine*, v.14, no.1, p.17-33, Menlo Park, USA. 1993. Disponível em: <<http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html>>. Acesso em 10 jan. 2006.
- DECKER, S. et al. The semantic Web: the roles of XML and RDF. *IEEE Internet Computing*, v. 4, no. 5, p. 63-74, set/out 2000. Disponível em: <<http://ieeexplore.ieee.org>>. Acesso em: 20 out. 2005.
- DING, Y.; FOO, S. Ontology research and development, Part 1 - A review of ontology generation. *Journal of Information Science*, v. 28, no. 2, p. 123-136. 2002.
- ENGELBART, C.; ENGLISH W. K. A research center for augmenting human intellect. *AFIPS Conference Proceedings of the 1968 Fall Joint Computer Conference*, San Francisco, v. 33, p. 395-410, 1968.
- ENGELSCHALL, R. S. *Apache Desktop Reference*. Boston: Addison Wesley, 2001.176p.
- FERNÁNDEZ MOLINA, J.C.; GUIMARÃES, J. A Ethical aspects of knowledge organization and representation in the digital environment: their articulation in professional codes of ethics. IN: López-Huertas, M.M., (Ed.). *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries*. Wurzburg: ERGON-Verlag, 2002. p. 487-492.
- FERNEDA, E. *Recuperação de Informação: análise sobre a contribuição da Ciência da Computação para a Ciência de Informação*. 2003. 137 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2003.
- FIELDING, R. Relative Uniform Resource Locators. *RFC Sourcebook*, no.1808, jun. 1995. Disponível em: <<http://www.networksorcery.com/enp/rfc/rfc1808.txt>>. Acesso em: 20 Set. 2005.

- FUJITA, M. S. L. *A Leitura Documentária do indexador: aspectos cognitivos e lingüísticos influentes na formação do leitor profissional*. 2003. 321f. Tese (Livre-Docência em Análise Documentária e Linguagens Documentárias Alfabéticas) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.
- GARCIA JIMÉNEZ. A. Instrumentos de representación del conocimiento: tesauros versus ontologías. *Anales de documentacion*, Mucia, n.7, p.79-95, 2004. Disponível em: <<http://www.um.es/fccd/anales/ad07/ad0706.pdf>>. Acesso em: 10 jan 2006.
- GILLIES, J.; CAILLIAU, R. *How The Web was Born*. New York: Osford University Press, 2000.
- GOLBECK et al. The National Cancer Institute's Thésaurus and Ontology. *Journal of Web Semantics*, v.1, no.1, p. 75-80, 2003.
- GROSOFF, B. et al. Description logic programs: Combining logic programs with description logics. In: *Proc. of WWW 2003*, Budapest, mai 2003, p. 48-57. Disponível em: <<http://www.cs.man.ac.uk/~horrocks/Publications/download/2003/p117-grosof.pdf>>. Acesso em: 16 set. 2005.
- GRUBER, T. R. *Toward principles for the design of ontologies used for knowledge sharing*. Knowledge Systems Laboratory, Stanford University, 1993. Disponível em: <<http://citeseer.ist.psu.edu/gruber93toward.html>>. Acesso em: 16 set. 2005.
- GUARINO, N. Some ontological principles for designing upper level lexical resources. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVOLUTION, 1., 1998, Granada. *Proceedings...* Granada: 1998.
- GUARINO, N.; GIARETTA, P. Ontologies and knowledge bases – towards a terminological clarification. In: N. MARS (Ed). *Towards very large knowledge bases: knowledge building and knowledge sharing*. Amsterdam: IOS Press, 1995. p. 25-32.
- HORROCKS, I., et al. Semantic web architecture: stack or two towers? In: FAGES, F.; SOLIMAN, S. (Ed.). *Principles and Practice of Semantic Web Reasoning*, (PPSWR 2005), no. 3703 in LNCS, 2005. p. 37-41.
- INTERNET WORLD STATS. *Top 20 Countries with the highest number of Internet users*. Jul. 2005. Disponível em: <<http://www.internetworldstats.com/top20.htm>>. Acesso em: 2 set. 2005.
- JACOBS, I. *About W3C: history*. 2006. Disponível em: <<http://www.w3.org/Consortium/history>>. Acesso em: 5 jan. 2006.
- JASPER, R.; USCHOLS, M. A Framework for understanding and classifying ontology applications. In: *KRR5-99*, Stockholm. 1999. Disponível em: <<http://sern.ucalgary.ca/KSI/KAW/KAW99/papers/Uschold2/final-ont-apn-fmk.pdf>> Acesso em: 10 de out. 2005.

JONES, D.; BENCH-CAPON, T.; VISSER, P. *Methodologies for ontology development*. 1998. Disponível em: <<http://www.iet.com/Projects/RKF/SME/methodologies-for-ontology-development.pdf>>. Acesso em: 08 jan. 2006.

KLEINROCK, L. UCLA to be first station in nationwide computer network. *UCLA Press Release*, July 1969. Disponível em: <<http://www.lk.cs.ucla.edu/REPORT/press.html>>. Acesso em: 20 dez. 2005.

KLYNE, G. et al. Resource Description Framework (RDF): concepts and abstract syntax. *W3C recommendation*, 10 Feb. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210>>. Acesso em: 15 set. 2005.

KOIVUNEN, M. R.; MILLER, E. *W3C Semantic Web Activity*. 2001. Disponível em: <<http://www.w3.org/2001/12/semweb-fin/w3csw>>. Acesso em: 10 out. 2005.

KRISHNAMURTHY, B.; REXFORD, J. *Redes para a Web*. Rio de Janeiro: Campus, 2001.

KRUK, S. R. et al. *JeromeDL reconnecting digital libraries and the semantic Web*. Disponível em: <http://www.marcont.org/marcont/pdf/www2005_jeromedl.pdf>. Acesso em: 10 out. 2005.

LANCASTER, F.W. *Indexação e resumo: teoria e prática*. Brasília: Briquet de Lemos, 2004. 452 p.

LARA FILHO, D. O fio de Ariadne e a arquitetura da informação na WWW. *DataGramaZero*, Rio de Janeiro, v. 4, n. 6, dez. 2003. Disponível em: <http://www.dgzero.org/dez03/Art_02.htm>. Acesso em: 15 set. 2005.

LARISSA, O.; SWICK, R. R. Resource Description Framework (RDF) model and syntax specification. *W3C recommendation*, 22 Feb. 1999. Disponível em: <<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>>. Acesso em: 15 set. 2005.

LASSILA, O.; MCGUINNES, D. L. *The role of frame-based representation on the semantic Web*. Jan. 2001. Disponível em: <<http://www.ksl.stanford.edu/people/dlm/etai/lassila-mcguinness-fbr-sw.html>>. Acesso em: 15 nov. 2005.

LE COADIC, Y-F. *A Ciência da informação*. Brasília: Briquet de Lemos, 1996.

LÉVY, P. *As Tecnologias da inteligência: o futuro do pensamento na era da informática*. Rio de Janeiro: Ed. 34, 1993, 208 p.

LÉVY, P. *Cibercultura*. São Paulo: Ed. 34, 1999, 264 p.

LIMA, V. A. M. *Terminologia, comunicação e representação documentária*. 1998. 118 f. Dissertação (Mestrado em Ciências) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1998.

LUTFI, A. A experiência brasileira. *Folha de São Paulo*. Caderno de Informática, São Paulo, 20 out. 1999. Disponível em: <<http://www1.folha.uol.com.br/fsp/informat/fr2010199919.htm>>. Acesso em: 18 set. 2005.

MACHADO, A. *Maquina e imaginário*. São Paulo: EDUSP, 1993.

MANOLA, F.; MILLER, E. RDF Primer. *W3C recommendation*, 10 Feb. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-primer-20040210>>. Acesso em: 16 set. 2005.

MCGUINNESS, D. L.; HARMELEN, F. OWL Web Ontology Language Overview. *W3C Recommendation*, 10 Feb. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-owl-features-20040210>>. Acesso em: 16 set. 2005.

MEY, E. S. A. Bibliotheca Alexandrina. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 1, n.2, p. 71-91. 2004. Disponível em: <<http://server01.bc.unicamp.br/seer/ojs/index.php>> Acesso em: 10 de out. 2005.

MILES, A.; BRICKLEY, D. SKOS Core Guide. *W3C Working Draft*, 02 Nov. 2005. Disponível em: <<http://www.w3.org/TR/swbp-skos-core-guide>>. Acesso em: 15 jan. 2006.

MILLER, E. An Introduction to the Resource Description Framework, *D-Lib Magazine*, mai. 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 10 set. 2005.

MILLER, E. *The semantic Web*. 2004. Disponível em: <<http://www.w3.org/2004/Talks/0120-semweb-umich/Overview.html>>. Acesso em: 16 set. 2005.

MOREIRA, A. *Tesouros e ontologias: estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o Método Analítico-Sintético*. 2003. 150 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2003.

MOREIRA, A.; OLIVEIRA, A. P. Contribuição da terminologia na modelagem de sistemas computacionais. *DataGramaZero*, Rio de Janeiro, v. 6, n. 5, out. 2005. Disponível em: <http://www.dgz.org.br/out05/Art_01.htm>. Acesso em: 20 set. 2005.

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A. P. O nível do conhecimento e os instrumentos de representação: tesouros e ontologias. *DataGramaZero*, v.5, n. 6, dez. 2004. Disponível em: <http://dgz.org.br/dez04/Ind_art.htm>. Acesso em: 20 set. 2005.

MUCHERONI, M. L. et al. SisProDiMeX - Um sistema distribuído de transporte de Imagens Médicas. In: CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE, 9., 2004, Ribeirão Preto. *Anais...*, 2004. v.1. p. 1-5.

NAKAMURA, E. T.; GEUS, P. L. *Segurança de redes em ambientes corporativos*. São Paulo: Futura, 2003.

PÉREZ AGÜERA, J. R. Automatización de tesauros y su utilización en la Web semántica. *textos universitarios de biblioteconomía i documentació*, n.13, 2004. Disponível em: <<http://www.ub.es/bid/13perez2.htm>>. Acesso em: 15 jan 2006.

PFÜTZENREUTER, E. *Aplicabilidade e desempenho do protocolo de transporte SCTP*. 2004. 119 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Santa Catarina, Florianópolis, 2004.

PINCEMIN, B. Thésaurus documentaires et ontologies: divergences et ressemblances. *Journée d'étude Web Sémantique*. Paris, 2003. Disponível em: <http://www-lli.univ-paris13.fr/membres/biblio/1195_pincemin_ws_0410.pdf>. Acesso em: 10 jan. 2006.

PIRES, H. F. A Produção morfológica do ciberespaço e a apropriação dos fluxos informacionais no Brasil. In.: COLOQUIO INTERNACIONAL DE GEOCRÍTICA, 7., 2005, Santiago de Chile. *Anais...* Santiago de Chile: 2005.

POLI, R. *Framing ontology*. 1999. Disponível em: <<http://www.formalontology.it>>. Acesso em: 30 set. 2005.

PRASAD, A. Creation of digital libraries in indian languages using UNICODE. In: *WORKSHOP ON DIGITAL LIBRARIES: THEORY AND PRACTICE*, 2003, Bangalore. Disponível em: <https://drtc.isibang.ac.in/bitstream/1849/72/2/I_unicode.pdf>. Acesso em: 16 set. 2005.

PRUD'HOMMEAUX, E.; SEABORNE, A. SPARQL query language for RDF. *W3C working draft*, 21 July 2005. Disponível em: <<http://www.w3.org/TR/2005/WD-rdf-sparql-query-20050721>>. Acesso em: 16 set. 2005.

QIN, J.; PALING, S. Converting a controlled vocabulary into an ontology: the case of GEM. *Information Research*, v. 6, no. 2, 2001. Disponível em: <<http://informationr.net/ir/6-2/paper94.html>> . Acesso em: 10 jan 2006.

RAMALHO, R. A. S.; VIDOTTI, S. A. B. G. ; FUJITA, M. S. L. Bibliotecas Digitais na era da Web Semântica: reflexões no âmbito da gestão de conteúdos informacionais. In: FUJITA, M. S. L. (Org.). *A dimensão social da biblioteca digital na organização e acesso ao conhecimento: aspectos teóricos e aplicados*. São Paulo:USP/IBICT, 2005, v. 2, p. 91-113.

RAYWARD, W. B. The origins of information science and the International Institute of Bibliography/International Federation for Information and Documentation (FID). *Journal of the American Society for Information Science*, no. 48, p. 289-300, Apr. 1997. Disponível em: <<http://alexia.lis.uiuc.edu/~wrayward/otlet/OriginsofInfoSci.htm>>. Acesso em: 8 jan. 2006.

SANTOS, A. L. Textualidade literária e hipertexto informatizado. In: ENCONTRO NACIONAL DA ABRALIC, 5., 1997, Florianópolis. *Anais...* Florianópolis: UFSC/CNPq, 1997.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 1, n. 1, p. 41-62, jan/jun. 1996.

SBC, Sociedade Brasileira de Computação. *Currículo de referência da SBC para cursos de graduação em Computação*. Campinas, 2003. Disponível em: <www.sbc.org.br/educacao>. Acesso em: 10 jan. 2006.

SEMANTIC WEB. *About semanticWeb.org*. 2002. Disponível em: <<http://www.semanticweb.org/about.html>>. Acesso em: 20 out. 2005.

SHANNON, C ; WEAVER, W. *The Mathematical theory of communication*. Urbana: University of Illinois Press, 1949. Disponível em: <<http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>>. Acesso em: 18 ago. 2005.

SHANNON, C. E. A Mathematical theory of communication. *The Bell System Technical Journal*, v. 27, p.379-423, 623-656, July/Oct. 1948.

SIQUEIRA, M. A. *XML na Ciência da Informação: uma análise do MARC21*. Marília, 2003, 134 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003.

SMITH, M. K. et al. OWL Web ontology language guide. *W3C recommendation*, 10 Feb. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-owl-guide-20040210>>. Acesso em: 16 set. 2005.

SOERGEL, D. Thesauri and ontologies in digital libraries: tutorial. In: *EUROPEAN CONFERENCE ON DIGITAL LIBRARIES*, (ECDL), 2002, Roma. Disponível em: <http://www.dsoergel.com/cv/B63_rome.pdf>. Acesso em: 10 jan. 2006.

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

STANTON, M. A Evolução das redes acadêmicas no Brasil: Parte 1 - da BITNET à Internet. *Boletim Bimestral sobre Tecnologia de Redes: RNP – Rede Nacional de Ensino e Pesquisa*, v. 2, n. 6, 1998. Disponível em: <<http://www.networksorcery.com/enp/rfc/rfc1738.txt>>. Acesso em: 20 set. 2005.

SYNAK, M.; KRUK, S. R. MarcOnt initiative the ontology for the librarian world. In: EUROPEAN SEMANTIC WEB CONFERENCE ESWC, 2., 2005, Heraklion, Grécia. *Proceedings...* Heraklion: 2005.

TEIXEIRA, C. The existence of the world: an introduction to ontology: de Reinhardt Grossmann. *Disputatio*, Lisboa, v.1. , n. 7, p. 54-58, nov. 1999.

TESCH JR., J.R. *XML Schema*. Florianópolis: Visual Books, 2002.

VIDOTTI, S. A. B. G. *O ambiente hipermídia no processo de ensino-aprendizagem*. 2001. 126 f. Tese (Doutorado em Educação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

VRANDEČIĆ, D. et al. *DLP: an introduction*. 2005. Disponível em: <<http://logic.aifb.uni-karlsruhe.de/download/dlpintro.pdf>>. Acesso em: 18 set. 2005.

WELTY, C., GUARINO, N. Supporting ontological analysis of taxonomic relationships. *Data and Knowledge Engineering*, v. 39, no.1, p. 51-74, 2001.

WIENER, N. *Cibernética*. São Paulo: EDUSP: Polígono, 1970.

WIRTH, A. L. J. *Utilizando na prática: internet e redes de computadores*. Rio de Janeiro: Alta Books, 2002.