

RACHEL CRISTINA VESÚ ALVES

WEB SEMÂNTICA:
uma análise focada no uso de metadados

Marília
2005

RACHEL CRISTINA VESÚ ALVES

WEB SEMÂNTICA:
uma análise focada no uso de metadados

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Informação, como parte das exigências para a obtenção do título de mestre em Ciência da Informação, da Faculdade de Filosofia e Ciências – Universidade Estadual Paulista - UNESP, Campus de Marília. Área: Informação, Tecnologia e Conhecimento. Linha: Informação e Tecnologia.

Orientador: **Plácida Leopoldina Ventura Amorim da Costa Santos**
Pesquisa financiada pela CAPES

Marília
2005

Alves, Rachel Cristina Vesú.

A474a Web Semântica: uma análise focada no uso de metadados / Rachel
Cristina Vesú Alves. -- Marília, R. C. V. Alves, 2005.
180 f. ; 30 cm

Dissertação (Mestrado em Ciência da Informação) – Faculdade de
Filosofia e Ciências – Universidade Estadual Paulista, 2005.

Bibliografia: f. 169-180.

Orientadora: Profª Drª Plácida Leopoldina Ventura
Amorim da Costa Santos

1. Web Semântica. 2. Metadados. 3. Representação da informação.
I. Autor. II. Título.

CDD 025.316

Dedicatória...

*À Deus por mais essa conquista.
Aos meus pais Beni e Mauro,
e minha irmã Roberta.*

AGRADECIMENTOS...

Agradeço a Deus por permitir a conquista de mais um objetivo e a realização de mais um sonho. Obrigada por eu ter chegado até aqui, por me inspirar e amparar nos momentos difíceis.

Aos meus pais, Mauro e Beni, minha eterna gratidão. Passamos por muitas provas durante o caminho... mas juntos conseguimos superá-las. Obrigada pela torcida, apoio, paciência, amor e carinho. Obrigada por me incentivarem, por estarem ao meu lado e por me ajudarem em todos os momentos. Eu amo vocês!! À minha irmã Roberta, agradeço a torcida, ajuda e o apoio de todas as horas. Obrigada por estar ao meu lado. Eu te amo!!

À Professora Doutora Plácida L. V. A. da Costa Santos, minha orientadora, que admiro muito por sua competência profissional e pela pessoa maravilhosa que é, quero dizer que foi um orgulho ter sido sua orientanda. Agradeço pelos valiosos ensinamentos, pelo incentivo, compreensão e carinho. Foi um grande prazer estar ao seu lado em mais este trabalho. Obrigada por tudo!!

À professora Silvana Ap. B. G. Vidotti, que acompanha meu trabalho desde a graduação, agradeço pelos ensinamentos e contribuições durante todo esse tempo, pelas sugestões da qualificação e da defesa, que foram de grande valor para a conclusão desta pesquisa.

Ao professor Edberto Ferneda, pelos apontamentos e contribuições durante a qualificação e defesa, que foram importantes para o término desta pesquisa.

Aos docentes do curso de Pós-Graduação em Ciência da Informação pelos valiosos ensinamentos transmitidos.

Ao pessoal da Pós-Graduação, colegas de turma e funcionários.

Aos funcionários da biblioteca que sempre me atenderam com muita simpatia, em especial a Luzinete, pelo auxílio no levantamento bibliográfico e na normalização.

À Sylvia do escritório de pesquisa, pela sua simpatia e disposição para ajudar a solucionar as dúvidas de normalização.

A CAPES pelo financiamento desta pesquisa, durante o período de novembro de 2003 a agosto de 2005.

Aos familiares que sempre torceram por mim, em especial a Renata, a minha avó, meus tios, irmãos, primos e sobrinhos.

Aos meus amigos pelo apoio e por entenderem meus momentos de ausência durante o desenvolvimento dessa dissertação, em especial agradeço à: Mara Patrícia, Fabiana Straioto, Lourdes Mariano, Lucilene, Kamila e Paulo pelo apoio e pelos momentos de alegria.

E por fim agradeço a todas as pessoas que direta ou indiretamente contribuíram para a realização desse trabalho.

Muito Obrigada!!!

O que sabemos é uma gota.
O que ignoramos é um oceano.
(Isaac Newton)

ALVES, R. C. V. *Web Semântica: uma análise focada no uso de metadados*. 2005. 180 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia Ciências, Universidade Estadual Paulista, Marília, 2005.

RESUMO

Atualmente a nossa sociedade, denominada sociedade da informação, vem sendo caracterizada pela valorização da informação, pelo uso cada vez maior de tecnologias de informação e comunicação e pelo crescimento exponencial dos recursos informacionais disponibilizados em diversos ambientes, principalmente na Web. Essa realidade trouxe algumas mudanças no acesso automatizado às informações. Se por um lado temos uma grande quantidade de recursos informacionais disponibilizados, por outro temos como consequência problemas relacionados à busca, localização, acesso e recuperação dessas informações em ambientes digitais. Nesse contexto, o problema que originou essa pesquisa está relacionado com a dificuldade na busca e na recuperação de recursos informacionais digitais na Web e a ausência de tratamento adequado para a representação informacional desses recursos. O maior desafio para a comunidade científica no momento está na identificação de padrões e métodos de representação da informação, ou seja, na construção de formas de representação do recurso informacional de maneira a proporcionar sua busca e recuperação de modo mais eficiente. Assim, a proposição apontada nesse trabalho como solução do problema refere-se ao estabelecimento da Web Semântica e a aplicação de padrões de metadados para a representação da informação, pois são consideradas como iniciativas importantes para proporcionar uma melhor estruturação e representação dos recursos informacionais em ambientes digitais. Com uma metodologia baseada na análise exploratória e descritiva do tema a partir da literatura disponível, apresenta-se uma análise da Web Semântica como uma nova proposta para organização dos recursos informacionais na Web e as ferramentas tecnológicas que permeiam sua construção, com enfoque no uso de metadados como elemento fundamental para proporcionar uma melhor representação dos recursos informacionais disponibilizados na Web e sua posterior recuperação. A proposta da Web Semântica é disponibilizar recursos informacionais melhor estruturados e representados, formando uma rede de informações conectadas que por meio de ferramentas tecnológicas, tais como: os agentes de software, a linguagem de marcação XML, arquitetura de metadados RDF, ontologias e, principalmente, padrões ou formatos e metadados. Como resultado pode-se destacar que a implementação da Web Semântica requer o trabalho conjunto das várias ferramentas tecnológicas estudadas e que proporcionará em pequena, média e grande escala a tão necessária estruturação e representação informacional dos recursos e consequentemente sua melhor recuperação. Além disso, foi possível verificar que as tecnologias da Web Semântica convergem para a área de Ciência da Informação, estabelecendo uma estreita relação na questão da representação do conhecimento, principalmente com relação ao uso de metadados que são considerados essenciais para se estabelecer uma boa representação dos recursos informacionais na rede. Sabendo que a representação da informação é necessária em qualquer ambiente para proporcionar uma recuperação mais eficiente, podemos considerar os metadados como ferramentas essenciais para estabelecer a representação dos recursos informacionais no ambiente da Web Semântica como instrumentos para a construção de uma rede de conhecimentos e recuperação da informação de modo mais eficiente.

Palavras-chave: Web Semântica, Metadados, Representação da Informação, Web, Ontologias, Arquitetura de Metadados RDF, Linguagem de Marcação XML.

ALVES, R. C. V. *Web Semântica: uma análise focada no uso de metadados*. 2005. 180 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia Ciências, Universidade Estadual Paulista, Marília, 2005.

ABSTRACT

Nowadays our society, named society of information, has been characterized by the valorization of information through the increasing use of the information and communication technologies and the exponential growth of the informational resources, available in various environments, mainly on the Web. This reality has brought some changes for the automated access to information. If we have a big amount of informational resources available at one side, on the other we have problems related to search, localization, access and recuperation of this information in digital environments as a consequence. In this context, the problem that originated this research is related to the difficulty on searching and recuperating digital informational resources on the Web, and the lack of adequate treatment for the informational representation of these resources. At the moment, the biggest challenge for the scientific community is to identify patterns and methods of representation of information, that is, the construction of forms of representation of the informational resource in order to provide its search and recuperation in a more efficient manner. So, the pointed proposition for the solution of the problem, in this paper, refers to the Semantic Web establishment and the application of metadata patterns to the representation of information, because they are considered an important initiative for providing a better structuring and representation of the informational resources in digital environments. With a methodology based on the exploratory and descriptive analysis of the theme, beginning from the available literature, it is possible to present a Semantic Web analysis as a new proposal for the organization of the informational resources on the Web, and the technological tools that permeate its construction, focusing the use of metadata as the fundamental element to provide a better representation of the informational resources available on the Web, and their posterior recuperation. The Semantic Web proposal is to make better structured and represented informational resources available, creating a net of information connected with the use of technological tools, such as: the software agents, the XML markup language, RDF metadata architecture, ontology, and, mainly, metadata patterns or formats. As a result, it is possible to highlight that the Semantic Web implementation requires working the various technological tools studied as a group, and that it will provide the so-needed structuring and representation of the informational resources at a small, medium and large scale, and consequently their better recuperation. Besides, it was possible to verify that the Semantic Web technologies converge to the Science of Information area, establishing a narrow relation to the matter of the representation of information, mainly related to the use of metadata that are considered essential for an efficient representation of the informational resources establishment on the Web. If we know that the representation of information is necessary in any environment to provide a more efficient recuperation, we can consider the metadata as being fundamental tools for establishing the representation of the informational resources in the Semantic Web environment as instruments for the construction of a knowledge net and the recuperation of the information in a more efficient way.

Keywords: Semantic Web, Metadata, Representation of Information, Web, Ontology, RDF Metadata Architecture, XML Markup Language.

LISTA DE SIGLAS

AACR2 - Anglo American Cataloguing Rules

CSS - Cascading Style Sheet

DC - Dublin Core

DCMI - Dublin Core Metadata Initiative

DTD - Document Type Definition

HTML - Hypertext Markup Language

HTTP - Hypertext Transfer protocol

MARC - Machine Readable Cataloging Format

MARXML - Machine Readable Cataloging Format in eXtensible Markup Language

MCF - Meta Content Framework

NISO - North American Standard Organization

RDF - Resource Description Framework

SGML - Standard Generalized Markup Language

URI - Uniform resource Identifier

URL - Uniform Resource Locator

W3C - World Wide Web Consortium

WWW - World Wide Web

Xlink - XML Linking Language

XML - eXtensible Markup Language

Xpath - XML Path Language

Xpointer - XML Pointer Language

XSL - eXtensible Stylesheet Language

XSTL - Extensible Stylesheet Language Transformation

LISTA DE FIGURAS

FIGURA 1: Arquitetura da Web Semântica comentada	30
FIGURA 2: Arquitetura mais simples de camadas da Web Semântica	35
FIGURA 3: Estrutura do documento HTML	64
FIGURA 4: Exemplo de um documento XML simples	68
FIGURA 5: Exemplo de uma DTD separada do documento XML	71
FIGURA 6: Exemplo de XML Schema	72
FIGURA 7: Exemplo de folha de estilo CSS	73
FIGURA 8: Exemplo de Xlink simples	74
FIGURA 9: Exemplo de <i>namespace</i>	75
FIGURA 10: Representações de um statement: grafo e tripla	89
FIGURA 11: Serialização em XML de descrições RDF	91
FIGURA 12: Serialização em XML abreviada de descrição RDF	91
FIGURA 13: Definição de tipos em RDF	92
FIGURA 14: Definição de tipos em RDF expressa em XML	92
FIGURA 15: Reitificação de um statement RDF	93
FIGURA 16: Aserções sobre um statement	94
FIGURA 17: Reitificação expressa em XML	95
FIGURA 18: Coleção Bag listando as medidas de temperaturas de uma região	96
FIGURA 19: Uma coleção Bag descrita em XML	96
FIGURA 20: Grafo de um Schema RDF	99
FIGURA 21: Definição de um Schema RDF em RDF/XML	100
FIGURA 22: Funcionamento da Web Semântica de acordo com a SemanticWeb.org	108
FIGURA 23: Tipologia de formatos de metadados	122
FIGURA 24: Relação entre identificador, recurso e representação	129
FIGURA 25: Elementos do padrão de metadados Dublin Core e sua categorização	133
FIGURA 26: Comparação entre elementos: Dublin Core Qualificado e Dublin Core não Qualificado	136
FIGURA 27: DC Qualificado	137
FIGURA 28: DC não Qualificado	138
FIGURA 29: Registro bibliográfico em formato de intercâmbio MARC 21	142
FIGURA 30: Representação com MARC 21	145
FIGURA 31: Representação de um recurso pelo formato MARCXML	148

SUMÁRIO

1 INTRODUÇÃO.....	11
1.1 DEFINIÇÃO DO PROBLEMA.....	12
1.2 PROPOSIÇÃO.....	13
1.3 OBJETIVOS.....	14
1.4 METODOLOGIA.....	15
1.5 JUSTIFICATIVA.....	17
2 WEB SEMÂNTICA: Uma nova proposta para organização e recuperação de recursos informacionais na rede.....	20
2.1 ARQUITETURA DA WEB SEMÂNTICA: o delineamento de uma nova Web.....	26
2.2 ONTOLOGIAS: conceitos e definições básicas.....	37
2.3 TIPOS E CARACTERÍSTICAS DAS ONTOLOGIAS.....	41
2.4 IMPORTÂNCIA DAS ONTOLOGIAS PARA A DEFINIÇÃO DE CONCEITOS SEMÂNTICOS NA REDE DE CONHECIMENTOS.....	53
3 ARMAZENAMENTO E ESTRUTURAÇÃO DOS RECURSOS INFORMACIONAIS NA WEB: o papel da linguagem XML.....	57
3.1 LINGUAGEM SGML – Standard Generalized Markup Language.....	61
3.2 LINGUAGEM HTML – HiperText Markup Language.....	63
3.3 LINGUAGEM XML – eXtensible Markup Language.....	67
3.3.1 CARACTERÍSTICAS DA LINGUAGEM XML E TECNOLOGIAS ASSOCIADAS.....	69
3.3.2 LINGUAGEM XML E A WEB SEMÂNTICA.....	76
4 INTEROPERABILIDADE NA WEB: uso da arquitetura de metadados RDF – Resource Description Framework.....	80
4.1 ARQUITETURA DE METADADOS: estabelecimento de interoperabilidade na rede.....	85
4.2 ARQUITETURA RDF: característica e estrutura.....	87
4.3 CROSSWALKS: ferramenta para o mapeamento entre formatos de metadados.....	103
4.4 INTEROPERABILIDADE NA REDE: trabalho conjunto entre aplicações e ferramentas.....	105
5 METADADOS PARA A REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS NA WEB SEMÂNTICA.....	107
5.1 METADADOS: origem, definição e características.....	110
5.2 FORMATOS DE METADADOS.....	124
5.2.1 FORMATOS DE METADADOS SIMPLES.....	126
5.2.2 FORMATOS DE METADADOS ESTRUTURADOS..	131
5.2.3 FORMATOS DE METADADOS RICOS.....	140
5.3 ALGUMAS CONSIDERAÇÕES SOBRE FORMATOS DE METADADOS SIMPLES, ESTRUTURADOS E RICOS.....	152

5.4 METADADOS: a chave para a representação de recursos informativos na web semântica.....	156
6 CONSIDERAÇÕES FINAIS.....	159
REFERÊNCIAS	169
BIBLIOGRAFIA CONSULTADA.....	177

1 INTRODUÇÃO

A nossa sociedade vem se modificando com os avanços tecnológicos dos últimos tempos e com a valorização que a informação vem ganhando a cada dia, transformando-se em foco de um novo paradigma.

Estamos em um momento, denominado por muitos como “a era da informação” ou a “era do conhecimento” (TAKAHASHI, 2000; BORGES, 2000), no qual a informação é considerada matéria-prima para o desenvolvimento social, econômico e cultural e aliada ao uso de tecnologia de informação passou a ser fator importante na construção de uma nova sociedade, também denominada “sociedade da informação” (TAKAHASHI, 2000; BAGGIO, 2000).

Nesse cenário, temos a Internet como uma ferramenta que tem revolucionado todas as áreas do conhecimento pela facilidade de disponibilização de informação e o volume de recursos que cresce a cada dia, e a World Wide Web (WWW), considerada como a maior fonte de informação de diversas áreas têm transformado os processos de geração e sua de informações.

Entretanto, essa crescente quantidade de informações disponibilizadas na rede vem causando certos problemas de busca e recuperação, como por exemplo, a falta de precisão nos recursos informacionais recuperados pelas ferramentas de busca (BARRETO, 1999). Cendón (2001) afirma que nem as melhores ferramentas de busca conseguem cobrir 60% das páginas disponibilizadas na rede e isto ocorre por três motivos: a crescente quantidade de informações disponibilizadas, problemas com indexadores e robôs de busca das ferramentas e falta de tratamento adequado do conteúdo dos recursos informacionais (ausência de representação e linguagem de marcação que expressem melhor o conteúdo dos recursos).

Para Marcondes e Sayão (2001, p. 26) a informação, na sociedade da informação, passa a ser insumo para qualquer atividade e para ela ser útil e relevante tem que estar disponível no momento certo, além disso, “De nada adianta a informação existir, se quem dela necessita não sabe a sua existência ou se ela não puder ser encontrada”.

Sendo assim, tendo como base as afirmações de Barreto (1999), Cendón (2001) e Marcondes e Sayão (2001), vemos que é de extrema importância concentrar esforços na tentativa de encontrar uma solução para esses problemas. O maior desafio atualmente para a comunidade científica refere-se ao tratamento dos recursos informacionais disponibilizados, pois a partir de um bom tratamento (representação da informação) é possível estabelecer técnicas mais eficazes de busca e recuperação das informações na rede.

A literatura aponta para solucionar essa questão o estabelecimento da denominada Web Semântica. Desenvolvida por Tim Berners-Lee (BERNERS-LEE, HENDLER, LASSILA, 2001), a Web Semântica tem sido indicada como um caminho para solucionar a representação dos recursos informacionais na Web, pois visa a proporcionar o acesso automatizado aos recursos informacionais, com base na estruturação e representação dos dados. Além disso, visa a estabelecer o contexto semântico em que o recurso se insere, para que possam ser desenvolvidas técnicas mais eficazes para a recuperação desses recursos informacionais na Web Semântica.

A Web Semântica será estabelecida por meio do trabalho conjunto de várias outras ferramentas tecnológicas, entretanto a base para sua construção está no uso de metadados para a representação dos recursos informacionais.

1.1 DEFINIÇÃO DO PROBLEMA

Nesse contexto, somos atores de um novo cenário, que se caracteriza pelo crescimento exponencial do número de informações disponíveis na Web, que juntamente com as mudanças tecnológicas também nos traz o problema que acabou dando origem a esta pesquisa: a dificuldade na busca e recuperação de recursos informacionais digitais na Web.

Além disso, esses fatores se agravam com a falta de tratamento adequado para a representação desses recursos, fazendo com que os usuários recuperem uma grande quantidade de documentos irrelevantes para suas necessidades.

Portanto, o maior desafio atualmente para a comunidade científica refere-se ao tratamento para que, conseqüentemente, possam ser melhoradas a busca e a recuperação das informações.

1.2 PROPOSIÇÃO

Resolver esse problema se apresenta como um desafio para a comunidade científica e a solução apontada está relacionada com o estabelecimento da Web Semântica, bem como a aplicação de metadados para garantir a representação dos recursos informacionais digitais.

No entanto, para o estabelecimento da Web Semântica, além do uso de Metadados, é necessário que outras ferramentas sejam utilizadas para que possam proporcionar uma otimização na busca e a recuperação entre sistemas ou repositórios de informação.

Essas ferramentas correspondem: ao uso de identificadores, tais como URI e UNICODE, para a identificação mínima dos recursos; as arquiteturas de metadados, para garantir uma maior interoperabilidade entre os dados e metadados distintos; as ontologias das

mais diversas áreas do conhecimento, para garantir uma definição dos conceitos envolvidos na representação; a linguagem de marcação XML, para proporcionar uma melhor estruturação do conteúdo dos recursos informacionais e dos dados e metadados estabelecidos nas outras ferramentas citadas; e por fim, o estabelecimento de regras que irão determinar aos agentes inteligentes a lógica para o entendimento dos dados.

1.3 OBJETIVOS

A Ciência da Informação como uma área que tem como objeto de estudo a informação desde sua geração até o seu uso busca-se analisar o desenvolvimento da Web Semântica e a aplicação de metadados para proporcionar a representação dos recursos informacionais digitais em repositórios informacionais como a Web, com o propósito de atingir os seguintes objetivos:

1.3.1 Objetivo geral

O desenvolvimento e implementação da Web Semântica irá ocorrer a partir do uso intensivo de metadados, sendo assim, esta pesquisa tem como objetivo principal estudar os aspectos que permeiam a construção da Web Semântica em relação com a Ciência da Informação, bem como o uso e aplicação de metadados para a representação de recursos

informativos, já que são considerados fatores principais para a representação da informação na atualidade.

1.3.2 Objetivos específicos

Nesse sentido, os objetivos específicos a serem abordados nesta pesquisa são os seguintes:

- Analisar as ferramentas tecnológicas que compõe a Web Semântica, destacando para a questão do uso das ontologias como responsáveis por garantir a semântica dos dados;
- Identificar as características da linguagem de marcação XML – eXtensible Markup Language, e sua importância para a estruturação dos recursos informativos na Web Semântica;
- Analisar o uso da arquitetura de metadados, em especial a arquitetura RDF – Resource Description Framework, para o estabelecimento da interoperabilidade sintática, estrutural e semântica.
- Analisar as características, o uso e a aplicação dos diferentes tipos de metadados e seus formatos para o estabelecimento da representação de recursos informativos e destacá-los como ferramentas essenciais para a construção da Web Semântica.

1.4 METODOLOGIA

No intuito de construir um conhecimento teórico sobre a Web Semântica, as ferramentas tecnológicas responsáveis pelo seu estabelecimento e principalmente o uso dos metadados para a representação dos recursos informacionais, o presente trabalho caracteriza-se por ser uma pesquisa de análise exploratória e descritiva do tema (CERVO, BREVIAN, 2003), na qual se buscou no referencial teórico publicado as principais questões estabelecidas no objetivo deste trabalho, para que pudessem ser localizadas as contribuições científicas sobre esse assunto. A adoção dessa metodologia permitiu abordar os aspectos mencionados nos objetivos específicos definidos anteriormente e proporcionar a compreensão e concretização do objetivo geral proposto.

Como procedimentos metodológicos seguiu-se os seguintes passos:

1. Levantamento bibliográfico: realizado em nível nacional e internacional em fontes bibliográficas primárias (livros, periódicos, anais de congresso, dissertações, teses e documentos eletrônicos da Internet, entre outros documentos congêneres), secundárias (Base de Dados textuais e referenciais como: Scielo, Web Spirs, Current Contents, Probe, Web of Science, Lisa, Periódicos Capes, Science Direct, entre outras) e terciárias (bibliografias, catálogos coletivos, guias de literatura, diretórios, índices e outros) da Ciência da Informação.
2. Adotou-se como abordagem inicial para a seleção dos documentos, os critérios de pertinência quanto aos assuntos presentes na pesquisa, aos idiomas português, inglês e espanhol e período de publicação limitado aos últimos dez anos, apenas como abordagem inicial, sendo que não houve limitação cronológica para documentos identificados.
3. Após o levantamento bibliográfico e seleção dos materiais foram realizadas as leituras e documentação dos textos selecionados, que proporcionam a criação de uma base teórica para um maior entendimento e definição da solução ao problema de pesquisa e dos

processos de tratamento da informação da informação por metadados, bem como as tecnologias que possibilitam uma recuperação da informação mais precisa pela implementação da Web Semântica.

Assim, esta pesquisa apresenta no presente capítulo, uma introdução ao tema principal da pesquisa, apresenta também a definição do problema e da hipótese para solucioná-lo, os princípios metodológicos que nortearam esta investigação científica, bem como a justificativa para sua realização, conforme visto a seguir.

1.5 JUSTIFICATIVA

O interesse para a realização desta pesquisa surgiu durante a graduação em Biblioteconomia, na participação do programa de Iniciação Científica CNPq/PIBIC e do Grupo de pesquisa novas Tecnologias em Informação com uma pesquisa intitulada “Análise dos padrões de descrição das informações para a organização de documentos eletrônicos: um estudo sobre metadados”. Essa pesquisa, que também originou o trabalho de conclusão de curso, tratava da questão da representação dos recursos em meio eletrônico com o uso de padrões de metadados. Há, portanto, um grande interesse pessoal para o desenvolvimento desta pesquisa por considerá-la uma complementação do tema anteriormente desenvolvido e por ser um assunto de importância atual.

Foi possível perceber durante a realização da pesquisa na Iniciação Científica a importância do uso de padrões de metadados para a estruturação e representação dos recursos disponíveis na rede, bem como sua importância para a recuperação da informação.

Portanto, se considerarmos que cada vez mais haverá informações disponibilizadas na rede e que serão necessários mecanismos que recuperem essas informações de modo mais eficiente, vemos que é de grande importância concentrar esforços em estudos que visam a representação dos recursos para que possa ser melhorada sua busca e recuperação.

Pretende-se com a realização desta pesquisa contribuir para a área da Ciência da Informação no sentido de proporcionar um referencial teórico aos profissionais da área sobre o tema Web Semântica e metadados, pois são apontados hoje como um caminho para proporcionar uma recuperação de modo mais eficiente por meio da representação de recursos informacionais digitais. Assim, a importância da realização desta pesquisa está em acompanhar os avanços da área de Ciência da Informação sobre o tema, a questão da representação dos recursos informacionais e sua recuperação na Web.

Além disso, a importância social na realização desta pesquisa está em fornecer subsídios para que profissionais da área possam desenvolver futuramente técnicas mais eficazes de recuperação da informação na rede, contribuindo para a diminuição das dificuldades de localização, busca, acesso e recuperação das informações.

Os serviços desenvolvidos pelo bibliotecário como, a responsabilidade de organizar, tratar, armazenar, recuperar e disseminar a informação de forma rápida e precisa, são de fundamental importância para o desenvolvimento de estudos relacionados aos metadados e a Web Semântica.

Quanto à organização desta dissertação, além do presente capítulo que trata da Introdução e que aborda questões iniciais como o problema da pesquisa, hipótese, justificativa, relevância social, relevância para a área de Ciência da Informação, objetivos, entre outros, a dissertação está estruturada da seguinte forma:

CAPÍTULO 2 - *WEB SEMÂNTICA: uma nova proposta para a organização e recuperação de recursos informacionais na rede.* É destacado a questão do estabelecimento da Web Semântica, a arquitetura responsável pelo seu estabelecimento, as tecnologias envolvidas que estão presentes nessa arquitetura e em especial as ontologias para o estabelecimento e definição da semântica na Web.

CAPÍTULO 3 - *ARMAZENAMENTO E ESTRUTURAÇÃO DOS RECURSOS INFORMACIONAIS NA WEB: o papel da linguagem XML.* Neste capítulo são tratadas algumas considerações sobre a linguagem SGML e HTML, no intuito de dar suporte para um melhor entendimento sobre os aspectos principais que envolvem a linguagem XML, suas tecnologias associadas para sua utilização, e sua relação com a Web Semântica no estabelecimento de uma melhor estruturação dos dados na rede.

CAPÍTULO 4 - *METADADOS PARA A REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS NA WEB SEMÂNTICA.* São abordados os metadados e formatos de metadados, os conceitos de cada um destes termos, suas características, bem como os formatos que mais se destacam entre as categorias identificadas. E ainda, a importância dos metadados para a representação dos recursos na Web Semântica.

CAPÍTULO 6 - *CONSIDERAÇÕES FINAIS.* São levantadas algumas considerações sobre os aspectos principais que envolvem o estabelecimento da Web Semântica, os metadados para a representação dos recursos informacionais na rede e a Ciência da Informação.

REFERÊNCIAS – Por fim estão apresentadas as referências utilizadas para o desenvolvimento desta pesquisa.

2 WEB SEMÂNTICA: uma nova proposta para a organização e recuperação de recursos informacionais na rede

As novas tecnologias de informação e comunicação, aliadas a crescente valorização da informação e do conhecimento, vêm provocando profundas transformações em nossa sociedade nos últimos tempos.

Inseridos nas diversas áreas, a informação e o conhecimento passaram a adquirir um novo valor e se transformaram em chaves do paradigma em que a tecnologia atua significativamente nos procedimentos de produção, armazenamento, transmissão, acesso, uso, disseminação e recuperação; e conseqüentemente, na organização e tratamento (representação) da informação.

Atualmente estamos presenciando uma mudança significativa, a passagem da sociedade para uma cultura voltada para o uso de tecnologias. Nesse cenário tecnológico a Internet apresenta-se como um mecanismo que modificou todas as áreas do conhecimento, devido a facilidade de disseminar a informação; e a World Wide Web (WWW), como a maior fonte de informações de diversas áreas, pois proporciona o acesso à uma enorme quantidade de recursos informacionais em diversos formatos de armazenamento.

A Web juntamente com a Internet vem expandindo a cada ano, não só em números de usuários, mas também na quantidade de recursos disponibilizados. Assim, presenciamos nos últimos tempos o crescimento exponencial do número de informações disponibilizadas na rede e a tendência é que este número aumente, pois a Internet e a Web continuarão a se consolidar como fonte de informação fundamental em diversas áreas do conhecimento (CENDÓN, 2000).

Criada sem o intuito de ser um espaço organizado, a Internet surgiu como proposta de

um sistema distribuído de comunicação entre computadores para possibilitar a troca de informações. Assim como a Internet, a Web também apresenta certa desorganização, mesmo oferecendo, por meio de seu sistema de hipertexto, interfaces mais amigáveis para a organização, disponibilização e acesso ao crescente repositório de documentos que se tornara a Internet. Conforme apontam Souza e Alvarenga (2004, p. 133)

Embora tenha sido projetada para possibilitar o fácil acesso, intercâmbio e a recuperação de informações, a *Web* foi implementada de forma descentralizada e quase anárquica; cresceu de maneira exponencial e caótica e se apresenta hoje como um imenso repositório de documentos que deixa muito a desejar quando precisamos recuperar aquilo de que temos necessidade.

Esse crescimento dos recursos informacionais disponibilizados na Web trouxe como consequência a dificuldade na sua recuperação. Na tentativa de facilitar a busca, localização e recuperação dos recursos foram sendo criados mecanismos que pudessem auxiliar esta tarefa. Esses mecanismos de busca estão presentes desde antes da popularização da Web e vem se modificando ao longo do tempo para atender a demanda de recuperação dos mais variados tipos de recursos informacionais.

Também chamados de ferramentas de busca ou pesquisadores, os mecanismos de busca são sites especializados em localizar informações na Web e existe atualmente um grande número de ferramentas para este fim (BRANSKI, 2000). Diferenciam-se em diversos aspectos, dentre eles podemos citar a forma de localização, descrição, indexação das páginas, os recursos disponíveis para a busca, a forma de recuperação e apresentação dos resultados, entre outros.

Não é objetivo deste trabalho tratar de modo detalhado o funcionamento das ferramentas de busca, pois isto já foi realizado em outros trabalhos, como por exemplo, o de Santarém Segundo (2004), mas apenas fornecer alguns aspectos que possam explicar a relação das atuais ferramentas de busca com as novas propostas de recuperação e organização de recursos informacionais na rede.

Existem três tipos básicos de ferramentas de busca, a seguir veremos resumidamente cada um deles:

- A. Diretórios: Caracterizam-se por serem ferramentas genéricas composta por categorias temáticas de assuntos amplos. Os recursos informacionais são selecionados, organizados e classificados de forma manual, ou seja, o tratamento dos sites é feito com ajuda de pessoas (CENDÓN, 2001). Cada diretório apresenta um critério para seleção dos sites, mas em geral o interessado envia uma breve descrição do conteúdo de seu site e solicita a inclusão do endereço da página no banco de dados do diretório. Caso seja aceito, o endereço (URL) será classificado na categoria que julgarem mais adequada (BRANSKI, 2000). Apresentam uma base de dados menor quando comparada com os motores de busca, no entanto, a recuperação é um pouco mais precisa.
- B. Motores de Busca: Os motores de busca ou índices surgiram quando a seleção manual de sites na Web se tornou dificultosa devido ao aumento de recursos informacionais na rede. Portanto, a indexação dos sites e a criação de seus bancos de dados é realizado de modo automático, suas bases são extremamente grandes e por isso a busca é feita por meio de palavras-chave. Essas ferramentas não organizam hierarquicamente os recursos, por isso podem apresentar um resultado com menor precisão quando comparado com os diretórios. Colecionam o maior número possível de recursos por meio do uso de softwares chamados robôs.
- C. Metamotores: Também chamados de Multibuscadores ou Metapesquisadores buscam simultaneamente em vários mecanismos de busca e não possuem um banco de dados próprio, obtêm as respostas dos mecanismos pesquisados individualmente e, então, apresentam aos usuários um resultado unificado. São mais indicados para buscas com termos únicos, quando não encontramos muitos resultados em outras ferramentas (BRANSKI, 2000). Geralmente realizam a busca na rede em uma única interface e

impossibilita o acesso às interfaces de refinamento de pesquisa de cada motor de busca. (CENDÓN, 2001).

A diferença básica entre esses três tipos de ferramentas de busca está relacionada com o modo de construção de suas base de dados, modo manual nos diretórios e modo automático nos mecanismos de busca.

Uma desvantagem que podemos destacar da indexação manual dos diretórios está relacionado ao tempo com que esse serviço é feito e o limite de informações processadas diariamente, que é bem menor quando comparada com as ferramentas de busca que utilizam robôs para a construção de suas bases de dados. Entretanto, as ferramentas que se utilizam de robôs também apresentam desvantagens, como a recuperação menos precisa de informações, mesmo possuindo um banco de dados relativamente grande.

Um aspecto apontado por Souza e Alvarenga (2004), está relacionado com a recuperação realizada pelas ferramentas que fazem uso de robôs de busca, em primeiro lugar a recuperação dos recursos informacionais atualmente na Web é feita por meio das palavras-chave contidas no conteúdo dos recursos. Em segundo lugar, não há nenhuma estratégia que seja satisfatória para proporcionar uma melhor indexação dos sites e conseqüentemente uma melhor recuperação desses recursos.

Para a construção das bases de dados das ferramentas de busca dois componentes diferentes, porém complementares, trabalham na busca, localização e recuperação das informações: os robôs e indexadores.

Os robôs de busca também chamados de *spiders*, *crawlers* ou agentes são softwares responsáveis pela busca e localização dos recursos informacionais, eles vasculham a Web utilizando estratégias variadas para se locomoverem de um site a outro. Essas estratégias muitas vezes não são claramente divulgadas, mas geralmente a busca começa pelos sites mais populares e a partir da homepage, vão seguindo os links e adicionando os endereços ao banco

de dados. Usam algoritmos próprios para determinar os links a serem seguidos e voltam aos sites regularmente para verificar as alterações e atualizar o sistema (CENDÓN, 2001; BRANSKI, 2000).

O indexador é responsável por retirar as informações necessárias das páginas para a construção do banco de dados da ferramenta de busca, tais como o endereço da página, títulos, resumos, tamanho do arquivo, o conteúdo integral das páginas, ou somente o título e as primeiras linhas do site, entre outros itens (CENDÓN, 2001; BRASKI, 2000). Os critérios para a indexação das páginas variam de acordo com a ferramenta utilizada, alguns motores de busca indexam cada palavra do texto visível nas páginas, palavras que ocorrem com frequência, palavras e frases mais importantes do título ou cabeçalhos ou nas primeiras linhas do texto etc (CENDÓN, 2001).

Esses critérios de indexação que os motores de busca estabelecem são fundamentais para a recuperação dos recursos informacionais para os usuários, pois o modo como a indexação é feita irá influenciar no resultado da busca. Se o termo de busca não estiver incluído na base de dados da ferramenta, o documento não será encontrado (CENDÓN, 2001).

Nesse caso há interferência de um fator importante, o uso de elementos adicionais, elementos que descrevem mais detalhadamente um recurso. Alguns recursos informacionais não fazem uso desse instrumento e acabam dificultando a sua recuperação que é feita apenas pelos critérios dos sistemas de indexação da ferramenta.

Marcondes e Sayão (2001, p. 26) apontam que nem sempre o resultado de uma busca na Web é satisfatório, na maioria das vezes encontramos muitos recursos que não estão relacionados com o termo de busca. Isso acontece por vários problemas, dentre os principais citam os seguintes:

[...] baixa qualidade da indexação, por ser feita automaticamente, que resulta em grande quantidade de informações recuperadas, a maioria sem relevância

(em termos de recuperação de informação, oferecem alta revocação, mas baixa precisão); cobertura parcial da Internet; as ferramentas de busca não são especializadas; indexam páginas HTML isoladas e não recursos; além disto, grande quantidade de informações disponíveis na Internet estão sob a forma de registros contidos em bases de dados, que ficam assim “escondidas”; estes registros são acessados somente por meio das interfaces destas bases de dados, o que pressupõe uma interação entre um usuário humano com a base de dados e, portanto, ficam inacessíveis aos programas robôs.

Barreto (1999), aponta ainda que um dos principais problemas está relacionado ao mecanismo de indexação que indexa as palavras contidas nos recursos informacionais sem levar em consideração a semântica do contexto onde está inserido. Existem ainda outros problemas citados por essa autora, entretanto, no contexto desta pesquisa esse seria o mais significativo, pois estabelecer o contexto semântico de um recurso é um fator chave para proporcionar uma boa recuperação na rede e atingir o desenvolvimento da Web Semântica.

Mesmo com a variedade de ferramentas disponíveis e apesar de estarem sempre em constante atualização em suas técnicas de busca, muitas vezes as ferramentas atuais não conseguem atender de modo satisfatório seus usuários. Apesar de toda tecnologia algumas limitações ainda ocorrem. Os principais empecilhos estão relacionados com a crescente quantidade de informações disponibilizadas; com as próprias limitações das ferramentas de busca (suas técnicas de busca e indexação dos sites) e com a falta de representação (tratamento) adequado dos recursos informacionais na rede, tanto na representação por uma linguagem de marcação que possibilite melhor visualização do conteúdo do recurso, como também na construção de formas de representação convencionadas pela Biblioteconomia, catalogação e metadados.

Partindo dessas limitações e da necessidade de se estabelecer uma contextualização dos recursos informacionais surge a Web Semântica. Apontada na literatura como um caminho para solucionar os problemas destacados nesta pesquisa, a Web Semântica apresenta um novo modo para organizar os recursos informacionais da Web e de acordo com seus idealizadores possibilitará expressar um maior significado das informações e proporcionar o

desenvolvimento de ferramentas de busca mais eficientes.

2.1 ARQUITETURA DA WEB SEMÂNTICA: o delineamento de uma nova Web

Com o intuito de melhorar a recuperação de recursos em ambientes informacionais como a Web, por exemplo, a proposta da Web Semântica é instituir um maior nível semântico na representação dos recursos informacionais, proporcionando assim, uma maior eficiência aos mecanismos de busca que trabalham com processamento automático de recursos na rede.

Os computadores e robôs de busca não interpretam palavras em um determinado contexto, portanto, não conseguem “entender” o conteúdo significativo de um recurso informacional. Santarém Segundo e Vidotti (2003, p. 3), apontam que,

Os computadores trabalham com processamento lógico, mas não são capazes de fazer associações de significados, diferentemente da mente humana que é capaz de juntar partes de informações dispersas e de estabelecer um novo contexto, identificando o significado das informações dispostas e assimilando um novo conhecimento.

Os seres humanos possuem a capacidade de interpretar, conseguem distinguir o sentido das palavras em um determinado contexto, isto não ocorre nas ferramentas de busca e principalmente nos robôs e indexadores, responsáveis pela localização e extração de informações do conteúdo dos recursos necessárias para a construção da base de dados das ferramentas. Não conseguem distinguir semanticamente o contexto em que a informação está inserida, por este motivo que muitas vezes não recuperamos recursos que correspondem às nossas necessidades.

Partindo desse princípio que a Web Semântica foi idealizada e se apresenta como uma nova forma de tornar conteúdos da Web mais significativos para computadores, trazendo não só uma revolução no modo como os recursos são disponibilizados atualmente, mas também

novas possibilidades para o surgimento de ferramentas de busca mais eficientes.

Com essa nova “organização” dos recursos informacionais disponibilizados na rede proposta pela Web Semântica seria possível implantar, em agentes inteligentes, regras para o raciocínio sobre os dados representados e definidos semanticamente.

Mas então o que vem a ser a Web Semântica? Web Semântica é o nome de um projeto criado por Tim Berners-Lee (BERNERS-LEE, HENDLER, LASSILA, 2001) e liderado pela W3C (World Wide Web Consortium) que pretende embutir inteligência e contexto a Web atual e possibilitar posteriormente uma melhor recuperação e uso da informação (SOUZA, ALVARENGA, 2004).

Trata-se de uma evolução da Web atual que pretende implantar, nos próximos dez anos, uma nova forma de desenvolvimento e utilização da Web tradicional, baseada no uso de tecnologias que proporcionem um maior significado na rede (MARTINS JÚNIOR, 2003; MOURA, 2002b). Muitos estudos definem o que vem a ser Web Semântica, por isso segue abaixo algumas definições encontradas na literatura:

Seu idealizador, Berners-Lee, aponta que “A Web Semântica não é uma Web separada, mas uma extensão da Web atual na qual as informações apresentam significados bem definidos e permite que computadores e pessoas possam trabalhar em cooperação” (BERNERS-LEE, HENDLER, LASSILA, 2001).

Para Palmer (2001) “A Web Semântica é uma rede de informações interligadas de tal modo que possa ser facilmente processada por máquinas, em escala global”.

Faria e Girardi (2002?) definem a Web Semântica como sendo,

[...] uma extensão da Web atual, que introduz uma estrutura e um significado para permitir a evolução de uma rede de documentos para uma rede de dados na qual toda a informação tem um significado bem definido para ser interpretada por computadores e humanos, aumentando assim a capacidade das máquinas de trabalhar em cooperação com as pessoas.

Esteban Villamizar (2002?) faz uma comparação entre a Web atual e a Web Semântica e diz que, a Web atual pode ser considerada como um conjunto de páginas conectadas entre si

e a Web Semântica deverá ser considerada como um conjunto de conceitos interrelacionados.

Codina (2003) define a Web Semântica como sendo,

[...] um conjunto de iniciativas, tecnológicas em sua maior parte, destinadas a criar uma futura World Wide Web na qual os computadores podem processar a informação, isto é, representa-la, gerenciá-la, como se os computadores possuíssem inteligência.

Sendo assim, a Web Semântica propõe-se a estruturar e dar semântica aos dados representados com o intuito de diminuir ou eliminar os problemas de recuperação já mencionados nesta pesquisa. Para isso apresenta uma estrutura que possibilitará a compreensão e gerenciamento do conteúdo dos recursos informacionais, por meio da valorização da semântica destes recursos e de agentes capazes de processar informações e trocar informações com outros programas (BERNERS-LEE, HENDLER, LASSILA, 2001; CUNHA, 2002; MOURA, 2002a).

A partir dessas definições podemos dizer que a Web Semântica seria uma extensão da Web atual que apresentaria recursos informacionais melhor estruturados e representados, ou seja, o conteúdo informacional destes recursos seriam melhor explicitados e definidos semanticamente, formando uma rede de informações conectadas que por meio de ferramentas tecnológicas, tais como os agentes de software, proporcionaria uma melhor recuperação de informação.

De acordo com Berners-Lee, Hendler e Lassila (2001), a Web Semântica trará estrutura para o conteúdo significativo dos recursos informacionais da Web e criará um ambiente onde os agentes de softwares possam realizar atividades sofisticadas como processar e entender os dados solicitados na busca, e assim, proporcionar uma recuperação mais eficiente para os usuários.

Para Moura (2002?a)

A Web Semântica é hoje um dos objetivos a longo prazo da W3C. Deverá se desenvolver num ambiente de acesso inteligente à informação heterogênea e distribuída, através de agentes de softwares. Estes agentes irão mediar e realizar o brokering entre, as necessidades de cada usuário e as fontes de

informação disponíveis, permitindo pesquisas mais acuradas e eficientes.

Sendo assim, o objetivo da Web Semântica, segundo os estudos de Rosa (2002) seria transformar o conteúdo atual da Web num formato que permita, não só humanos, mas também que agentes inteligentes compreendam o significado das informações e possam recuperar e manipular esta informação de modo mais lógico.

Segundo Faria e Girardi (2002?), um dos desafios da Web Semântica,

[...] é criar uma linguagem que seja capaz de expressar ao mesmo tempo o significado dos dados e definir regras para raciocinar sobre os mesmos, de forma a deduzir novos dados e regras e, permitir que regras existentes em sistemas de conhecimento possam ser exportadas para Web.

Para Souza e Alvarenga (2004), o projeto da Web Semântica provê a criação e implantação de padrões tecnológicos para permitir o compartilhamento de informações entre sistemas de informação. Sendo assim, para que a Web Semântica seja implementada, é necessário o trabalho conjunto de várias ferramentas tecnológicas que propiciem uma melhor estruturação e representação dos dados.

Isso pode ser melhor visualizado quando observamos a arquitetura da Web Semântica proposta pelos seus idealizadores. Existem alguns esquemas que demonstram a arquitetura da Web Semântica, entretanto será destacado nesta pesquisa o esquema da W3C complementado pelas explicações de Moura (2002b), conforme pode ser visto na Figura 1 a seguir:

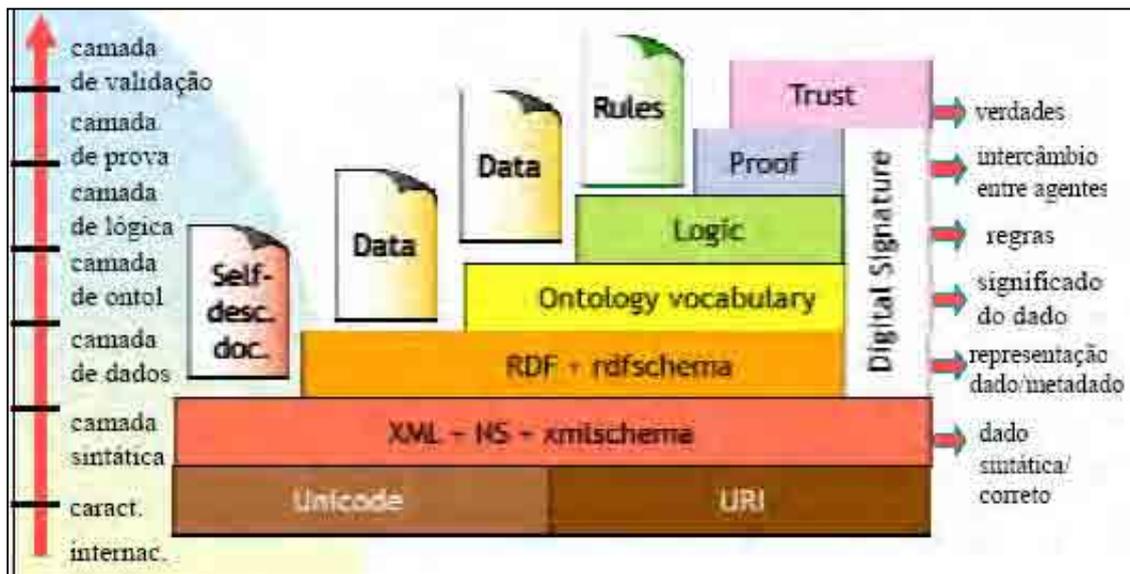


FIGURA 1: Arquitetura da Web Semântica comentada.

FONTE: Moura (2002b).

Como pode ser visto no esquema da Figura 1 a arquitetura da Web Semântica apresenta sete camadas, cada uma com uma ferramenta e tecnologia diferente, conforme a explicação abaixo:

▪ CARACTERÍSTICA INTERNACIONAL

É a camada base da arquitetura da Web Semântica. É composta pela URI (Uniform Resource Identifier) e UNICODE que são padrões para a descrição e estabelecimento de identificadores universais do recurso e códigos internacionais de dados (SANTARÉM SEGUNDO, 2004). Esses dois elementos são responsáveis pelo estabelecimento de uma identificação mínima dos recursos na rede, como por exemplo, a localização de um recurso pela URL (Uniform Resource Locator). Segundo Rosa (2002) uma URI,

[...] estabelece uma forma padrão para a identificação de recursos. [...] Através da utilização de URI faz-se a referência para recursos representados na Web Semântica. No contexto da Internet, o conceito de URI já é bem utilizado. Na Web é utilizado um tipo de URI chamado URL. Através da URL é possível endereçar documentos utilizando protocolos específicos da Internet como *http* e *ftp*.

Já o Unicode é definido mais detalhadamente por Rosa (2002) como sendo,

[...] uma linguagem que define uma forma padrão para a representação de caracteres. Unicode proporciona uma forma única para a representação de um caracter não importando a plataforma, o programa nem a linguagem que está sendo utilizada. A utilização de Unicode na Web Semântica proporciona a capacidade de troca de símbolos de maneira universal, requisito fundamental para o sucesso desta nova proposta de representação de informação na Internet.

▪ CAMADA SINTÁTICA

Composta pela linguagem XML, pelo uso de *namespaces* e pelo XML *Schema*, essa camada é responsável pelo estabelecimento correto da sintaxe de descrição dos dados. Além disso, a linguagem XML proporciona uma melhor estruturação, não só dos recursos informacionais, mas também dos dados e metadados que representam o recurso. Sendo assim, a linguagem XML apresenta-se como fundamental, pois enfoca mais o conteúdo dos recursos e não somente a sua forma de apresentação, possibilitando aos agentes de software uma melhor visualização dos dados. De acordo com Rosa (2002) a XML é,

[...] uma linguagem de marcação de fácil compreensão e legível por humanos e por agentes de software, constitui-se como a base de apoio da Web Semântica, pois proporciona a interoperabilidade entre agentes, além da flexibilidade e extensibilidade necessária para a representação dos dados.

Além disso, "XML proporciona o padrão para a representação das estruturas, na qual todas as linguagens, para expressar significado da Web Semântica, serão baseadas" (ROSA, 2002).

O XML *Schema* é definido por Rosa (2002) como uma ferramenta que,

[...] permite a definição e a descrição de estruturas e de conteúdos de documentos XML. Através dessa linguagem, define-se o formato válido de um documento XML, incluindo quais elementos e atributos são permitidos ou não, quais são as suas localizações, o número de ocorrências de cada elemento e outras características. Ou seja, proporciona mecanismos para a definição de gramáticas para correção de documentos XML.

Já os *namespaces*, segundo Rosa (2002) podem ser considerados como,

[...] um método para qualificar nomes de elementos e atributos usados em documentos XML, através da associação de referências URI. Através desse mecanismo de espaço de nomes, é possível a combinação de documentos com a utilização de vocabulário compartilhado. Através do mecanismo de espaço de nomes definido em XML, é possível compartilhar a reutilizar a

definição de outros esquemas XML sem que haja problemas de colisão de nomes.

Os *namespaces* serão muito utilizados na Web Semântica, tanto nas ontologias como também na arquitetura RDF.

▪ CAMADA DE DADOS

Essa camada está diretamente relacionada com a representação, o processamento e a codificação dos metadados. Para isso estão presentes nessa camada a arquitetura de metadados RDF e o RDF Schema, que são ferramentas responsáveis por expressar significados e promover a interoperabilidade entre metadados e padrões ou formatos de metadados (SANTARÉM SEGUNDO, 2004; MARTINS JÚNIOR, 2003).

Segundo Rosa (2002) o RDF,

[...] é uma linguagem para representação de informação na Web. Trata-se de uma infra-estrutura que fornece a habilidade para codificação, troca e reutilização de metadados. RDF define um modelo de dados para descrição de semântica de dados para o entendimento pelo computador. É o fundamento para o processamento de metadados (informação sobre informação).

Com uma função semelhante ao XML *Schema* tratado no item anterior, o RDF *Schema*,

[...] é uma linguagem que define a estrutura válida para dos documentos RDF. RDF e RDF Schema são recomendações do consórcio W3C que definem o padrão para a representação de metadados. São a base de todas as linguagens para expressar semântica da Web Semântica, devido à adoção pelo consórcio W3C (ROSA, 2002).

É importante destacar que tanto na camada de dados como na camada sintática explicada anteriormente, está implícito o uso de metadados e formatos ou padrões de metadados para promover a representação dos recursos informacionais.

▪ CAMADA DE ONTOLOGIA

Essa camada é responsável pelo estabelecimento do significado dos dados, ou seja,

pelo estabelecimento da semântica dos dados descritos e representados pelos metadados. As ontologias presentes nessa camada estabelecem não só os esquemas ontológicos a serem seguidos por certa comunidade, mas também as definições de significados dos conceitos a serem utilizados para a representação de um recurso. Essa camada é importante, pois além de ter a definição dos significados e semântica dos dados é nela que estão estabelecidos os esquemas classificatórios utilizados pelos agentes de softwares (SANTARÉM SEGUNDO, 2004). As ontologias presentes nessa camada serão tratadas mais detalhadamente ainda neste capítulo.

▪ CAMADA LÓGICA

A potencialidade da Web Semântica pode ser comprovada nessa camada, pois teve como base as camadas responsáveis pela estruturação, representação e estabelecimento semântico dos dados. A camada lógica é responsável por proporcionar uma busca e recuperação mais eficientes devido ao uso de agentes, regras e mecanismos de inferência sobre os dados e metadados. Esse conjunto de regras de inferência utilizados pelos agentes de softwares foram baseadas nas descrições estabelecidas nas camadas mais inferiores e que são utilizadas para relacionar e processar informações (SANTARÉM SEGUNDO, 2004; AFONSO, 2001).

De acordo com Rosa (2002), podemos dizer então que,

A camada de Lógica proporciona a definição de semântica em linguagem formal habilitando a execução de serviços inteligentes. É composta principalmente por regras de inferência, com as quais os agentes poderão se utilizar para relacionar e processar informação.

▪ CAMADA DE PROVA

A camada de prova é responsável pelo intercâmbio entre agentes, para isso, está relacionada com as diversas definições lógicas estabelecidas na camada lógica que serão processadas pelos agentes para a construção da prova. “Uma vez que se constrói um sistema

que segue a lógica definida, podem-se seguir as ligações semânticas para construir a prova” (FERNEDA, 2003, p. 119). De acordo com Rosa (2002),

De posse das regras de inferência da camada imediatamente inferior a esta (camada de prova), os agentes podem ter mais poder para raciocinar sobre conceitos e relacioná-los na camada de ontologia. Esta é a camada na qual pode-se obter explicações (provas) sobre as respostas dadas por agentes que consomem alguma informação com o objetivo de verificar se a dedução foi correta.

▪ CAMADA DE VALIDAÇÃO

A última camada da Web Semântica é responsável pelo estabelecimento de verdades, ou seja, pelo estabelecimento de autenticidade, confiabilidade e validade dos dados na Web Semântica (SANTARÉM SEGUNDO, 2004). Essa camada fornecerá aos agentes que raciocinam sobre os dados a garantia de que a informação ou recursos informacional recuperado é verdadeiro e autêntico. De acordo com Rosa (2002),

A camada de confiança (*Trust*) conjuntamente com a camada de assinatura digital (*digital signature*) proporciona mecanismos para prevenção de inconsistências na Web Semântica. Através de aplicações criadas neste nível, é possível criar agentes que saibam dizer, identificar e validar algum tipo de informação. Trata-se de outra característica importante da Web Semântica e muito importante no ambiente da Internet, na qual blocos de dados encriptados podem ser utilizados para garantir a autenticidade das fontes e a confiabilidade da informação que os agentes consultam.

As camadas pertencentes ao grupo Digital Signature (assinatura digital) já descritas, são tecnologias em aperfeiçoamento, contudo, necessárias para garantir a integridade, validade e autenticidade dos dados utilizados pelos agentes na Web Semântica.

Em seu trabalho, Miller (2001) afirma que as tecnologias que envolvem a implementação da Web Semântica estão sendo estudadas pela W3C e ainda estão em desenvolvimento. Alguns grupos de trabalho da W3C estão unificando tecnologias nas camadas mais baixas da Web Semântica, onde já é possível implementar algumas aplicações. Entretanto, as camadas mais altas, tais como lógica, prova e confiança, exigem mais pesquisas, um maior consenso e a junção de informações de demonstrações experimentais.

Miller (2001) aponta também que quanto mais metadados forem utilizados e quanto

mais ricos em representação forem, haverá uma quantia maior de oportunidades de recuperação e aplicações a serem desenvolvidas nos diversos seguimentos que utilizarão a Web Semântica, como o caso de repositórios informacionais.

É a partir dessas considerações de Miller que será encaminhado este trabalho. Sendo assim, a arquitetura da Web Semântica será tratada de modo mais simplificado conforme expõe o seguinte esquema de Santarém Segundo (2004, p. 113):

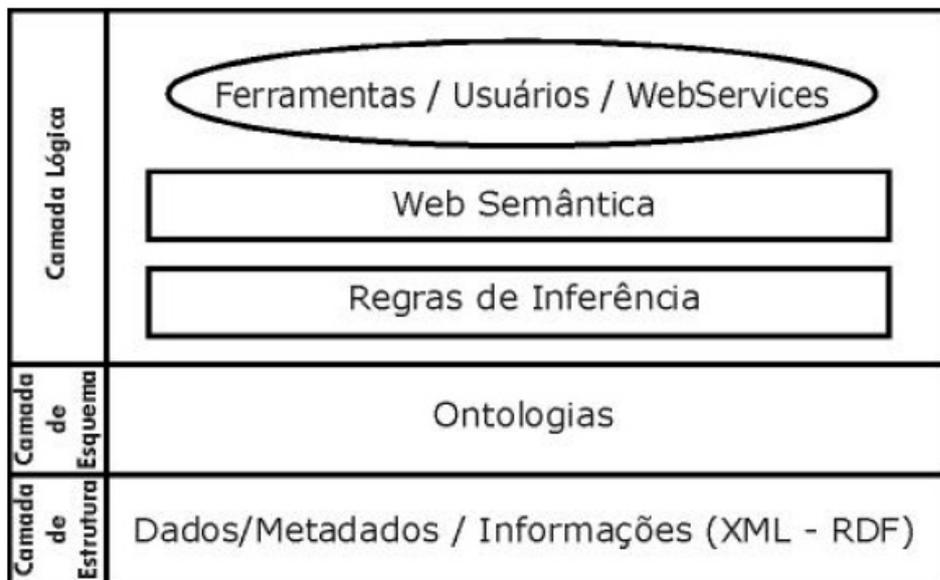


FIGURA 2: Arquitetura mais simples de camadas da Web Semântica.
 FONTE: Santarém Segundo (2004, p. 113).

No esquema da Figura 2 a arquitetura da Web Semântica é apresentada em três camadas principais: a camada de estrutura, composta por dados e metadados, estruturados pela linguagem XML e expressados pela arquitetura RDF; a camada de esquema, composta pelas ontologias que definem o significado semântico dos dados e os esquemas ontológicos; e a camada lógica, compondo a Web Semântica propriamente dita e que apresenta as regras de inferências utilizadas pelos agentes, bem como as ferramentas de busca, os usuários e WebServices.

Como no outro esquema da Figura 1, cada camada requer uma tecnologia responsável

por uma tarefa, entretanto, é preciso destacar que mesmo sendo diferentes e designadas para tarefas distintas, todas estas tecnologias são utilizadas em conjunto para se estabelecer a Web Semântica.

De acordo com Berners-Lee, Hendler e Lassila (2001), o intuito da Web Semântica é fornecer a estrutura necessária para o conteúdo significativo dos recursos informacionais na Web e criar um ambiente onde agentes de softwares possam resolver tarefas sofisticadas de busca e recuperação. Por isso o desafio da Web Semântica é promover um meio para expressar os dados significativos e regras lógicas sobre esses dados para que agentes de software e sistemas possam obter trocas de informações.

Portanto, para que isso ocorra, todas as camadas da Web Semântica devem ser desenvolvidas e implementadas em conjunto, pois, a lógica deve ser bastante adequada para expressar e descrever propriedades complexas para que os agentes de software possam raciocinar sobre elas e não considerá-las contraditórias. Além disso, para o funcionamento correto da Web Semântica é preciso que os computadores tenha acesso aos recursos informacionais devidamente estruturados, representados e definidos semanticamente para que possa haver a inferência e o entendimento por parte dos agentes e conseqüentemente proporcionar a busca e a recuperação automática mais eficientes (BERNERS-LEE, HENDLER, LASSILA, 2001).

Um bom exemplo dado por Codina (2003) sobre a recuperação de recursos informacionais na Web Semântica e que ilustra toda sua potencialidade é o seguinte: se fizermos uma busca sobre o tema “é possível evitar a guerra?” iremos encontrar também “é possível conseguir a paz?”, isto ocorre, pois a semântica da primeira pergunta é a mesma da segunda. Isso não ocorre atualmente, pois quando buscamos “como evitar a guerra” não encontramos recursos informacionais sobre “como conseguir a paz”.

Entretanto, essa situação só irá se realizar quando forem criados programas que “[...]”

coletem o conteúdo da Web de diversas fontes, processem estas informações e compartilhem os resultados com outros programas. Estes programas são os agentes” (SOUZA, ALVARENGA, 2004, p. 137).

Ainda de acordo com Souza e Alvarenga (2004, p. 138),

A efetividade desses agentes de software vai aumentar exponencialmente à medida que mais conteúdo marcado semanticamente e passível de ser ‘entendido’ por máquinas estiver disponível. A Web Semântica promete esta sinergia: mesmo os agentes que não tenham sido expressamente desenhados para trabalhar em conjunto poderão trocar informações entre si, quando houver semântica embutida nestes dados.

E essa semântica embutida nesses dados só existirá a partir da representação dos recursos informacionais e da determinação dos significados que serão estabelecidas nas ontologias.

Sendo assim, o maior enfoque deste trabalho será dado às camadas mais baixas da Web Semântica, onde encontramos um maior avanço nos estudos e pesquisas, além de serem as camadas diretamente ligadas com a representação dos recursos informacionais, com a estruturação, definição semântica e estabelecimento de interoperabilidades.

Como a potencialidade da Web Semântica será vista a partir de conteúdos marcados semanticamente, o próximo item tratado neste capítulo serão as ontologias, que são consideradas fundamentais para o estabelecimento da semântica na rede.

2.2 ONTOLOGIAS: conceitos e definições básicas

As ontologias são ferramentas utilizadas na Web Semântica para o estabelecimento da comunicação entre humanos e agentes de softwares, pois elas determinam o significado e os conceitos que representam os recursos informacionais na rede e são importantes para o

estabelecimento de uma rede de conhecimentos na Web.

De acordo com os vários autores estudados, a palavra ontologia foi herdada da filosofia e significa, segundo Novello (2002?), "[...] uma explicação sistemática da existência". Segundo Moura (2002b) a ontologia "lida com a natureza e organização da realidade". Entretanto, o termo ontologia vem sendo empregado na Web Semântica de modo diferente do significado adotado na filosofia. Ligado a Ciência da Computação, em especial na comunidade de Inteligência Artificial, o termo está relacionado à "tudo que existe deve poder ser representado por um formalismo" (MOURA, 2002b).

De acordo com os estudos de Almeida e Bax (2003), vários autores definem e explicam o que é ontologia. Entretanto, ainda não há um consenso para a definição do termo, pois apresenta uma pluralidade discursiva que se altera de acordo com a comunidade onde o termo está sendo aplicado. Por esse motivo, serão destacadas aqui as definições que mais se adequam a esta pesquisa.

De acordo com Almeida e Bax (2003, p. 08),

Historicamente o termo ontologia tem origem no grego "*ontos*", ser, e "*logos*", palavra. O termo original é a palavra aristotélica "*categoria*", que pode ser usada para classificar alguma coisa. Aristóteles apresenta categorias que servem de base para classificar qualquer entidade e introduz ainda o termo "*differentia*" para propriedades que distinguem diferentes espécies do mesmo gênero. A conhecida técnica de herança é o processo de mesclar *differentias* definindo categorias por gênero.

Para Sowa (1999) "O assunto ontologia é um estudo de categorias de coisas existentes em um determinado domínio. O produto de tal estudo, chamado ontologia, é um catálogo de tipos de coisas existente em um domínio de interesse". Esse catálogo de tipo de coisas é apresentado em categorias em um mesmo domínio.

De acordo com Gruber (1996) uma ontologia é uma especificação explícita de uma conceitualização, na qual apresenta definições que se associam aos nomes de entidades no domínio em que se insere. Essas definições estão relacionadas, por exemplo, a classes, relações, funções e axiomas formais que restringem a interpretação. Sendo assim, uma

ontologia seria uma declaração de uma teoria lógica.

Nessa definição encontra-se relacionado o termo conceitualização que, de acordo com Almeida e Bax (2003), seria o conjunto de relacionamentos determinado por uma coleção de objetos, conceitos e outras entidades existentes em um domínio para formar uma rede conceitual. Outro termo que está relacionado com a definição de ontologia seria relação intencional, que pode ser explicado como "[...] uma lista de características do conceito" (ALMEIDA; BAX, 2003, p. 08).

Guarino (1996) aponta que as ontologias compartilham conceitos que estão especificados na forma de vocabulários que determinam o sentido para as palavras em um determinado domínio, formando assim, uma teoria lógica na qual os agentes de software terão acesso no momento da inferência dos dados. Sendo assim, podemos dizer que as ontologias definem uma linguagem por meio de um conjunto de termos para ser utilizada na formulação de consultas (ALMEIDA, BAX, 2003).

Os conceitos apresentados neste capítulo servirão de base para entendermos o significado das ontologias. Entretanto, o conceito adotado para esta pesquisa será o de Borst (1997, apud ALMEIDA, BAX, 2003, p. 09), que afirma que, “Uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”. Almeida e Bax fazem uma importante explicação,

Nessa definição, “formal” significa legível para computadores; “especificação explícita” diz respeito a conceitos, propriedades, relações, funções, restrições, axiomas, explicitamente definidos; “compartilhado” quer dizer conhecimento consensual; e “conceitualização” diz respeito a um modelo abstrato de algum fenômeno do mundo real (ALMEIDA, BAX, 2003, p. 09).

Unindo a definição de Borst e as explicações de Almeida e Bax, podemos dizer que para esta pesquisa ontologia seria uma especificação formal e explícita de conceitos com suas propriedades, funções, valores e relações, legíveis por máquina e que podem ser compartilhados por uma determinada comunidade na qual esta ontologia esteja sendo adotada.

Podemos dizer ainda que as ontologias definem as relações entre conceitos e estabelecem regras lógicas de raciocínio sobre eles, proporcionando um entendimento e uma comunicação entre pessoas e também define o conteúdo processado pelos agentes de software através de terminologias consensuais. (ROSA, 2002).

Assim, para a área de tecnologia, em especial para a comunidade de pesquisadores da Web, podemos dizer que as ontologias, usadas muitas vezes no plural, vem sendo empregadas para designar a determinação de conceitos e suas relações em um mesmo domínio (MARINO, 2001).

Complementando a definição de Borst (1997) com as explicações de Souza e Alvarenga (2004), podemos dizer então que as ontologias são vocabulários que especificam e definem conceitos, entidades, classes, propriedades, funções, valores. Ou seja, na Web as ontologias definem o significado dos dados que representam os recursos informacionais e suas relações com outros recursos e conceitos em um mesmo domínio do conhecimento.

Essas características se assemelham aos tesouros utilizados à décadas na Biblioteconomia, para a definição de vocabulários controlados, normalização e padronização dos termos adotados em uma comunidade para a definição de conceitos.

O propósito de uma ontologia é possibilitar o compartilhamento de informações semânticas formando um vocabulário comum a certa comunidade. Para Souza e Alvarenga (2004, p. 137), "O objetivo de sua construção é a necessidade de um vocabulário compartilhado para se trocarem informações entre os membros de uma comunidade, sejam eles humanos ou agentes inteligentes".

Inserida em uma determinada comunidade, a ontologia irá representar semanticamente os recursos pertencentes a esse domínio e fornecendo a base para se estabelecer a interoperabilidade semântica. Do ponto de vista da Web Semântica o significado das ontologias continua sendo o mesmo definido por Borsot (1997), e podemos verificar isto na

afirmação de Santarém Segundo (2004), que aponta que as ontologias na Web Semântica estabelecem uma espécie de relação entre categorias de conceitos e definições de um domínio particular, possibilitando um entendimento e compartilhamento do conhecimento entre pessoas e programas de aplicações.

Novello (2002?) aponta que o uso de ontologias na Web Semântica permite integrar sistemas inteligentes no nível do conhecimento. Em outras palavras, permite estabelecer o compartilhamento do conhecimento, representado semanticamente segundo os critérios temáticos estabelecidos em uma comunidade, para que possa haver a interoperabilidade em nível semântico (ontológico). O uso de ontologias na Web, segundo Moura (2002a),

[...] permite que agentes de softwares compreendam a semântica embutida nas definições e vocabulários especificados com respeito a um domínio, sem ambigüidades, viabilizando o intercâmbio de informações através de consultas.

Sendo assim, para um melhor entendimento sobre as potencialidades das ontologias serão tratados no próximo item os seus tipos e as características que tem o intuito de proporcionar o estabelecimento de semântica na rede.

2.3 TIPOS E CARACTERÍSTICAS DAS ONTOLOGIAS

As ontologias são criadas em várias comunidades de interesse, por este motivo são diferentes uma das outras, mas possuem algumas características e componentes básicos em grande parte de sua estrutura. De acordo com Souza e Alvarenga (2004, p. 09),

Os componentes básicos de uma ontologia são classes (organizada em uma taxonomia), relações (representam o tipo de intenção entre os conceitos de um domínio), axiomas (usados para modelar sentenças sempre verdadeiras) e instâncias (utilizadas para representar elementos específicos, ou seja, os próprios dados).

Além desses componentes básicos, os autores Tello (2002?) e Esteban Villamizar (2002?) destacam outros componentes que auxiliam na representação do conhecimento e que são os seguintes:

- 1 Conceitos: são idéias básicas que se pretende formalizar, ou seja, conceitos que vão definir um certo objeto.
- 2 Relações: representam as conexões, vínculos entre conceitos pertencentes a um domínio. Essas relações vão formar a taxonomia do domínio, onde são determinadas: classes, sub-classes, partes etc.
- 3 Funções: são os tipos de relações que se estabelecem entre vários elementos de uma ontologia, que no geral cumprem o mesmo papel, ou seja, desempenham a mesma função.
- 4 Instâncias: utilizadas para representar objetos por meio de um conceito em um determinado ambiente ou domínio.
- 5 Axiomas: são teoremas, regras ou afirmações que declaram as relações que os elementos de uma ontologia devem cumprir. Eles permitem inferir conhecimento que não estão indicados nas taxonomias de uma ontologia.

Para Bézivin (1998), “Uma ontologia define o que deveria ser extraído de um sistema para construir um determinado modelo deste sistema”. Para isso, possui as seguintes propriedades:

- 1 Compartilhamento: “[...] significa que um acordo deve existir entre diferentes agentes baseado no acordo de ontologias comuns, isto é, devem ter o mesmo entendimento sobre um dado conceito”, em outras palavras, esta propriedade é baseada no entendimento sobre os conceitos que definem um recurso e que serve para a comunicação entre agentes de softwares dos sistemas, um acordo que define o mesmo conceito de um recurso

(BÉZIVIN, 1998).

2 Filtragem:

está ligada a abstração, onde consideram-se modelos de abstração. Esses modelos, por definição, levam em consideração somente parte da realidade, e a vantagem de utilização está na habilidade de deixar de lado muitas características indesejáveis. Uma ontologia define o que deveria ser extraído de um sistema de forma a constituir um determinado modelo desse sistema (BÉZIVIN, 1998).

Em resumo esses modelos de abstração em uma ontologia definem o que deveria ser extraído de um sistema.

Além dos componentes e propriedades, as ontologias podem conter também informações de naturezas distintas, Moura (2002b) aponta que essa natureza pode ser de três tipos, conforme descritas abaixo:

- 1 Informações de natureza terminológica: que se caracterizam pelo conjunto básico de conceitos e relações.
- 2 Informações de natureza assertiva: que se caracterizam por conter conjuntos de axiomas assertivas aplicadas em conceitos e relações, ou seja, axiomas que definem regras.
- 3 Informações de natureza pragmática: que se caracteriza por uma camada de ferramentas que contém um conjunto de informações pragmáticas que não se enquadram nos dois tipos acima, informações pragmáticas possibilitam uma interpretação de uma determinada informação em um contexto.

Almeida e Bax (2003) e Jones, Bench-Capon e Visser (1998) apontam que as abordagens para a construção de ontologias são ainda mais artesanais do que científicas, pois não há uma proposta unificada entre as comunidades que as constroem. Diante disso são citados no trabalho de Almeida e Bax (2003) várias metodologias, ferramentas, linguagens e métodos de avaliação que podem ser utilizados na construção de ontologias. De modo resumido esses

itens se caracterizam da seguinte forma:

- 1 Metodologias: são desenvolvidas com o objetivo de sistematizar e unificar os diversos métodos existentes para a construção e a manipulação de ontologias. Essas metodologias estabelecem regras para a construção de ontologias individuais e em grupo, para o aprendizado sobre estruturas de outras ontologias e para a integração de ontologias variadas.
- 2 Ferramentas para a construção de ontologias: a construção de ontologias trata-se de uma tarefa dispendiosa e que por isso critérios devem ser bem definidos.
- 3 Linguagens para a construção de ontologias: as linguagens necessitam de critérios que atendam diversos aspectos, tais como operadores axiomas, declarações etc, que auxiliem na construção das ontologias.
- 4 Métodos de avaliação para ontologias: para avaliar as ontologias são necessários alguns critérios sobre os conceitos e definições que compõem as ontologias, mecanismos de integração, formalismo de representação do conhecimento, avaliação técnica etc. Ainda são critérios de avaliação os seguintes itens destacados por Almeida e Bax (2003, p. 13):
 - Verificar a estrutura ou arquitetura da ontologia: as definições são construídas seguindo os critérios de projeto?
 - Verificar a sintaxe das definições: existem estruturas ou palavras-chave sintaticamente incorretas nas definições?
 - Verificar o conteúdo das definições: o que a ontologia define ou não? O que define incorretamente? O que pode ser inferido e o que não pode?

Moura (2002a) afirma que no contexto da Web o uso de ontologias é crucial, pois “[...] permite que agentes de *software* compreendam a semântica embutida nas definições e vocabulários especificados com respeito a um domínio, sem ambigüidades, viabilizando o intercâmbio de informações através de consultas”.

A criação de ontologias na Web ocorrerá com o uso de linguagens para seu desenvolvimento e seu uso depende do domínio e a comunidade onde a ontologia está sendo criada. Moura (2002b) destaca que as linguagens de ontologias para Web irão atuar em

conjunto com RDF, XML, Topic Maps. Algumas dessas linguagens são: XOL (XML - based Ontology Exchange Language), SHOE (Simple HTML Ontology Extensions), OIL (Ontology Inference Layer), DAML+OIL (DARPA Agent Markup Language + OIL - é uma linguagem para a representação de ontologias que possui embutidos RDF/RDFS) e OWL (Web Ontology Language). Contudo, não é objetivo tratar aqui dessas linguagens, mas apenas citá-las como ferramentas importantes para o estabelecimento de semântica na rede.

Uma outra questão relacionada às ontologias refere-se aos Topic Maps, que segundo Ahmed et al (2001) citado por Moura (2002b) é uma abordagem que tem como objetivo,

[...] identificar assuntos de interesse de uma área de conhecimento e construir um mapa de tópicos, onde um tópico corresponde a um assunto. Esse mapa é então enriquecido com outros tópicos, ocorrências e associações, formando uma imensa rede de conhecimento.

Segundo Moura (2002b) essa tecnologia apresenta maior capacidade de representar semântica e é utilizada para descrever estruturas de conhecimento, índices eletrônicos, esquemas de classificação, encontrar informações na Web, entre outras e a linguagem para a criação dos Topic Maps é a XTM - XML Topic Maps.

Diante da necessidade de se estabelecer maior semântica na Web, vários projetos estão sendo realizados em várias áreas do conhecimento utilizando ontologias. Na área de recuperação da informação Almeida e Bax (2003, p. 11), apontam que os projetos adotam as ontologias para serem utilizadas em projetos de domínios, ou seja, em comunidades de interesses específicos e bem variados. De modo geral esses projetos utilizam ontologias para: formalizar o conhecimento de uma comunidade, ou seja, representam formalmente o conteúdo para que este possa ser compartilhado e recuperado pelos usuários desta comunidade.

Além das características apontadas até agora sobre as ontologias, podemos dizer que elas podem ser categorizadas de acordo com seus tipos. Conforme os estudos de Almeida e Bax (2003, p. 10), essas categorias estão relacionadas ao nível, grau de formalismo, aplicação, estrutura, conteúdo e podem ser dos seguintes tipos:

1) Quanto ao nível

- a) ontologias de domínio: "Reutilizáveis no domínio, fornecem vocabulário sobre conceitos, seus relacionamentos, sobre atividades e regras que os governam".
- b) ontologias de tarefa: "Fornecem um vocabulário sistematizado de termos, especificando tarefas que podem ou não estar no mesmo domínio".
- c) ontologias gerais: "Incluem um vocabulário relacionado a coisas, eventos, tempo, espaço, casualidade, comportamento, funções etc".

2) Quanto ao grau de formalismo

- a) ontologias altamente informais: "Expressa livremente em linguagem natural".
- b) ontologias semi-informais: "Expressa em linguagem natural de forma restrita e estruturada".
- c) ontologias semiformais: "Expressa em uma linguagem artificial definida formalmente".
- d) ontologias rigorosamente formal: "Os termos são definidos com semântica formal, teoremas e provas".

3) Quanto à aplicação

- a) ontologias de autoria neutra: "Um aplicativo é escrito em uma única língua e depois convertido para uso em diversos sistemas, reutilizando-se as informações".
- b) ontologias como especificação: "Cria-se uma ontologia para um domínio, a qual é usada para documentação e manutenção no desenvolvimento de *softwares*".
- c) ontologias de acesso comum à informação: "Quando o vocabulário é inacessível, a ontologia torna a informação inteligível, proporcionando conhecimento compartilhado".

dos termos".

4) Quanto à estrutura

a) ontologias de alto nível: "Descrevem conceitos gerais relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objetivo, evento, ação etc.) os quais são independentes do problema ou domínio".

b) ontologias de domínio: "Descrevem o vocabulário relacionado a um domínio, como, por exemplo, medicina ou automóveis".

c) ontologias de tarefa: "Descrevem uma tarefa ou atividade, como, por exemplo, diagnósticos ou compras, mediante inserção de termos especializados na ontologia".

5) Quanto ao conteúdo

a) ontologias terminológicas: "Especificam termos que serão usados para representar o conhecimento em um domínio (por exemplo, os léxicos)".

b) ontologias de informação: "Especificam a estrutura de registros de bancos de dados (por exemplo, os esquemas de bancos de dados)".

c) ontologias de modelagem do conhecimento: "Especificam conceitualizações do conhecimento, têm uma estrutura interna semanticamente rica e são refinadas para uso no domínio do conhecimento que descrevem".

d) ontologias de aplicação: "Contêm as definições necessárias para modelar o conhecimento em uma aplicação".

e) ontologias de domínio: "Expressam conceitualizações que são específicas para um determinado domínio do conhecimento".

f) ontologias genéricas: "Similares às ontologias de domínio, mas os conceitos que as definem são considerados genéricos e comuns a vários campos".

g) ontologias de representação: "Explicam as conceitualizações que estão por trás dos formalismos de representação do conhecimento".

A categorização apresentada por Almeida e Bax (2003, p. 10) está relacionada com as funções dos diferentes tipos de ontologias ou os tipos de conhecimentos que representam e após destacá-las podemos verificar que na área de Biblioteconomia algumas ferramentas que fazem parte do dia a dia do bibliotecário podem ser consideradas como esquemas ontológicos.

Isso ocorre, pois não existe representação da informação sem uma estrutura ontológica, pois para se estabelecer conceitos semânticos, que são regidos pelas estruturas ontológicas, é preciso haver previamente uma forma de representação que também siga uma estrutura ontológica para proporcionar a representação da informação, para posteriormente estabelecer a semântica.

Entre as ferramentas utilizadas pelo bibliotecário podemos citar: o código de catalogação AACR2 e o formato MARC para a catalogação de recursos informacionais e os Tesouros para a definição de conceitos de um determinado domínio. A seguir apresentamos uma breve explicação de cada ferramenta e em quais estruturas ontológicas se apresentam, conforme as categorias estudadas:

A. **AACR2**: O Anglo American Cataloguing Rules em sua segunda edição, constitui-se como um conjunto de regras e normas para o estabelecimento de uma padronização na representação de diversos recursos informacionais. Caracteriza-se por ser um código abrangente e detalhado, e devido a sua aceitação passou a ser utilizado no ensino de catalogação nos cursos de graduação em biblioteconomia brasileiros; além de ser considerado como um código internacional para a construção de formas de representação bibliográfica. O AACR2 comporta a descrição de qualquer tipo de informação

independente do suporte, caracteriza-se por possuir um formalismo em sua estrutura de representação, pois estabelece, por meio de suas regras, uma relação semântica entre os elementos descritos, além disto, apresenta uma estrutura coerente, lógica e de fácil memorização. De acordo com a tipologia de Almeida e Bax (2003), podemos considerar que o AACR2 pertence aos seguintes tipos de ontologias:

- 1) Quanto ao nível
 - ontologias de domínio: "Reutilizáveis no domínio, fornecem vocabulário sobre conceitos, seus relacionamentos, sobre atividades e regras que os governam".
- 2) Quanto ao grau de formalismo
 - ontologias semiformais: "Expressa em uma linguagem artificial definida formalmente".
- 3) Quanto à aplicação
 - ontologias de autoria neutra: "Um aplicativo é escrito em uma única língua e depois convertido para uso em diversos sistemas, reutilizando-se as informações".
 - ontologias de acesso comum à informação: "Quando o vocabulário é inacessível, a ontologia torna a informação inteligível, proporcionando conhecimento compartilhado dos termos".
- 4) Quanto a estrutura
 - ontologias de alto nível: "Descrevem conceitos gerais relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objetivo, evento, ação etc.) os quais são independentes do problema ou domínio".
- 5) Quanto ao conteúdo
 - ontologias de informação: "Especificam a estrutura de registros de

bancos de dados (por exemplo, os esquemas de bancos de dados)".

B. **MARC**: O formato MARC – Machine Readable Cataloging ou catalogação legível por computador, foi desenvolvido na década de 60 pela LC - Library of Congress, para possibilitar a catalogação e intercâmbio de registros bibliográficos legíveis por máquina. Atualmente denominado MARC 21, este formato compreende a orientação para a estrutura de formas de representação de diversos recursos informacionais, inclusive recursos disponíveis em meio eletrônico. De acordo com Ferreira (2002) o Formato MARC 21 é um padrão amplamente usado na área biblioteconômica para estabelecer a estrutura de representação, importação e exportação de dados bibliográficos. A versão mais atual desse formato está relacionada a linguagem XML, sendo denominado MARC XML. De acordo com Ferreira (2002) geralmente o conteúdo dos elementos que compõem o registro MARC21 é definido por regras e normas externas, como por exemplo, o AACR2. Por esse motivo, apresenta uma estrutura de representação lógica e coerente, que estabelece uma relação semântica entre os elementos descritos, assim como ocorre com o AACR2. As categorias e tipos de ontologias do Formato MARC21 são as seguintes:

- 1) Quanto ao nível
 - ontologias de domínio: "Reutilizáveis no domínio, fornecem vocabulário sobre conceitos, seus relacionamentos, sobre atividades e regras que os governam".
- 2) Quanto ao grau de formalismo
 - ontologias semiformais: "Expressa em uma linguagem artificial definida formalmente".
- 3) Quanto à aplicação

- ontologias de autoria neutra: "Um aplicativo é escrito em uma única língua e depois convertido para uso em diversos sistemas, reutilizando-se as informações".
- ontologias de acesso comum à informação: "Quando o vocabulário é inacessível, a ontologia torna a informação inteligível, proporcionando conhecimento compartilhado dos termos".

4) Quanto a estrutura

- ontologias de alto nível: "Descrevem conceitos gerais relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objetivo, evento, ação etc.) os quais são independentes do problema ou domínio".

5) Quanto ao conteúdo

- ontologias de informação: "Especificam a estrutura de registros de bancos de dados (por exemplo, os esquemas de bancos de dados)".

C. **Tesauros**: Definidos por Dodebei (2002, p. 59),

[...] os tesauros reúnem conceitos que embora sejam representados por símbolos lingüísticos, têm restrito o significado. [...] para cada conceito só pode existir uma representação simbólica, designada de "termo" ou "descriptor".

Em outras palavras, as terminologias designam o significado dos conceitos individuais utilizados na representação. Os tesauros são controlados por normas internacionais e apresentam classes e relações (generalização, especialização, relação partitiva, instância, relacionamento e associações), tais como as ontologias. Os tesauros são controlados por normas internacionais e são destinados a proporcionar somente a representação semântica do conteúdo dos recursos informacionais. Mas também possui uma estrutura lógica, coerente, pois além de definir significados para os conceitos, estabelece a relação semântica entre eles. Assim, com base nas categorias e tipos de ontologias destacados por Almeida e Bax (2003, p.

10) podemos dizer que os tesouros se enquadram nas seguintes:

1) Quanto ao nível

- ontologias de domínio: "Reutilizáveis no domínio, fornecem vocabulário sobre conceitos, seus relacionamentos, sobre atividades e regras que os governam".

2) Quanto ao grau de formalismo

- ontologias semiformais: "Expressa em uma linguagem artificial definida formalmente".
- ontologias rigorosamente formal: "Os termos são definidos com semântica formal, teoremas e provas".

3) Quanto à aplicação

- ontologias de autoria neutra: "Um aplicativo é escrito em uma única língua e depois convertido para uso em diversos sistemas, reutilizando-se as informações".
- ontologias de acesso comum à informação: "Quando o vocabulário é inacessível, a ontologia torna a informação inteligível, proporcionando conhecimento compartilhado dos termos".

4) Quanto à estrutura

- ontologias de alto nível: "Descrevem conceitos gerais relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objetivo, evento, ação etc.) os quais são independentes do problema ou domínio".
- ontologias de domínio: "Descrevem o vocabulário relacionado a um domínio, como, por exemplo, medicina ou automóveis".

5) Quanto ao conteúdo

- ontologias rigorosamente formal: "Os termos são definidos com semântica

formal, teoremas e provas".

- ontologias terminológicas: "Especificam termos que serão usados para representar o conhecimento em um domínio (por exemplo, os léxicos)".
- ontologias de informação: "Especificam a estrutura de registros de bancos de dados (por exemplo, os esquemas de bancos de dados)".
- ontologias de modelagem do conhecimento: "Especificam conceitualizações do conhecimento, têm uma estrutura interna semanticamente rica e são refinadas para uso no domínio do conhecimento que descrevem".
- ontologias de domínio: "Expressam conceitualizações que são específicas para um determinado domínio do conhecimento".
- ontologias genéricas: "Similares às ontologias de domínio, mas os conceitos que as definem são considerados genéricos e comuns a vários campos".
- ontologias de representação: "Explicam as conceitualizações que estão por trás dos formalismos de representação do conhecimento".

Como pode ser visto, a categorização de Almeida e Bax (2003) nos mostra que as ferramentas responsáveis por estabelecer representações dos diversos recursos informacionais na área de biblioteconomia, seguem esquemas ontológicos para possibilitar em suas representações o estabelecimento de semântica.

2.4 IMPORTÂNCIA DAS ONTOLOGIAS PARA A DEFINIÇÃO DE CONCEITOS SEMÂNTICOS NA REDE DE CONHECIMENTOS

A evolução da tecnologia acarretou a descoberta de novos meios de organizar e estruturar dados e informações na Web, conseqüentemente estes novos meios necessitam de novas ferramentas.

De acordo com Santarém Segundo (2004) os sistemas atuais de representação do conhecimento são centralizados em uma comunidade ou domínio, exigindo que todos desta comunidade compartilhem as mesmas definições de conceitos. Entretanto, com a globalização da informação, muitas dessas comunidades passaram a ter a necessidade de compartilhamento de seus dados e informações e para que isto ocorresse era preciso que essas definições também fossem compartilhadas com outras comunidades. Com isso, a necessidade de uma ferramenta que proporcione uma representação precisa dos dados por meio de conceitos semânticos é imprescindível.

Essas ferramentas, denominadas ontologias, representam o conhecimento explícito por meio de uma conceitualização, ou seja, o conhecimento registrado e são consideradas por muitos autores como um dos suportes para o estabelecimento da Web Semântica, pois unida às ferramentas apresentadas em suas camadas e principalmente ao uso de metadados, irá possibilitar a interoperabilidade semântica, o estabelecimento de uma rede de conhecimentos e conseqüentemente uma melhor recuperação da informação na Web.

Moura (2002a) afirma que "A vantagem de uma ontologia é de se lidar com conceitos, representando-os formalmente, e de se livrar de problemas inerentes ao vocabulário da linguagem natural tais como homonímia, sinonímia, metonímia, etc".

Em outras palavras, podemos dizer que as ontologias definem formalmente os conceitos a serem utilizados em uma comunidade, evitando problemas com palavras que possuem a mesma forma escrita e mesma pronúncia, mas que apresentam significado diferente; resolve problemas de palavras sinônimas, bem como estabelece também a relação

entre essas palavras; soluciona problemas de palavras usadas fora de seu contexto semântico, devido à definição formal de conceitos.

Conforme aponta Tello (2002?) para que possamos estabelecer a Web Semântica, é preciso que o conhecimento esteja representado de forma a não haver ambigüidades, ou seja, de forma que seja legível pelos agentes de softwares responsáveis pela busca e recuperação das informações.

Moura (2002a) afirma que no contexto da Web, em especial da Web semântica, o uso de ontologias é crucial, pois,

[...] permite que agentes de software compreendam a semântica embutida nas definições e vocabulário especificados com respeito a um domínio, sem ambigüidades, viabilizando o intercâmbio de informações através de consultas.

Podemos dizer então, que as ontologias proporcionam um caminho para representar o conhecimento em nível semântico dos recursos disponibilizados na rede e possibilitam, entre outras coisas, uma recuperação da informação com maior significado, devido ao estabelecimento formal dos conceitos; e possibilita a interoperabilidade em nível semântico de recursos e informações na rede.

Em outro estudo, Moura (2002b) ressalta que as ontologias podem ser consideradas como uma ferramenta que proporcionará a comunicação entre humanos e máquinas (ou agentes), comunicação esta tão necessária para o estabelecimento da Web Semântica.

Entretanto, é preciso que os agentes de softwares tenham acesso não somente às definições estabelecidas nas ontologias, mas também a uma coleção de recursos informacionais devidamente estruturadas e representadas. Por isso, a importância do uso das outras ferramentas e tecnologias presentes nas demais camadas da Web Semântica.

Sendo assim, será abordado nesta pesquisa não só a camada responsável pela definição do significado dos dados, mas também as ferramentas presentes na camada esquema (metadados, linguagem XML e arquitetura RDF), que são responsáveis pela estruturação e

representação dos recursos informacionais na Web Semântica. Nesse sentido, o próximo capítulo tratará da questão da linguagem XML.

3 ARMAZENAMENTO E ESTRUTURAÇÃO DOS RECURSOS INFORMACIONAIS NA WEB: o papel da linguagem XML

Nos primórdios da humanidade todo o conhecimento humano era transmitido por meio da comunicação oral. Porém, perpetuar a memória da humanidade por meio de registros do conhecimento também são características inerentes aos seres humanos, tanto quanto a oralidade. Na busca por uma forma de registrar informações que eram passadas de geração para geração, a humanidade desenvolveu vários tipos de escrita, suporte e técnicas até chegarmos ao que conhecemos hoje.

Com o desenvolvimento da imprensa por Gutenberg houve o que podemos chamar de primeira revolução da informação, pois a invenção dos tipos móveis possibilitou um aumento significativo no número de informações publicadas em papel.

Hoje podemos dizer que estamos presenciando uma segunda revolução da informação, tão importante como a que ocorreu na era de Gutenberg com a invenção da imprensa, graças ao desenvolvimento das tecnologias de informação e comunicação, em especial de informática, que possibilitaram o acesso a um número muito grande de informações.

O acesso às informações em meio eletrônico já é uma realidade no cotidiano de muitas pessoas, mas no início da era dos computadores essa nova tecnologia foi utilizada principalmente para processamento de dados e cálculos matemáticos. Somente com o passar do tempo e com o aperfeiçoamento da tecnologia de informática foi possível utilizar os computadores para armazenar, recuperar e intercambiar informações em grande escala.

Essa realidade tornou-se possível graças a implementação da Internet, o desenvolvimento da World Wide Web (WWW) e o uso de técnicas como as linguagens de

marcação que possibilitaram o acesso a grandes quantidades de informações armazenadas em meio eletrônico.

Muitas pessoas desconhecem como um documento é criado, armazenado e acessado na WWW. Na realidade não é necessário saber disto para fazer uso das informações disponíveis, mas conhecer o modo como os documentos são construídos e armazenados é importante para o estabelecimento e desenvolvimento de padrões que possibilitem uma melhor recuperação dessas informações nas redes de comunicação.

Para que as informações pudessem estar disponíveis e para que houvesse armazenamento, recuperação e intercâmbio de recursos na rede era necessário o uso de padrões de identificação para a criação destes documentos em meio eletrônico. Nesse sentido, foram sendo criadas as chamadas linguagens de marcação.

Porém, é importante lembrar que historicamente a palavra marcação já era utilizada fora do meio eletrônico para indicar destaque no texto, como negrito e sublinhado por desenhistas e datilógrafos, mostrando qual parte do texto seria representada. Assim, o termo passa a ser utilizado também para designar determinados destaques em meio eletrônico. É o que explica Almeida (2002, p. 6),

Como a formatação e a impressão de textos se tornaram automatizadas, o termo foi estendido para todos os tipos de códigos de marcação em textos eletrônicos. Todos os textos impressos são codificados com sinais de pontuação, uso de letras maiúsculas e minúsculas, regras para a disposição do texto na página, espaço entre as palavras, etc. Estes elementos são um tipo de “marcação”, cujo objetivo é ajudar o leitor na determinação de onde uma palavra termina e onde outra começa, ou identificar características estruturais (por exemplo, cabeçalhos) ou simples unidades sintáticas (por exemplo, parágrafos e sentenças). Codificar ou “marcar” um texto para processamento por computadores é também um processo de tornar explícito o que é conjetural. Indica como o conteúdo do texto deve ser interpretado. Dessa forma, por “linguagem de marcação”, entende-se um conjunto de convenções utilizadas para a codificação de textos. Uma linguagem de marcação deve especificar que marcas são permitidas, quais são exigidas, como se deve fazer distinção entre as marcas e o texto e qual o significado da marcação.

De acordo com Bax (2001, p. 33), as linguagens de marcação tiveram sua origem na necessidade de disponibilizar as informações na rede,

A informação e o computador são parceiros antigos, mas a intensificação e democratização do seu uso, aliadas à abstração sempre crescente do nível de interação e troca de informações, criaram terreno propício para a origem das chamadas linguagens de marcação. Este fenômeno foi marcante na primeira década dos anos 90, com o aparecimento da Web. Estas linguagens permitem a construção de padrões públicos e abertos que estão sendo criados para se tentarem maiores avanços no tratamento da informação; elas minimizam o problema de transferência de um formato de representação para outro e liberam a informação das tecnologias de informação proprietárias.

No início da Web havia somente sistemas proprietários, ou seja, sistemas fechados que possibilitam o acesso às informações somente dentro dele mesmo com o uso de software e hardware específicos. Com a criação de padrões públicos ou abertos, foi possível um maior avanço no tratamento da informação e o acesso a informações de modo mais democrático.

De acordo com Bax (2001, p. 32), as linguagens de marcação funcionam do seguinte modo:

Estas linguagens identificam, de forma descritiva, cada “entidade informacional” digna de significado presente nos documentos, como, por exemplo, parágrafos, títulos, tabelas ou gráficos. A partir destas descrições, os programas de computador podem melhor compreender e, em consequência, melhor tratar ou processar a informação contida em documentos eletrônicos.

Segundo Bax (2001) existem dois tipos de marcação: a marcação procedimental e a marcação descritiva. A marcação procedimental do texto está relacionada às marcas inseridas nos documentos, tanto na forma implícita como na forma explícita, que indicam como o texto do documento deve ser apresentado (fontes, caracteres, como o texto deve estar disposto na página, entre outros). A marcação descritiva está relacionada com o uso de “Tag(s)” (etiquetas ou marcas) descritivas que indicarão principalmente a informação existente em um documento e não apenas sua apresentação física, o intuito é que o conteúdo fique separado do estilo da apresentação do texto, proporcionando assim, uma melhor visualização da informação.

Uma Tag descritiva (etiqueta ou marca) determina o início e o fim de um texto ou qualquer outra informação relacionada a ele como parágrafo, título etc. Como afirma Bax

(2001, p. 33) uma Tag “[...] é tudo o que não for considerado conteúdo em um documento. Elas indicam a função (o propósito) da informação no documento, em vez de como ela deve ser apresentada, ou seja, sua aparência física”.

O uso de Tag(s) descritivas permite tratar cada unidade de informação atribuindo-lhes características específicas e possibilitando uma maior estruturação da informação (BAX, 2001). No entanto, o ideal é que o conteúdo do documento esteja separado do estilo usado para sua apresentação, pois isto possibilita recuperar informações com maior significado.

Nesse sentido, partindo do ponto de vista das linguagens de marcação, podemos dizer que os documentos são constituídos por três partes: conteúdo, estrutura e estilo (formatação). “O conteúdo é a informação propriamente dita, a estrutura define como se dá a organização da informação, ou das idéias, no documento e o estilo define o visual da apresentação das informações ao usuário” (BAX, 2001, p. 34).

A vantagem dessa separação entre conteúdo, estrutura e estilo de acordo com Bax (2001, p. 34, grifo do autor) é que,

Tal distinção ou separação promove, ou acaba se revertendo em uma simplificação, pois o autor não tem mais que se preocupar *a priori* com o “visual” da informação, podendo dedicar-se exclusivamente ao conteúdo e à estrutura de apresentação das idéias no documento. Dessa forma, o texto se manterá bem “mais limpo”, sem uma infinidade de códigos que não dizem respeito ao conteúdo da informação, podendo ser mais facilmente compreendido pelo homem.

Portanto, o uso das linguagens de marcação acaba proporcionando um melhor gerenciamento das informações tanto no fator mencionado acima, como também quando permite uma maior liberdade da informação com o uso de padrões não proprietários, assim a informação pode ser convertida de um padrão para outro independentemente do tipo de mídia que está veiculada (monitor, celulares, impressoras, interpretador braile, televisão etc) (BAX, 2001).

Das linguagens de marcação existentes a que prevalece na maioria dos documentos disponibilizados hoje na Internet é a HTML (HiperText Markup Language). No entanto, essa

linguagem foi baseada em um padrão mais geral, a SGML (Standard Generalized Markup Language). Para entendermos um pouco melhor o que significam, serão tratadas a seguir algumas características dessas linguagens, o que facilitará um maior entendimento da linguagem XML (eXtensible Markup Language), apontada na camada estrutura da arquitetura da Web Semântica como necessária à estruturação dos dados.

3.1 LINGUAGEM SGML – Standard Generalized Markup Language

A linguagem SGML foi criada no final da década de 60 e passou a ser considerada como um padrão da ISO em 1986 (GUIMARÃES, 2004; BAX, 2001). Criada por pesquisadores da IBM, o objetivo da SGML era “[...] construir um sistema portátil (i.é., independente de sistema operacional, formatos de arquivos, etc) para o *intercâmbio e manipulação de documentos*” (GUIMARÃES, 2004).

De acordo com Almeida (2002), a linguagem SGML é considerada como um padrão não proprietário de código aberto e, segundo Guimarães (2004), se caracteriza por não restringir os documentos em uma única aplicação, estilo ou sistema de processamento. Para Almeida (2002, p. 6) “Um dos objetivos do SGML é garantir que documentos codificados de acordo com suas regras possam ser transportados de um ambiente de hardware e software para outro, sem perda de informação”.

Apesar de comumente ser chamada de linguagem, a SGML pode ser considerada como uma metalinguagem, ou seja, uma linguagem para a definição e criação de outras linguagens. Em outras palavras, a SGML é uma linguagem *auto descritiva*, ela não possui um conjunto determinado de Tag(s), mas possibilita definir qualquer conjunto, sendo que cada

documento possui uma especificação formal por meio de uma definição de tipo de documento ou DTD(s) – Document Type Definition (BAX, 2001). As DTD(s) definem como as Tag(s) devem ser interpretadas, se há alguma regra para restrições de determinadas Tag(s) e quais são essas regras; e até a ordem em que elas devem aparecer no documento (BAX, 2001).

Por ser um sistema de marcação generalizada, os objetivos da SGML estão relacionados à estruturação rigorosa para apresentação do conteúdo do documento e não somente ao estilo ou layout que o documento irá apresentar. Conforme afirma Guimarães (2004, grifo do autor),

- a marcação de um documento deve descrever a estrutura do documento e outros atributos do mesmo, em vez de especificar o *processamento* a ser feito no mesmo,
- a marcação deve ser definida rigorosamente, de forma que sistemas formais como programas possam ser usados para processar o documento.

Quanto às características da linguagem, a SGML permite que o usuário crie suas próprias Tag(s), sem a imposição de nenhum conjunto específico, pois o uso das DTD(s) irá determinar como serão essas Tag(s) especificando as características de uso, organização e construção do documento.

De acordo com Santarém Segundo (2004) um documento SGML apresenta três camadas:

- Estrutura: que define como ocorre a organização da informação no documento, por meio da DTD. A estrutura fornece, portanto, detalhes específicos das Tag(s) dos caracteres e como o padrão será aplicado ao documento;
- Conteúdo: é toda a informação contida em documento. Essas informações estão dispostas nas Tag(s), que por sua vez definem cada parte do conteúdo como título, parágrafos, figuras etc;
- Estilo: está relacionado com a forma de apresentação do documento, no entanto, o estilo não é definido pela linguagem SGML e sim por um padrão de estilo denominado

DSSSL (Document Style Semantic and Specification Language) que deve ser usado junto com a SGML, pois ele irá padronizar o estilo da sintaxe, semântica e layout do documento.

Apesar de apresentar algumas vantagens, tais como melhor compartilhamento da informação e portabilidade dos dados, a SGML apresenta certa complexidade para ser utilizada amplamente na Web.

Por esse motivo, mesmo sendo usada no início dos anos 80 por algumas organizações, tornou-se necessária a criação de uma linguagem mais simples e fácil de ser utilizada para a criação de documentos na Web. Tendo como base a SGML, foi desenvolvida então a linguagem HTML, que será melhor explicada a seguir.

3.2 LINGUAGEM HTML – HiperText Markup Language

Criada especialmente para ser uma linguagem simples e fácil de ser utilizada, a HTML se transformou em uma linguagem extremamente popular. Atualmente a maioria das páginas da WWW estão em HTML.

A HTML é uma linguagem de marcação que basicamente está voltada para a estruturação e apresentação visual de documentos da Web (GUIMARÃES, 2004). Isso possibilitou o acesso às informações de modo simples e em qualquer arquitetura computacional (CASTRO, 2001).

Tendo sua origem na linguagem SGML, a “HTML possui um grupo de *tags* predefinidos, concebidos com a função de organizar a informação a ser transferida por meio de páginas Web” (BAX, 2001, p. 35).

Estruturalmente um documento HTML é um conjunto hierárquico de elementos que são demarcados por duas Tag(s) iniciais e finais. De acordo com Santarém Segundo (2004, p. 41), “Um documento HTML é delimitado pelas Tag(s) <HTML> e </HTML> e dividido em cabeçalho e corpo. O cabeçalho é delimitado pelas Tags <HEAD> e </HEAD> e o corpo pelas tags <BODY> e </BODY>”. Como pode ser visto no exemplo abaixo:

```

<HTML>
  <HEAD>.....
  .....
  </HEAD>
  <BODY>.....
  .....
  </BODY>
</HTML>

```

FIGURA 3: Estrutura do documento HTML.
FONTE: do autor.

- Cabeçalho <HEAD> </HEAD>: além de conter informações sobre o documento, como o título que é mostrado no alto da janela do browser por exemplo, o cabeçalho contém informações importantes como as Tag(s) <META>. Essa Tag não apresenta função de apresentação do documento, mas são Tag(s) que facilitam a descrição das informações nos documentos auxiliando na recuperação pelos robôs de busca. Seria adequado que essas Tag(s) fossem sempre usadas para garantir um pouco mais de descrição do conteúdo das páginas HTML, pois é por meio delas que as ferramentas de busca recuperam as informações para a construção de suas bases de dados. Infelizmente poucos desenvolvedores utilizam esse recurso. Por isso muitas vezes recuperamos informações irrelevantes para nossas necessidades (SANTARÉM SEGUNDO, 2004).
- Corpo do documento <BODY> </BODY>: esta Tag contém as informações que serão mostradas para o usuário. Para que o texto possa ser visualizado com parágrafos, cores, links etc, é necessário o uso de outras Tag(s) para determinar essas funções dentro do

corpo do documento. Um exemplo são as Tag(s) `....` que inserem negrito no texto (SANTARÉM SEGUNDO, 2004). No entanto, na linguagem HTML as Tag(s) que vão dentro das Tag(s) `<BODY>` e `</BODY>` não apresentam função semântica, ou seja, não indicam o conteúdo, apenas indicam como o conteúdo deve se apresentado.

Como já foi dito, a linguagem HTML contribuiu muito para a popularidade da Web por ser uma linguagem de grande utilização pelos desenvolvedores de sites. De acordo com Bax (2001, p. 36),

Contrariamente à SGML, que é um padrão complexo e difícil de implementar, a grande vantagem de HTML é sua relativa facilidade em ser entendida pelo usuário da Web e de ser processada, mesmo em diferentes navegadores. Este aspecto foi o principal responsável pela explosão da Web. Paradoxalmente, a falta de flexibilidade acabou se revelando uma força da linguagem e seu fator popularizador.

Como pode ser visto, as características da linguagem HTML traz certa limitação e uma falta de flexibilidade na troca mais efetiva de informações na Web (BAX, 2001). Diante dessa limitação, a cada nova versão foram sendo inseridas novas Tag(s) e atributos de estilos na tentativa de melhorar a linguagem, mas por estarem relacionadas à forma de apresentação do documento essas atualizações fizeram com que a linguagem HTML ficasse inadequada para sua formatação e de difícil leitura para o homem (BAX, 2001).

Na tentativa de definir padrões mínimos na linguagem HTML, a World Wide Web Consortium ou W3C, que é uma organização que cuida do desenvolvimento e manutenção dos padrões da Web, definiu na versão 4.0 da linguagem HTML o uso das chamadas folhas de estilo ou CSS (Cascading Style Sheet) (BAX, 2001).

Criada no final de 1996, a CSS - Cascading Style Sheet é uma linguagem usada para definir estilos, ou seja, é uma linguagem desenvolvida especialmente para a formatação do conteúdo dos documentos (BAX, 2001; CASTRO, 2001).

As folhas de estilo podem ser usadas e ligadas ao documento HTML de quatro modos: externo (apesar de estarem localizadas no cabeçalho do documento HTML o arquivo para as

folhas de estilo são criadas em um arquivo independente e a ligação com o documento HTML é feita por um link para esse arquivo), incorporado (as especificações das folhas de estilo aparecem diretamente no cabeçalho do documento HTML e são definidas pelas Tag(s) <STYLE> e </STYLE>), inline (a indicação aparece no corpo do documento, ou seja, no conteúdo e somente a Tag especificada apresenta a formatação indicada), importado (as folhas de estilo estão em outra área da Internet e são ligadas ao documento por meio de um link, sem necessidade de cópia) (SANTARÉM SEGUNDO, 2004).

O ideal para os documentos na Web, de acordo com as recomendações da W3C, é que seu conteúdo fique separado da sua forma de apresentação, ou seja, as folhas de estilo irão definir com os elementos deverão ser mostrados nos navegadores, permitindo que o conteúdo do documento fique livre das Tag(s) que apenas marcam como o texto deve ser apresentado (BAX, 2001). Essa é uma das vantagens do uso de folhas de estilo como a CSS em documentos HTML, porém Santarém Segundo (2004) aponta ainda as seguintes vantagens: aplicação de diferentes estilos em um mesmo documento, fácil manutenção do documento, consistência e uniformidade no arranjo do documento.

Apesar de existir a possibilidade do uso de folhas de estilo para a separação do conteúdo com a forma de apresentação do documento e apesar de ter sido a responsável pela popularização da WWW, a linguagem HTML ainda se apresenta com uma certa limitação em determinados casos. Diante da grande quantidade de informações disponíveis na rede atualmente, foi preciso se pensar em uma outra linguagem que pudesse deixar mais claro o conteúdo do documento.

Foi desenvolvida então a linguagem XML, que está sendo indicada atualmente como necessária para garantir um melhor armazenamento e compartilhamento das informações, principalmente quando falamos no desenvolvimento da Web Semântica. A seguir será melhor explicada a linguagem XML.

3.3 LINGUAGEM XML – eXtensible Markup Language

A linguagem XML foi criada em 1996 pelo W3C. Teve como base para sua criação a linguagem SGML e foi desenvolvida no intuito de atender as novas necessidades de gerenciamento do crescente número de informações da Web.

XML é similar a HTML em vários aspectos: é expressa em arquivos de texto (ASCII), foi criada no intuito de armazenar e transmitir dados; além de fazer o uso de Tag(s) iniciais e finais que qualificam cada unidade de informação. No entanto, sua diferença está no fato de não possuir um número fixo de Tag(s) e por se preocupar mais com o conteúdo do documento e não somente com sua forma de apresentação. Bax (2001, p. 36, grifo do autor) afirma que,

[...] diferentemente de HTML, XML não propõe um número fixo de marcas. Um elemento XML pode ser marcado da forma que o autor do documento bem entender, ou seja, com o termo que melhor descreve a informação na sua opinião [...].

Também como já foi visto, ao invés de descrever como os dados devem ser *mostrados*, as marcas indicam o que cada dado *significa*. Qualquer agente (humano ou de *software*) que receba este documento pode decodificá-lo e usar os dados como lhe convier.

Podemos ver então que a linguagem XML se dedica principalmente ao conteúdo do documento e as formas de apresentação deste documento ficam a cargo de outras ferramentas que serão vistas mais adiante e que podem ser associadas à linguagem. Isso significa que a XML contém em suas Tag(s) o significado do conteúdo a ser disponibilizado, fornecendo pistas semânticas sobre o significado dos dados. Na Figura 4 é apresentado o exemplo de um documento simples estruturado pela linguagem XML:

```

<?xml version = "1.0" encoding = "ISO - 8859 - 1"?>
<!--livro.xml -->
<?xml: stylesheet type = "text/css" href = "livro.css"?>

<! DOCTYPE livro SYSTEM "livro.dtd">
<livro> Referência para a assinatura de Base de Dados.
    <título>O livro depois do Livro</título>
    <autor>Giselle<lastname>Beiguelman</lastname></autor>
    <local>São Paulo</local>
    <editora>Peirópolis</editora>
    <ano>2003</ano>

</livro>

```

FIGURA 4: Exemplo de um documento XML simples.
FONTE: do autor.

A XML torna o conteúdo do documento mais visível, pois não contém códigos de formatação misturados com o conteúdo. Sua estrutura mostra claramente o significado das informações, conforme ilustra o exemplo da Figura 4. O conteúdo está contido entre as Tag(s) <livro>...</livro> e sua forma de apresentação (formatação) está no uso da folha de estilo CSS indicada na Tag “<?xml: stylesheet type = “text/css” href = “livro.css”?>”.

A linguagem XML está sendo apontada hoje como uma das ferramentas importantes para gerenciar e facilitar o acesso ao grande volume de informações disponíveis na rede, pois tem o potencial de tornar mais fácil as tarefas de armazenamento, tratamento, recuperação e intercâmbio das informações.

Isso está relacionado com as vantagens que a linguagem XML apresenta, tais como: possibilita um maior enfoque ao conteúdo do documento e não na forma de apresentação (estilo); por ser uma linguagem melhor estruturada permite o acesso ao conteúdo do documento de modo mais específico; permite a visualização de um mesmo documento sob formas diferentes por meio da utilização das folhas de estilo (flexibilidade); o uso padronizado da XML permite uma maior interoperabilidade entre os dados na Web; não possui Tag(s) fixas, possibilitando a criação de Tag(s) quando necessário (extensibilidade); permite certa automação para os agentes de softwares, desde que os documentos sejam bem

formados; o uso da XML possibilita representar (indicar) precisamente a semântica da informação; permite uma conexão entre documentos criando uma rede de conexão de conhecimentos etc.

3.3.1 CARACTERÍSTICAS DA LINGUAGEM XML E TECNOLOGIAS ASSOCIADAS

Uma das características importantes da XML é a extensibilidade, que permite criar novas Tag(s) para a representação semântica do documento de acordo com as necessidades do usuário. Outra característica é o enfoque dado ao conteúdo do documento e não somente a sua forma de apresentação, como acontece na linguagem HTML (MARTÍNEZ GONZALEZ, 2000). Além disso, o conteúdo do documento em HTML fica disposto em um texto corrido, trazendo dificuldades aos agentes de software em reconhecer o significado dos dados. Já na linguagem XML, o conteúdo do recurso aparece separado da formatação (estilo) sendo organizados mais claramente, facilitando assim, a interpretação pelos agentes de software.

Por ter sido criada especialmente para a recuperação e intercâmbio de dados, a XML permite, entre outras coisas, um melhor armazenamento e descrição da informação. Isso acontece pelo fato da XML apresentar uma sintaxe rígida, exigindo que regras sintáticas sejam seguidas para um documento ser bem formado (CASTRO, 2001).

Quando se fala em documentos bem formados quer dizer que se trata de documentos com Tag(s) corretas, se há uma Tag aberta é preciso haver uma para fechar. Isso não ocorre com a linguagem HTML que apresenta muitas vezes Tag(s) abertas, sem a Tag que corresponde ao seu fechamento. Essa falta de padronização na HTML acaba dificultando o

tratamento automático dos dados. Como a linguagem XML é mais rígida esse tipo de problema não ocorre, o que acaba sendo uma vantagem (CASTRO, 2001). No entanto, é necessário que além de bem formado um documento em XML também seja válido e essa validade é dada pela definição de tipo de documento ou DTD.

Uma aplicação XML é criada pelo usuário sem limitações quanto ao número e tipos de Tag(s) criadas. No entanto, para que esses dados sejam definidos e válidos é necessário o uso de “Esquemas”, ou seja, conjuntos de regras para definir os elementos e atributos permitidos, compatíveis e necessários na linguagem de marcação criada por cada pessoa. São, portanto, ferramentas importantes para manter a consistência dos documentos.

Além das características citadas, ainda existe uma outra que está relacionada às tecnologias associadas a essa linguagem. Castro (2001) afirma que a XML é uma linguagem simples e as tecnologias que estão associadas à ela formam o diferencial para que ela ofereça melhor qualidade que as outras linguagens. Essas tecnologias são: DTD, XML *Schema*, XSL, Xlink, Xpointer, Xpath, *Namespaces* etc, e serão brevemente explicadas a seguir.

A. DTD – Document Type Definition (definição de tipo de documento): A DTD é responsável pela modelagem de dados. Por meio dela é possível determinar quais os elementos o documento terá e como será a ligação entre eles. A DTD pode ser interna ou externa ao documento. Sendo externa é possível ser usada por diversos documentos (CASTRO, 2001). De acordo com Castro (2001, p. 41),

Uma DTD deve definir regras para cada elemento e atributo que apareça no documento XML. Se não, o documento XML não será considerado válido. Se em algum ponto for necessário adicionar elementos ao documento XML, também é necessário adicionar suas definições à DTD correspondente (ou criar uma nova DTD, se preferir).

O exemplo da Figura 5 ilustra uma DTD externa ao documento Livro exemplificado na

Figura 4:

```

<?xml version = "1.0" encoding = ISO - 8859 - 1"?>
<!--livro.dtd -->
<?xml: stylesheet type = "text/css" href = "livro.css"?>

<!ELEMENT livro (título, autor+, edição?, editora?)>
<!ELEMENT título (CDATA)>
<!ELEMENT autor (CDATA)>
<!ELEMENT local (CDATA)>
<!ELEMENT editora (CDATA)>
<!ELEMENT ano (CDATA)>

```

FIGURA 5: Exemplo de uma DTD separada do documento XML.
FONTE: do autor.

Neste exemplo, podemos visualizar os elementos que formarão as Tag(s) do documento XML pertencentes ao exemplo da Figura 4, ou seja, a definição de tipo de documento, por meio do estabelecimento dos elementos que formar.

B. XML Schema: o uso desta tecnologia vem sendo recomendado pela W3C em substituição das DTD(s), que embora tenham uma sintaxe específica são um pouco limitadas. Já o *XML Schema* permite controle maior sobre o conteúdo de um documento XML, por definir tanto elementos globais (para serem utilizados da mesma forma em todo o documento XML) e elementos locais (com significado particular em determinado contexto) (CASTRO, 2001). O *XML Schema* proporciona meios para definir a estrutura, o conteúdo e a semântica dos documentos e apresenta uma vantagem sobre as DTD (s) devido às diferenças que apresentam. Segundo Castro (2001, p. 69, grifo do autor):

[...] as DTDs são escritas em uma sintaxe que tem pouca relação com XML e que não podem ser analisadas com um parser XML. Em segundo lugar, todas as declarações em uma DTD são globais, o que significa que você não pode definir dois elementos diferentes com o mesmo nome, mesmo se aparecerem em contextos separados.

Finalmente, e talvez o mais importante, as DTDs não podem controlar que *tipo* de informação determinado elemento ou atributo podem conter.

A vantagem do uso de *XML Schema* é permitir um maior controle sobre o conteúdo do recurso. Na Figura 6 podemos visualizar um exemplo de um *XML Schema* (XML, 2001):

```

<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="poema">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="titulo" type="xs:string"/>
        <xs:element name="autor" type="xs:string"/>
        <xs:element name="data" type="xs:string"/>
        <xs:element name="corpo" type="Tcorpo"/>
      </xs:sequence>
      <xs:attribute name="tipo"
        type="xs:string"
        use="optional"
        default="soneto"/>
    </xs:complexType>
  </xs:element>
... </xs:schema>

```

FIGURA 6: Exemplo de XML Schema.

FONTE: XML (2001).

Disponível em: <<http://www.di.uminho.pt/~jcr/AULAS/micei2002/Slides/XML%20Schema.ppt>>.

O exemplo da Figura 6 trata de um *Schema XML* de um poema em forma de soneto, escrito em XML. Neste exemplo, as Tag(s) do *Schema XML* define não só os elementos, tais como: título, autor, data, corpo; mas também a seqüência dos elementos e o tipo de documento que será (poema do tipo soneto).

C. XSL (eXtensible Stylesheet Language) e CSS (Cascading Style Sheet): XSL e CSS são folhas de estilo, ou seja, um conjunto de regras que se aplicam ao documento para determinar a apresentação ou o estilo a ser visto pelo usuário. As folhas de estilo especificam ao navegador ou *browser* as características da apresentação do documento como: tipo de cor, tamanho da letra, tipo de fonte, espaçamento entre outros. A vantagem do uso de folhas de estilo é que as indicações para formatação e apresentação do documento ficam separadas do conteúdo, facilitando uma formatação futura e tornando o conteúdo mais visível (CASTRO, 2001). A diferença entre essas duas folhas de estilo é que a XSL é uma especificação mais atual que a CSS para se estabelecer a forma de apresentação dos documentos. Como ainda não está completamente estabelecida, a W3C

dividiu a XSL em duas partes: a XSL-FO (para Formatos de Objetos), que ainda é uma tecnologia que está em fase de conclusão; e a XSLT (para Transformação) que é utilizada juntamente com as folhas de estilo em CSS (CASTRO, 2001). Como não foi possível encontrar exemplos da XSL por ser uma tecnologia nova, será apresentado somente um exemplo da folha de estilo em CSS, conforme pode ser visto na Figura 7:

```
livro
{ display :block; font – family:Verdana; font
– size:12pt; }
título
{ display :block; margin-top:1em;
font_weigh:bold; }
autor
{ display :block; background-color:teal;
font-style: italic; color:white; }
edição
{ display : inline; }
editora
{ display : nome; }
```

FIGURA 7: Exemplo de folha de estilo em CSS.
FONTE: do autor.

O exemplo da Figura 7 mostra que para cada elemento descritivo: livro, título, autor, edição, editora, existe um estilo de formatação específico. A folha de estilo CSS determina, conforme o exemplo, que o elemento livro seja escrito em “verdana” com fonte tamanho “12”.

D. Xlink (XML Linking Language), Xpointer (XML Pointer Language) e Xpath (XML Path Language): estas três tecnologias são responsáveis pela determinação dos nós de ligação em um ou vários documentos XML. De acordo com Castro (2001), os designers da XML “[...] optaram por criar um método mais robusto que lhe permite criar links multidirecionais, controlar como e quando os links são ativados e muito mais”. Martínez González (2000) destaca que as características do Xlink são: extensibilidade, inclusão de informações adicionais sobre a semântica do recurso e a relação entre os outros recursos,

possibilidade de criação de links bidirecionais, links múltiplos, possibilidade de criação de links fora do documento e entre fragmentos de documentos. Associado ao Xlink existe o Xpointer que pode identificar atributos do tipo ID e fazer um link para o valor deste atributo. Além disso, pode percorrer o documento XML até chegar ao elemento desejado (CASTRO, 2001). Juntamente com esses dois recursos existe também o Xpath que é “[...] um sistema para descrever os conjuntos de nós especificando seu local no documento XML [...]” (CASTRO, 2001, p. 153). Essas tecnologias tornaram a linguagem XML mais completa. Entretanto, é preciso destacar que o XPath está associado a folha de estilo XSL, que é uma tecnologia ainda em desenvolvimento e que o Xlink e o XPointer são tecnologias que não são compatíveis com os principais navegadores, por isto não é possível verificar como funcionam realmente. No entanto, dentre essas tecnologias, é possível visualizarmos um exemplo simples do uso de um Xlink utilizado pela linguagem XML, conforme está apresentado na Figura 8 (CASTRO, 2001, p. 226):

```
<endangered>_species xmlns: xlink =  
<http://www.w3c.org/1999/xlink>  
<animal>  
<name language="English"> Tiger </name>  
...  
<source xlink:type= "simple" xlink:href=  
  "http://www.worldwildlife.org/species/species.cfm?sectio  
  nid=120&newspaperid=21"  
  Xlinkrole="information source" xlink:title= "Source of  
  Information" xlink: show="replace"  
  Xlink:actuate= "onRequest"/>  
...
```

FIGURA 8: Exemplo de Xlink simples.

FONTE: Castro (2001, p. 226).

O exemplo da Figura 8 mostra um xlink do tipo simples, utilizado para fazer a conexão para um arquivo externo.

E. Namespaces: segundo Castro (2001, p. 122) *Namespaces* pode ser definido como,

[...] um conjunto de elementos e atributos relacionados, identificados por um nome que compartilha uma URL comum. *Namespaces* são utilizados com mais frequência para distinguir elementos nomeados semelhantes, declarados globalmente, uns dos outros. (Elementos declarados localmente são geralmente tornados exclusivos por seu contexto).

Se os *namespaces* vão distinguir os elementos, então podemos dizer que de certa forma sua função é identificar o conjunto de elementos e atributos presentes em um documento XML. Segue abaixo um exemplo de *namespace* (CASTRO, 2001, p. 113):

```
<xsd:schema xmlns:xsd =  
  "http://www.w3c.org/2000/10/XMLSchema/">  
  <xsd:element name = "name" type = "xsd:string"/>  
  <xsd:element resource = "source" type = "xsd:string"/>  
  <xsd:element name = "river">  
    <xsd:complexType>  
      <xsd:sequence>  
        <xsd:element ref = "name"/>  
        <xsd:element ref = "source"/>  
      ...  
    <xsd:element>  
    ...
```

FIGURA 9: Exemplo de *namespace*.
Fonte: Castro (2001, p. 113)

O exemplo da figura 9 ilustra a distinção de elementos que compõem um documento XML e que compartilham uma URL comum para sua especificação.

Não é intuito deste capítulo tratar da linguagem XML e suas tecnologias associadas em todos os seus detalhes, mas sim fornecer informações básicas e introdutórias ao assunto, na tentativa de destacar a importância e a potencialidade que a XML nos oferece.

Como pode ser visto pelas características descritas acima, a XML exige uma padronização minuciosa para a criação de seus documentos. Para Castro (2001) isso poderá criar um obstáculo para sua utilização na construção de páginas pessoais.

No entanto, apesar de minuciosa essa linguagem possibilita um melhor tratamento do conteúdo dos recursos informacionais e se apresenta como uma linguagem ideal aos profissionais da informação para o gerenciamento e intercâmbio de grande quantidade de informações na rede devido à sua padronização (CASTRO, 2001).

A linguagem XML tem sido apontada como uma necessidade para melhorar a questão do armazenamento, recuperação, compartilhamento e troca de informações, pois proporciona uma estruturação do conteúdo do recurso de modo mais organizado e detalhado. Porém, seu uso efetivo ainda não se faz presente em toda a Web.

Bax (2001) comenta que a linguagem HTML é limitada por apenas apresentar a informação e possibilitar somente uma marcação estrutural e não uma marcação semântica. A linguagem SGML, por sua vez, é muito complexa para ser amplamente utilizada, por isto aponta a linguagem XML como um caminho para se resolver esses dois extremos e afirma que,

XML parece ser um bom compromisso entre a flexibilidade em termos de representação informacional e a simplicidade necessária para se tornar uma ferramenta ubíqua na Web.

Pode-se dizer que a passagem de uma marcação estrutural com HTML para uma marcação semântica com XML é uma fase importante no esforço para se transformar a Web de um espaço global de *informação* em uma rede universal de *conhecimento*. (BAX, 2001, p. 37, grifo do autor).

Bax (2001) ainda afirma que o nascimento de estruturas de marcação de dados mais ricas, ou seja, que melhoram a descrição dos dados irá promover um melhor armazenamento, compartilhamento e processamento das informações disponibilizadas na Web. E a linguagem XML possui o potencial para fazer com que isto ocorra.

3.3.2 LINGUAGEM XML E A WEB SEMÂNTICA

A semântica na Web não está relacionada apenas com ao conteúdo do recurso, mas também como os recursos se relacionam entre si. Para que isso ocorra é preciso o acesso a coleções estruturadas de informações e essa estruturação terá início com o uso de uma linguagem de marcação adequada, ou seja, uma linguagem que possibilite o acesso ao conteúdo estruturado do documento sem interferência de como estes dados serão apresentados. Por este motivo é que a linguagem XML vem sendo considerada essencial para o armazenamento e a estruturação dos recursos informacionais na rede.

Para a Web Semântica funcionar efetivamente é preciso que as informações disponíveis na rede estejam estruturadas. O primeiro passo para a estruturação dos dados começa com a linguagem de marcação.

Sendo assim, podemos destacar que a linguagem XML estabelece uma relação com a Web Semântica sob três aspectos: semântico, sintático e estrutural.

O primeiro aspecto está relacionado à semântica contida nos documentos. Cada tipo de documento possui um determinado conjunto de termos que representam conceitos específicos que terão maior significado para uma determinada comunidade. Bax (2001, p. 37) afirma que “Quanto maior a comunidade, menor é o conjunto de definições compartilhadas; quanto menor e mais focalizada a comunidade, maior será esse conjunto”, essa afirmação está relacionada com a criação de ontologias que definem a semântica na Web.

Considerando essa afirmação, Bax (2001, p. 37, grifo do autor) estabelece uma relação da necessidade de estar definindo a semântica de uma comunidade utilizando para isto a linguagem XML e aponta que,

Como a semântica depende das definições estabelecidas em uma comunidade específica, é razoável que, para se melhorar a comunicação nestas comunidades, deva existir uma abertura nas linguagens para as definições específicas de cada comunidade. XML torna isso possível, ou seja, torna-se viável se capturarem *ontologias* comunitárias sob a forma de DTD's e assim promover uma descentralização natural do controle das especificações das linguagens de marcação.

De acordo com Moura (2002a), o uso de vocabulários específicos, ontologias e padrões de metadados são recursos necessários para assegurar a interoperabilidade semântica¹ na Web. Diante disso, a XML pode proporcionar as condições para que isto ocorra por ser uma linguagem voltada para o conteúdo do documento.

O segundo aspecto da relação da linguagem XML com a Web Semântica refere-se à sintaxe e o uso de metadados. Por ser uma linguagem extensível, ou seja, permite a criação de novas etiquetas; é possível incluir nessas Tag(s) elementos de metadados que descrevam e representem o conteúdo do recurso.

Nesse caso, a XML permite, devido a essa extensibilidade, a inclusão de uma variedade de tipos de padrões ou formatos de metadados existentes, possibilitando assim que os documentos criados nessa linguagem apresentem uma melhor representação. Além disso, a linguagem XML, “[...] considerada pela W3C como a linguagem mais importante para a representação e troca de dados na Web” irá determinar como os metadados deverão ser codificados para a transferência de informações na rede, favorecendo assim, a interoperabilidade sintática² (MOURA, 2002a).

O terceiro aspecto está voltado para a questão estrutural. Para que haja um intercâmbio de dados e interoperabilidade na rede é preciso o uso de arquiteturas de metadados para garantir a interoperabilidade entre padrões distintos.

Nesse caso, a interoperabilidade estrutural³ especifica como os recursos estão organizados, juntamente com os tipos de recursos envolvidos e os possíveis valores para cada tipo (MOURA, 2002a). Ainda de acordo com Moura (2002a, grifo do autor),

¹ Semântica, também chamada de ciência do significado, é a parte da lexicologia que estuda a significação das palavras, ou seja, está relacionada com a definição do significado preciso das palavras (JOTA, 1976).

² Sintática, relacionada a sintática que trata dos signos e suas combinações e a sintaxe, que estuda as palavras e as relações que elas estabelecem uma com as outras (relação de concordância, subordinação, disposição ou ordem em um determinado ambiente ou sistema) (JOTA, 1976).

³ Estrutural, ligada a estrutura de um sistema, em outras palavras está relacionada aos modelos de organização e relação entre os componentes de um sistema (JOTA, 1976).

Esta característica pode beneficiar-se dos recursos providos pelo modelo RDF (Resource Description Framework), onde um esquema pode dividir vocabulários distintos através da utilização de *namespaces* XML, proporcionando a troca de informações na Web.

O uso de padrões de metadados e arquiteturas de metadados com a linguagem XML, garantem o intercâmbio de informações estruturadas na rede (MOURA, 2002a). Nesse sentido, vemos que a XML facilita os três tipos de interoperabilidade: semântica, sintática e estrutural, tão necessárias para o estabelecimento da Web Semântica. Por este motivo é que a XML apresenta-se atualmente como um requisito para armazenamento da informação na Web.

Acredita-se que a representação da informação possa promover uma integração e intercâmbio entre recursos heterogêneos distribuídos na rede, e isto só poderá ocorrer de modo ideal a partir do uso da linguagem XML e principalmente do uso de metadados (Moura, 2002a). Portanto, será tratada no próximo capítulo a questão da interoperabilidade dos dados, que proporcionará esse intercâmbio de dados e metadados, por meio de uma linguagem como a XML.

4 INTEROPERABILIDADE NA WEB: uso da arquitetura de metadados RDF – Resource Description Framework

Podemos perceber, pelos estudos sobre metadados e XML, que os métodos de representação e organização da informação passaram por mudanças, que incluem não somente os métodos e a tecnologia usada na criação de formas de representação de documentos e informações, mas também nos padrões, que são essenciais para a busca e recuperação da informação.

Apesar de estarmos vivenciando evoluções tecnológicas, é preciso lembrar que a essência do tratamento e recuperação de informações continua a mesma. Continuamos, por meio de normas, regras e padrões, buscando maneiras para realizar uma representação e uma organização da informação de modo eficiente. O que tem mudado, no entanto, é o meio tecnológico onde a informação está inserida e necessariamente a alteração nas formas de organização e de representação.

Nesse sentido, a característica marcante desse novo cenário fundamentado em ambientes e infraestrutura para informação digital, está baseada em uma única palavra: interoperabilidade, ou seja, padrões que promovam o intercâmbio de informações entre diferentes plataformas e sistemas heterogêneos.

A necessidade de compartilhamento da quantidade de informações que vem sendo produzida nos últimos anos é uma realidade que tem desafiado profissionais de várias áreas do conhecimento, principalmente no tratamento de ambientes altamente distribuídos e heterogêneos como a Web.

Promover a tão desejada interoperabilidade na Web ou em outro ambiente eletrônico esbarra em questões típicas para qualquer processo de integração de recursos. Questões essas

relacionadas com o uso de formatos de metadados destinados a promover a representação do recurso, linguagem de marcação XML para uma melhor estruturação e representação dos recursos; e ferramentas importantes para o estabelecimento de interoperabilidade sintática, semântica e estrutural, tais como as arquiteturas de metadados tratadas neste capítulo e as ontologias, tratadas anteriormente no capítulo dois.

Contudo, antes de uma abordagem mais profunda sobre o tema, é preciso esclarecer, primeiramente, o significado do termo interoperabilidade.

De acordo com o Glossário de Termos Técnicos do *Institute for Telecommunication Sciences*, a interoperabilidade pode ser definida como,

A habilidade de sistemas, unidades, ou forças de prover serviços e aceitar serviços de outros sistemas, unidades ou forças, e assim fazer uso desses serviços permitindo estabelecer uma troca para que possam atuar efetivamente juntos (INTEROPERABILITY, 1996, tradução nossa).

Ou ainda, “A condição alcançada entre sistemas de comunicação-eletrônica [...] quando podem trocar satisfatoriamente informações ou serviços entre sistemas ou entre seus usuários” (INTEROPERABILITY, 1996, tradução nossa), podendo ser considerada como a compatibilidade entre diferentes sistemas ou plataformas.

Siqueira (2003) aponta que podemos encontrar diversas definições para interoperabilidade dependendo do contexto em que o termo está inserido, como por exemplo, definições de interoperabilidade no contexto de banco de dados, softwares, ambientes Web, arquiteturas de redes entre outros.

Para o contexto dessa pesquisa, podemos utilizar a definição de interoperabilidade adotada por Siqueira (2003, p. 38), que a define como sendo,

[...] a capacidade de compartilhamento de informações entre softwares, independente da estrutura de armazenamento dos dados usada em seu banco de dados, ou seja, a estrutura de armazenamento dos dados não impede a troca de informações entre instituições.

Em síntese, podemos dizer que interoperabilidade é a capacidade de compartilhamento de informações em diferentes sistemas e que, por meio de algumas ferramentas como

linguagem de marcação adequada, uso de metadados e arquiteturas de metadados, estas informações registradas e armazenadas em diferentes estruturas e em diferentes comunidades do conhecimento poderão ser intercambiadas e trocadas nestes sistemas, fazendo com que haja um trabalho conjunto entre sistemas.

Após a análise dessas definições conclui-se que a interoperabilidade não é uma novidade entre os bibliotecários que tem, durante décadas, intercambiado registros catalográficos. A interoperabilidade apresenta-se com um novo nome para designar o intercâmbio de informações, agregando novas tecnologias, que proporcionam o aumento da velocidade e novos tipos de serviços, ampliando o uso de padrões, esquemas e formatos que propiciam estas operações.

Mas para que a interoperabilidade seja realmente viabilizada é preciso atender algumas características que são imprescindíveis para a troca de informações em redes de comunicação, estas características estão divididas em três níveis: semântico, estrutural e sintático que serão apresentados em seguida:

A. INTEROPERABILIDADE SEMÂNTICA: de acordo com Moura (2002a)

A interoperabilidade semântica possibilita compreender o significado de cada elemento descritor do recurso, juntamente com as associações nele embutidas. O uso de vocabulários específicos, ontologias e/ou padrões de metadados são essenciais para assegurar esse tipo de interoperabilidade.

Nessa afirmação podemos identificar segundo Marino (2001), dois subníveis da interoperabilidade semântica: a epistemológica, que trata do significado dos elementos descritores do formato, bem como das relações nele existentes; e a ontológica, que trata do uso de ontologias, vocabulários controlados e padrões de metadados para o estabelecimento dos significados dos dados representados.

Um exemplo desse tipo de interoperabilidade está na análise semântica da Área 1 (Título e Indicação de Responsabilidade) do AACR2 e do Campo 245 (Título e Indicação de

Responsabilidade) do MARC 21, que trazem informações sobre indicação de título do recurso informacional e as pessoas e/ou organizações responsáveis pela criação do conteúdo intelectual de um item documentário. Baseado nesse exemplo podemos dizer que esse nível de interoperabilidade trata do estabelecimento de significados entre elementos descritivos de um padrão e/ou estrutura comum para uma determinada comunidade.

Entretanto, é preciso lembrar que esse nível promove, em parte, a interoperabilidade em nível semântico, mas a compreensão plena do conteúdo é dada por outras ferramentas, como as ontologias. (MOURA, 2002b).

B. INTEROPERABILIDADE ESTRUTURAL: de acordo com Moura (2002a) "A interoperabilidade estrutural especifica como os recursos estão organizados, juntamente com os tipos envolvidos e os possíveis valores para cada tipo". Em outras palavras, podemos dizer que esse requisito está relacionado com as características da estrutura dos elementos descritivos que compõe um padrão ou formato de metadados. De acordo com Barreto (1999, p. 85) a interoperabilidade estrutural,

[...] define cada elemento componente de um padrão de metadados, descreve os seus tipos, a escala de valores possíveis para esses elementos e os mecanismos utilizados para se agrupar (ou relacionar) esses elementos de modo a que possam ser processados de forma automática. Quanto mais complexa for a estrutura do padrão de metadados, mais complexo deve ser o *modelo de dados* empregado para descrevê-la.

Podemos dizer então que a interoperabilidade estrutural proporciona uma representação da estrutura descritiva dos padrões ou formatos de metadados, desde os mais simples, passando pelos intermediários, como o padrão de metadados Dublin Core (DC)⁴, até

⁴ Formato ou padrão de metadados para a representação dos recursos para fins de localização. Será abordado com mais detalhes no capítulo 5.

os mais complexos como o formato MARCXML⁵. Esse nível estabelece meios para que o intercâmbio seja realizado de forma adequada e coerente.

C. INTEROPERABILIDADE SINTÁTICA: de acordo com Barreto (1999, p. 85) "A sintaxe provê uma linguagem comum para representar a estrutura do metadado". Para Moura (2002a) "A interoperabilidade sintática determina como os metadados devem ser codificados para a transferência de informações", pois determina as combinações e relações que os dados e metadados estabelecem uns com os outros.

Para garantir a interoperabilidade sintática, Barreto (1999) e Moura (2002a) apontam a necessidade do uso da linguagem XML para gerenciar a troca de informações, que é considerada atualmente pela W3C como uma linguagem importante para o auxílio na representação e troca de informações na Web. Portanto, a interoperabilidade sintática irá determinar a forma de estruturação dos elementos, ou seja, a lógica que estabelece como devem ser apresentadas as informações.

Como pode ser percebido, esses três níveis vão atuar em conjunto, e considerados como princípios de interoperabilidade, pois operam simultaneamente. Entretanto, de acordo com Marino (2001), o estabelecimento da interoperabilidade semântica, considerada por ele como a mais importante é o grande desafio para a promoção da integração entre recursos e informações na rede. Em seu trabalho, que visa ao tratamento da integração de recursos com a mesma semântica organizada em estruturas diferentes, Marino (2001, p. 02) aponta que há uma diferença no conceito de interoperabilidade semântica que deve ser explicada e até mesmo considerada como sendo portadora de subníveis,

Embora usualmente referenciado como um problema de interoperabilidade semântica, percebe-se a existência de dois níveis semânticos: *semântica epistemológica* e *semântica ontológica*. Semântica epistemológica foca na

⁵ Formato ou padrão de metadado específico da área de biblioteconomia responsável pela representação de recursos informacionais diversos originando um registro bibliográfico que permite o intercâmbio e localização dos recursos. Será tratado com mais detalhes no capítulo 5.

representação das associações e dependências entre os objetos do mundo real, enquanto que a semântica ontológica foca no significado preciso dos símbolos utilizados para representar objetos do mundo real.

Marino (2001) ressalta que os conflitos entre diferentes formas de organização da informação estão relacionados com os problemas de interoperabilidade semântica epistemológica, pois irá tratar da representação das associações e dependências entre os recursos na rede, ou seja, dos recursos que se relacionam entre si. Isso ocorrerá por meio do uso de arquiteturas de metadados. Já a interoperabilidade semântica ontológica, que trata dos significados precisos dos símbolos que representam um recurso, ou seja, do significado dos termos que representam um recurso, será estabelecida por meio das ontologias, que foi tratada no capítulo 2, mas que atuará conjuntamente com a arquitetura de metadados e os metadados.

4.1 ARQUITETURAS DE METADADOS: estabelecimento de interoperabilidade na Web

Os metadados trazem diversas vantagens para os usuários, pois por meio de uma representação padronizada dos recursos informacionais disponíveis em meio eletrônico, proporcionam o acesso mais amplo aos conteúdos, facilitam a busca, integram e compartilham recursos heterogêneos (GILLILAND-SWETLAND, 1999; ORTIS-REPISO JIMÉNEZ, 1999), permitindo ainda o controle e a administração de recursos, oferecendo informações sobre a utilização do recurso, informações sobre disponibilidade do recurso, além de promover a disseminação do recurso por meio da descrição de seu conteúdo. Sendo assim, a necessidade de seu uso na Web vem sendo divulgada como uma forma de representação importante para a recuperação das informações na Internet.

Diante do desenvolvimento de vários formatos de metadados é preciso usar ferramentas que integrem estes diversos tipos como as arquiteturas de metadados que permitirão a interoperabilidade na rede.

Desenvolvidas para garantir a interoperabilidade entre diversos padrões de metadados, as arquiteturas de metadados possuem o seguinte propósito, “[...] representar e dar suporte ao transporte de uma variedade de esquemas de metadados em ambiente distribuído, promovendo interoperabilidade nos três níveis (sintático, estrutural e semântico) [...]” (MOURA, 2002a).

Essas arquiteturas proporcionam suporte à codificação e ao transporte, ou seja, a interoperabilidade de metadados distintos, através de estruturas flexíveis. Segundo Ianella (1998) e Barreto (1999), o propósito da arquitetura de metadados é promover a codificação e transporte dos mais variados tipos de metadados, bem como a interoperabilidade nos níveis: semântico, sintático e estrutural.

De modo geral, as arquiteturas de metadados possibilitam que a sintaxe do metadados varie conforme requisições semânticas e práticas de uma determinada comunidade, que a responsabilidade pela gerência do metadado ocorra segundo os interesses da comunidade de especialistas, a proporção da interoperabilidade semântica e da extensibilidade de modo que as ferramentas de pesquisa possam acessar e manipular metadados de forma seletiva; o controle, de forma independente do acesso a conjuntos de metadados distintos que se referem a um mesmo objeto; e acomodam, de forma flexível, novos conjuntos de metadados, sem exigir mudanças nos metadados existentes, nem nos programas que os utilizam (BARRETO, 1999).

A exemplo dos padrões de metadados, também foram criadas várias arquiteturas de metadados, que de modo geral, apresentam o mesmo requisito: promover a interoperabilidade entre padrões de metadados distintos. Dentre as arquiteturas existentes que estão sendo

propostas para a implementação na Web, podemos citar: Arquitetura Warwick, MCF – Meta Content Framework, RDF – Resource Description Framework e arquitetura de modelagem de quatro níveis.

Dentre as arquiteturas citadas, a RDF vem sendo indicada e recomendada pelo W3C como a arquitetura que melhor promove a interoperabilidade na rede, pois unida aos metadados, ao uso da linguagem XML e às ontologias, consegue promover, de modo flexível, a interoperabilidade nos três níveis abordados nesta pesquisa: semântico, estrutural e sintático. Serão tratadas a seguir algumas das características dessa arquitetura.

4.2 ARQUITETURA RDF: características e estrutura

Desenvolvida pelo W3C (World Wide Web Consortium) a arquitetura RDF, de acordo com Lassila e Swick (1999), constitui-se como uma arquitetura para processar metadados e promover a interoperabilidade entre aplicações que trocam informações na Web. A RDF possibilita o processamento automatizado de recursos em várias áreas do conhecimento, tais como: na descoberta de recursos, descreve as relações entre recursos representados na rede, auxilia os agentes de software na troca e compartilhamento de informações, entre outras aplicações (LASSILA, SWICK, 1999).

Seu objetivo, de acordo com Barreto (1999, p. 103), é “[...] prover interoperabilidade entre aplicações que necessitam trocar informações estruturadas na Web”, tendo também a potencialidade para promover a interoperabilidade entre recursos que não estão vinculados a um domínio específico ou plataforma computacional e descritos por padrões de metadados diferentes (MARINO, 2001).

De acordo com Cunha (2002), o principal objetivo da arquitetura RDF,

[...] é definir um mecanismo para descrever recursos não vinculados a um domínio específico de aplicação. RDF facilita o intercâmbio de informações, que podem ser interpretadas por máquinas, entre aplicativos via Web, permite adicionar semântica formal para a Web e também, o compartilhamento de conhecimento.

A arquitetura RDF, em sua criação, recebeu influência de várias fontes e comunidades de estudos e pesquisa. As principais estão relacionadas às comunidades que buscam a padronização da Web pelo uso de metadados, estruturação de documentos pelo uso de linguagens de marcação como HTML, XML e SGML; áreas de tecnologia: programação orientada a objeto, linguagem de modelagem, banco de dados, além das áreas de biblioteconomia, representação do conhecimento (MARINO, 2001; MOURA, 2002a).

Segundo Lassila e Swick (1999), o RDF é uma ferramenta básica para promover interoperabilidade entre recursos na Web por meio do processamento de metadados. Por esse motivo, podemos ver que a aplicação da arquitetura RDF nas áreas acima trará uma grande vantagem nos aspectos relacionados à representação descritiva dos recursos, bem como a sua localização e interoperabilidade na Web.

Estruturalmente, podemos dizer que a arquitetura RDF é composta de duas especificações: o modelo RDF básico e o RDF *Schema* que irão proporcionar a flexibilidade necessária para a interoperabilidade dos dados na rede.

A) O MODELO RDF BÁSICO

De acordo com Marino (2001, p. 33) podemos dizer que "Tecnicamente, RDF não é uma linguagem, mas um modelo de dados para descrição de recursos com mais semântica, através da adoção de metadados". Sendo assim, para representar os recursos esse modelo é composto de quatro tipos de objetos que serão descritos a seguir, conforme aponta Marino (2001, p. 33):

- **Resources:** representam o universo de objetos que podem ser descritos. Todo recurso necessita de um Uniform Resource Identifier (URI) associado [...].
- **Literals:** representam os tipos de dados que o valor de uma propriedade pode assumir. Os tipos mais usuais de literais são os do tipo string.
- **Properties:** representam os aspectos do recurso a serem descritos. Propriedades podem ser visualizadas como atributos de recursos e neste sentido correspondem a pares de atributo-valor. Propriedades também são utilizadas para descrever relacionamentos entre recursos. Neste sentido, o modelo de dados RDF se assemelha ao modelo de Entidade-Relacionamento. Cada propriedade tem um significado específico, definem seus valores permitidos, os tipos de recursos que podem descrever, e seus relacionamentos com outras propriedades.
- **Statements:** representam a relação entre recurso, uma de suas propriedades e o valor que essa propriedade pode assumir.

Os *statements*, elementos básicos para a construção do modelo RDF, são constituídos por triplas (predicate, [subject], [object]), em que *predicate* seria a propriedade, *subject* o recurso e *object* o valor da propriedade. Essa tecnologia permite que valores sejam misturados, garantindo assim, "[...] maior flexibilidade ao modelo na representação de estruturas mais complexas" (MARINO, 2001, p. 34). O modelo de dados RDF também pode ser representado na forma de grafos "[...] um conjunto de nós conectados por arcos rotulados, onde os nós representam os recursos Web e os arcos representam as propriedades destes recursos" (MARINO, 2001, p. 35). Na Figura 10 temos um exemplo que demonstra como seria um recurso estruturado pelo modelo de dados RDF e representado na forma de grafo e tripla (MARINO, 2001, p. 35):

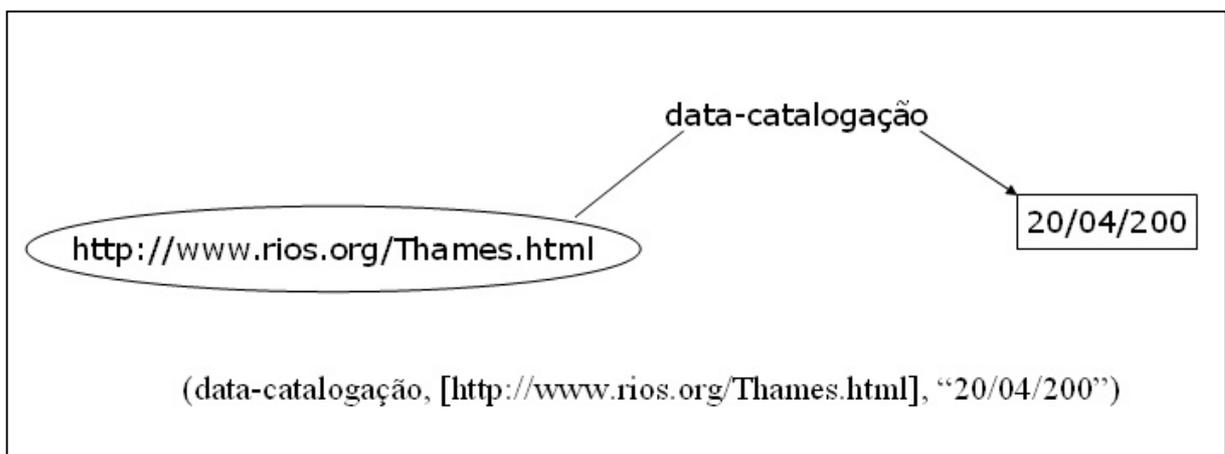


FIGURA 10: Representações de um statement: grafo e tripla.

FONTE: Marino (2001, p. 35).

A Figura 10 ilustra um *statement* que representa a seguinte estrutura: um recurso identificado pela URI do tipo URL “<http://www.rios.org/Thames.html>”, apresenta uma propriedade (data-catalogação), que apresenta o seguinte valor (20/04/2000). Em outras palavras, esse *statement* expressa a seguinte relação: “o documento <http://www.rios.org/Thames.html> foi catalogado em 20/04/2000” (MARINO, 2001, p. 35).

Além dos *statements* existem outros componentes do modelo básico RDF, que devem ser mencionados, são eles: a XML como linguagem de especificação da sintaxe RDF, definição de tipos, mecanismos de reitificação e definições de coleções, conforme será visto a seguir.

a) XML COMO LINGUAGEM DE ESPECIFICAÇÃO DA SINTAXE RDF

A possibilidade de expressar o modelo de dados RDF pela linguagem XML é um dos principais aspectos que faz com que esta arquitetura tenha sido recomendada como ideal pelo W3C para promover a interoperabilidade na Web. Isso ocorre, pois a XML é hoje uma linguagem “[...] amplamente aceita no contexto da interoperabilidade sintática” e promove, por meio de seus *Namespaces*, a mistura (intercâmbio) “[...] de diferentes padrões de metadados para compor descrições de recursos dentro de um mesmo documento” (MARINO, 2001, p. 36).

Duas sintaxes são propostas para expressar o modelo de dados RDF em XML: a “[...] serializada, que expressa toda a potencialidade do modelo RDF; e a abreviada, que inclui construtores adicionais para expressar de forma mais compacta o modelo RDF” (MARINO, 2001, p. 36). Entretanto, os dois tipos de sintaxes promovem descrições equivalentes e produzem o mesmo modelo de dados RDF. Essas descrições podem ser visualizadas na

Figura 11 como exemplo do modelo RDF serializado (MARINO, 2001, p. 36) e na Figura 12 o modelo RDF abreviado (MARINO, 2001, p.37):

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://www.recursoshidricos/esquema#"
  <rdf:Description about="http://www.rios.org/Thames.html">
    <s:data-catalogação>20/04/2000</s:data-catalogação>
  </rdf:Description>
</rdf:RDF>
```

FIGURA 11: Serialização em XML de descrições RDF.
FONTE: Marino (2001, p.36).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://www.recursoshidricos/esquema#"
  <rdf:description about="http://www.rios.org/Thames.html"
  s:data-catalogação="20/04/2000"/>
</rdf:RDF>
```

FIGURA 12: Serialização em XML abreviada de descrições RDF.
FONTE: Marino (2001, p.37).

Segundo Marino (2001, p. 36-37), podemos explicar os exemplos das Figuras 11 e 12 da seguinte forma:

A primeira linha do código indica o documento XML e a versão da linguagem. A segunda demarca o trecho RDF do documento e indica, com os prefixos “rdf:” e “s:”, a localização dos vocabulários que definem os elementos utilizados. As demais linhas representam a declaração RDF que descreve o documento, com marcadores precedidos dos prefixos “rdf:” e “s:”, cuja semântica é descrita no vocabulário associado ao prefixo. Assim, o marcador “rdf: Description about” indica que haverá uma descrição referente ao documento identificado pela URI <http://www.rios.org/Thames.html>, e que a semântica do elemento *Description* encontra-se definida no vocabulário associado ao prefixo “rdf:”. O marcador “s: data-catalogação” indica que o documento tem uma propriedade chamada “data-catalogação”, cujo valor é 20/04/200 e cuja semântica está definida no vocabulário associado ao prefixo “s:”.

b) DEFINIÇÃO DE TIPOS

Além dessas características, o modelo de dados RDF apresenta algumas primitivas importantes que são utilizadas para melhor descrever os recursos. Um exemplo seria a primitiva *rdf: type*, que permite: indicar o tipo de dados de certo recurso, estabelecendo uma relação de instanciação entre dois elementos (um elemento é instância do outro). De acordo com Marino (2001, p. 38) "Esse mecanismo é responsável por permitir inserir, em uma mesma descrição, dado e metadado". A definição de tipos pode ser visualizada na forma de grafo, conforme a Figura 13 e expressa em XML, conforme a Figura 14 (MARINO, 2001, p. 38):

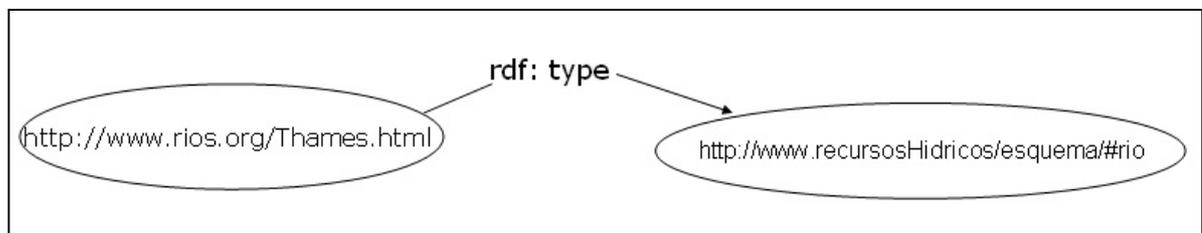


FIGURA 13: Definição de tipos em RDF.
FONTE: Marino (2001, p. 38).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:s="http://www.recursoHidricos/esquema#">
  <rdf:Description about="http://www.rios.org/Thames.html">
    <rdf:type resource="http://www.recursoHidricos/esquema/#rio"/>
    <s:data-catalogação>20/04/2000</s:data-catalogação>
  </rdf:Description>
</rdf:RDF>
```

FIGURA 14: Definição de tipos em RDF expressa em XML.
FONTE: Marino (2001, p. 38).

De acordo com o exemplo da Figura 13, a primitiva *rdf: type* especifica que o recurso “Thames” é do tipo “rio”. Expresso em XML, esse mesmo exemplo, ilustrado pela Figura 14, apresenta o indicador de fragmento (#) que foi incluído na primitiva *rdf: type*, para indicar que todas as propriedades estão relacionadas somente com um componente contido no recurso e não ao recurso todo (MARINO, 2001).

c) MECANISMO DE REITIFICAÇÃO

O mecanismo de reitificação permite uma aproximação entre níveis diferentes de abstração de dados em um nível comum. Esse mecanismo permite que qualquer *statement* seja considerado como um recurso "Desta forma é possível aninhar descrições obtendo assim, descrição sobre descrição, requisito fundamental em gerência de metadado" (MARINO, 2001, p. 39). Essa característica também é denominada "descrição de ordem maior". A reitificação em RDF significa expressar um *statement* com quatro propriedades, conforme a explicação a seguir:

- *Subject*: identifica o recurso sendo descrito pelo statement modelado.
- *Predicate*: identifica a propriedade original no statement modelado.
- *Object*: identifica o valor da propriedade no statement modelado.
- *Type*: descreve o tipo do novo recurso. Todos os statements reitificados são instâncias de "rdf: statement" (MARINO, 2001, p. 39).

O mecanismo de reitificação pode ser representado na forma de grafo, conforme a Figura 15 (MARINO, 2001, p. 39):

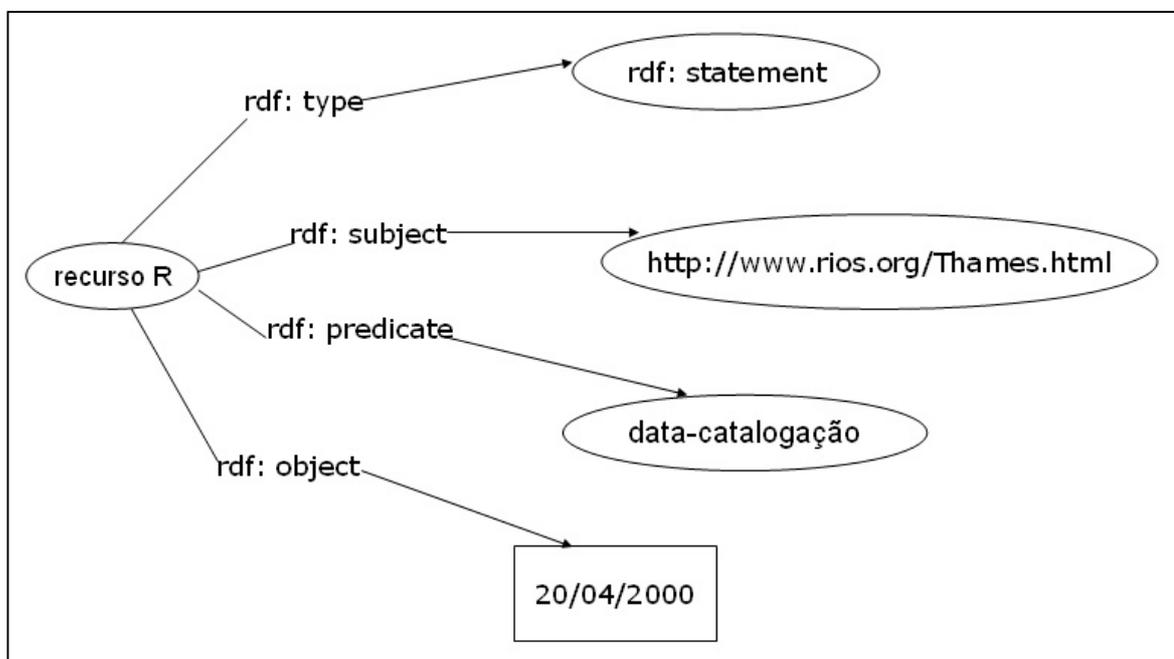


FIGURA 15: Reificação de um statement RDF.

FONTE: Marino (2001, p. 39).

O exemplo da Figura 15 ilustra uma descrição sobre descrição, onde um recurso R, que é do tipo *statement*, está relacionado com a descrição de outro recurso modelado por outro *statement*, onde *subject* “<http://www.rios.org/Thames.html>” seria esse recurso, *predicate* “data-catalogação” a propriedade do recurso e *object* “20/04/2000” o valor da propriedade. É possível também estabelecer asserções sobre o recurso, conforme o exemplo da Figura 16 e 17 (MARINO, 2001, p. 40):

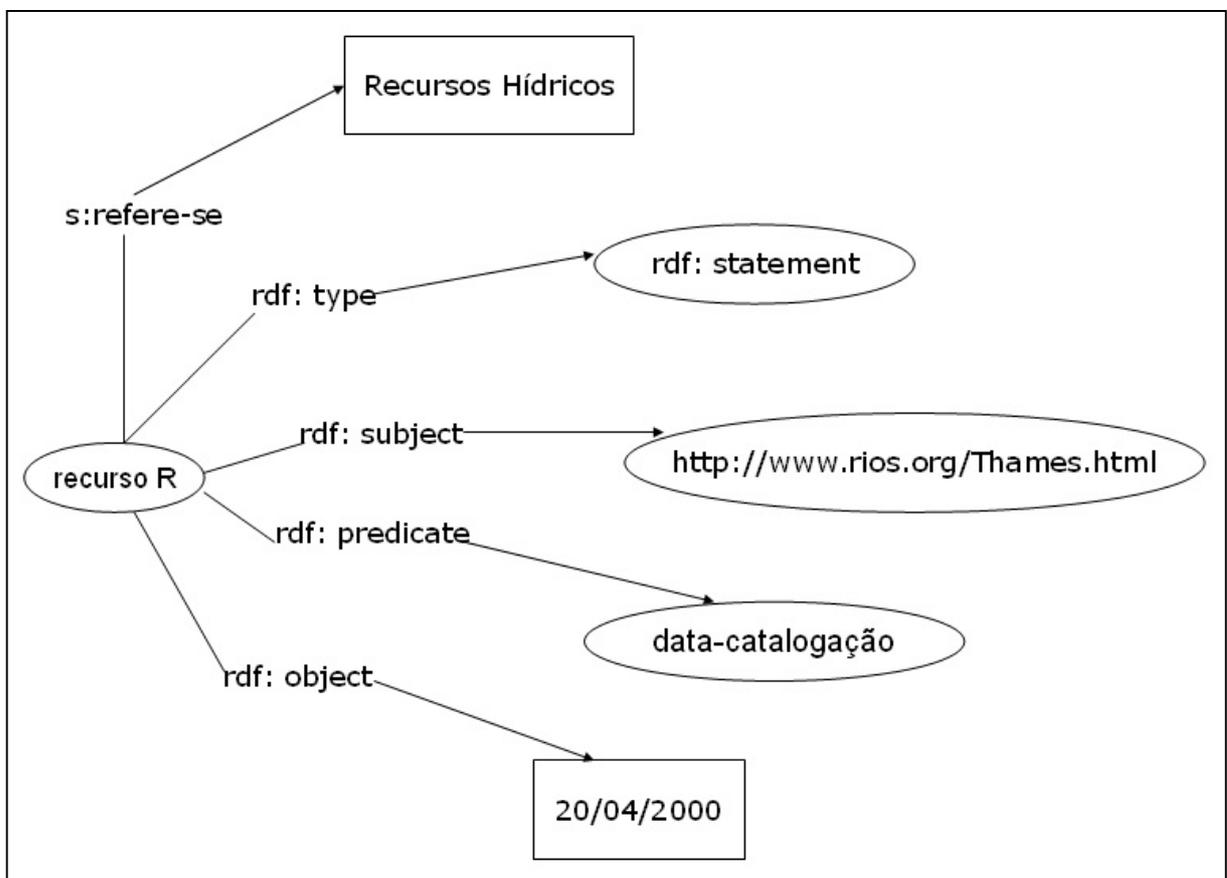


FIGURA 16: Asserções sobre um statement.
FONTE: Marino (2001, p. 40).

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://www.recursoHidricos/esquema#">
  <rdf:Description>
    <rdf:subject resource="http://www.rios.org/Thames.html"/>
    <rdf:predicate resource="http://www.recursoHidricos/esquema#data-
catalogação"/>
    <rdf:object>20/04/2000</rdf:object>
    <rdf:type resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#
Statement"/>
    <s:refere-se>Recursos Hídricos</s:refere-se>
  </rdf:Description>
</rdf:RDF>

```

FIGURA 17: Reitificação expressa em XML.

FONTE: Marino (2001, p. 41).

Após a reitificação é possível fazer asserções sobre o *statement* de ordem maior, conforme apontam as Figuras 16 e 17, onde o documento “http://www.rios.org/Thames.html”, catalogado em “20/04/200”, refere-se a propriedade “Recursos Hídricos” (MARINO, 2001).

d) DEFINIÇÃO DE COLEÇÕES

Ainda segundo Marino (2001, p. 41) “[...] o modelo de dados RDF oferece mecanismos que possibilitam a criação de coleções de recursos ou valores, atendendo a situações onde o valor de uma propriedade é um conjunto de valores ou de recursos”. O modelo RDF possui os seguintes tipos básicos que possibilita a criação de coleções:

[...] *bag*, que representa uma lista não ordenada de recursos ou valores; *sequence*, que representa uma lista ordenada de recursos ou valores; e *alternative*, que representa uma lista de valores alternativos para o valor de uma propriedade. Valores repetidos são possíveis somente nas coleções do tipo *bag* e *sequence* (MARINO, 2001, p. 41-42).

Essa declaração de coleções pode ser visualizada na Figura 18 em forma de grafo e na Figura 19 expressa em XML, conforme os exemplos de Marino (2001, p. 41 e 42):

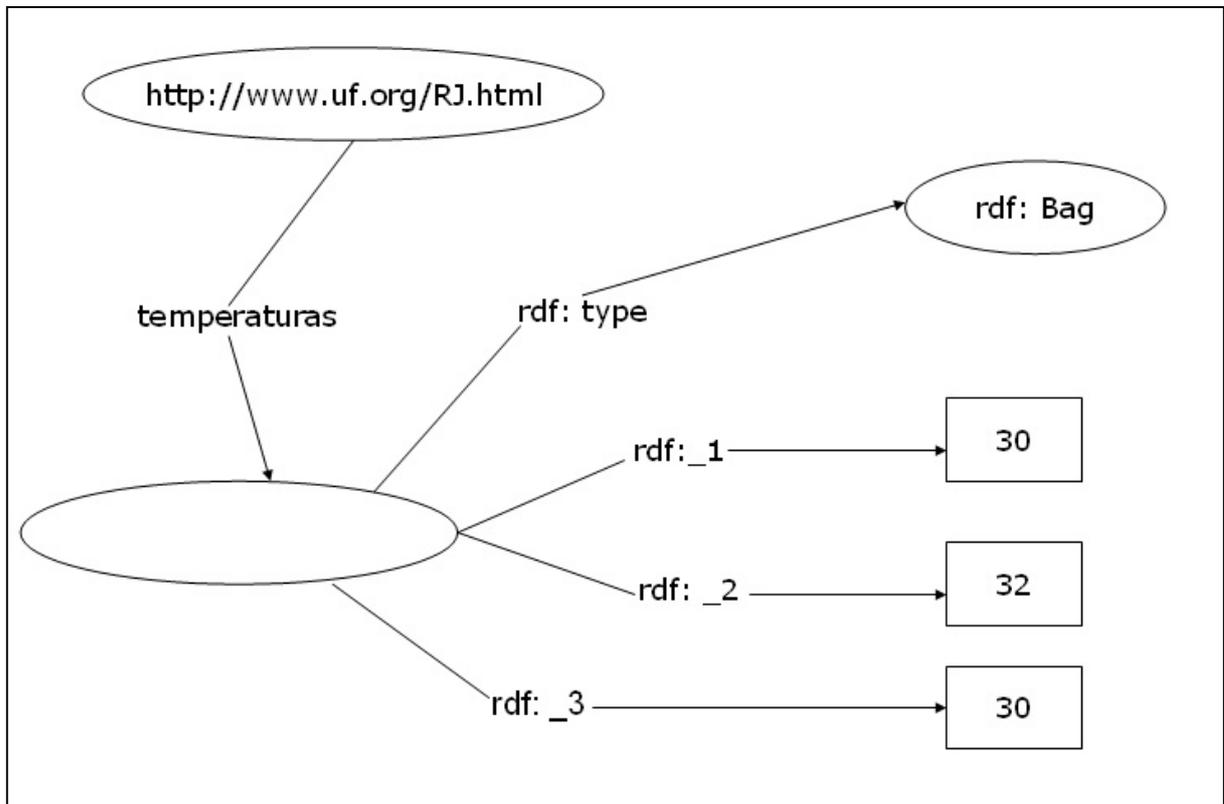


FIGURA 18: Coleção Bag listando as medidas de temperaturas de uma região.
 FONTE: Marino (2001, p. 41).

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:s="http://www.meioambiente.projeto/esquema#">
  <rdf:Description about="http://www.uf.org/RJ.html">
    <s:temperaturas>
      <rdf:Bag>
        <rdf:li
resource="http://www.resultados.experimento/temperaturas/30"/>
        <rdf:li
resource="http://www.resultados.experimento/temperaturas/32"/>
        <rdf:li
resource="http://www.resultados.experimento/temperaturas/30"/>
      </rdf:Bag>
    </s:temperaturas>
  </rdf:Description>
</rdf:RDF>

```

FIGURA 19: Uma coleção Bag descrita em XML
 FONTE: Marino (2001, p. 42).

As Figuras 18 e 19 mostram que: a coleção RDF do tipo `rdf: Bag` ilustra que a Federação RJ: “`http://www.uf.org/RJ.html`” possui um conjunto de valores de medidas de temperaturas. O `rdf: type` indica o tipo de coleção: uma instância `rdf: Bag`. Cada componente da coleção é rotulado de forma única pelos elementos do conjunto de ordinais (denominado Ord.) e representados por `rdf:_1`, `rdf_2`, ...`rdf_n`. (MARINO, 2001).

B) RDF SCHEMA

O *RDF Schema* complementa o modelo RDF básico, por meio de um mecanismo,

[...] que provê um sistema de tipos básicos para uso em modelos RDF, que aliado aos mecanismos de reitificação e namespaces, permite que comunidades de descrição de recursos possam criar e compartilhar seus próprios vocabulários (MARINO, 2001, p.42).

O *RDF Schema* é um mecanismo que promove na Web a interoperabilidade em nível semântico (epistemológico), pois “[...] representa a definição de um conjunto de propriedades com a semântica correspondente de um recurso” (MARINO, 2001, p. 41-43). Isso ocorre por meio da definição de propriedades (atributos) nas classes onde os recursos se enquadram, facilitando as descrições dos recursos em ambiente Web . O *RDF Schema* é constituído por: *Class* (classe), *subClass* (subclasse) e *Resource* (recurso).

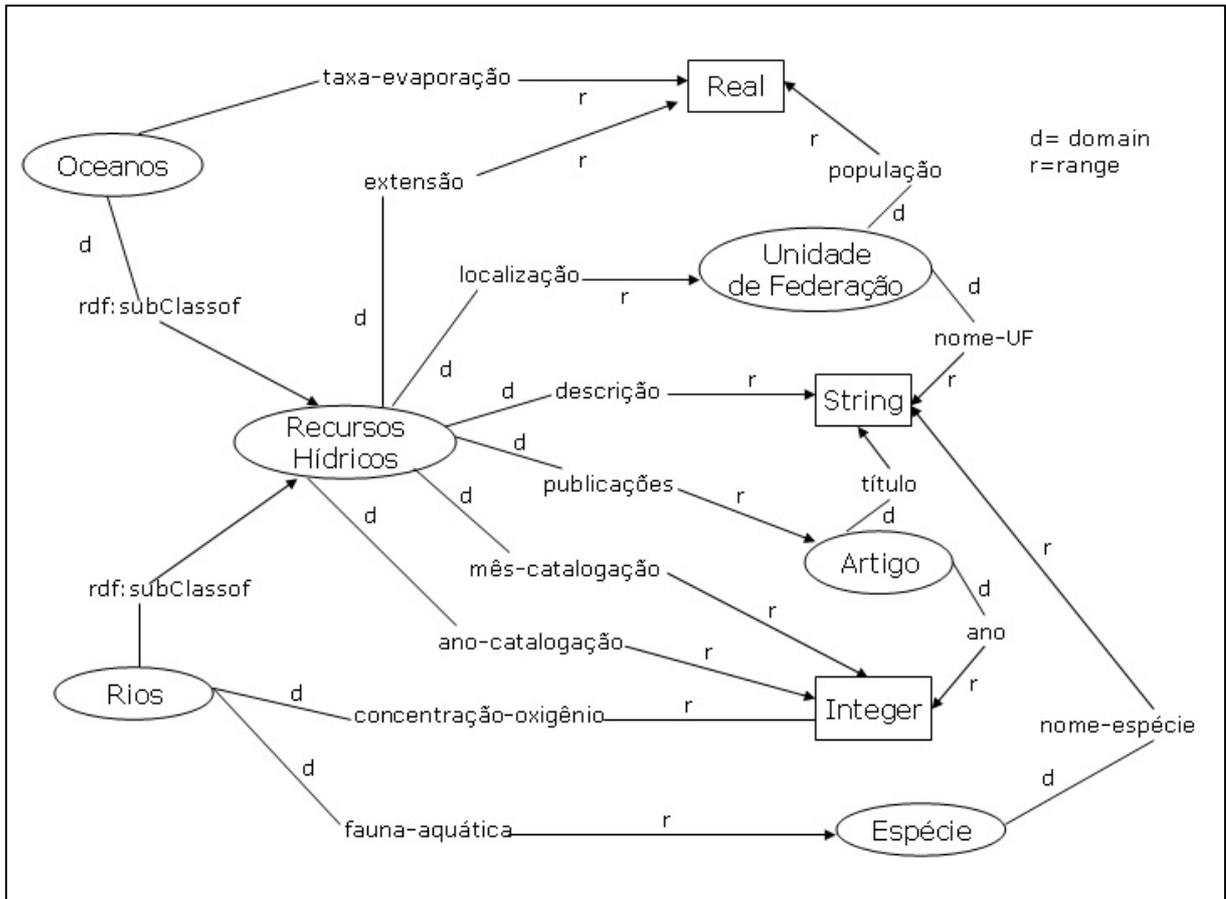
- Classe: é um recurso que comporta uma grande extensibilidade, pois “[...] se pode herdar as definições de esquemas já existentes, especializando os metadados de uma determinada comunidade, provendo assim o reuso e o compartilhamento destes esquemas” (MARINO, 2001, p. 43).
- Propriedades: “[...] possibilitam expressar relacionamentos entre Classes e suas instâncias ou superclasses. Relacionamentos entre propriedades também são permitidos, obtendo-se assim, uma hierarquia de propriedades” (MARINO, 2001, p. 44).
- Restrições: “O mecanismo permite associar restrições junto às propriedades de um recurso” (MARINO, 2001, p. 45). Em geral essas restrições estão relacionadas ao

domínio de uma propriedade e os valores que uma propriedade pode assumir (MARINO, 2001).

É importante lembrar que o RDF *Schema* possibilita apenas relações binárias entre recursos e propriedades e que não possibilita a interoperabilidade semântica em todos os níveis (epistemológico e ontológico), sendo necessário ser associado à ontologia para promover efetivamente a interoperabilidade nestes dois níveis. Sobre essa característica da arquitetura RDF, Marino (2001, p. 48) aponta que,

O mecanismo RDF Schema tem sido associado à modelagem ontológica de domínios, à medida que permite, através de um vocabulário distinto, a definição de modelos de objetos com semântica completamente definida para um domínio particular de interesse. Entretanto, conforme apresentado, este mecanismo provê somente uma semântica estrutural, permitindo a definição de um conceito em termos de suas propriedades, das restrições impostas e estas propriedades, dos relacionamentos entre estas propriedades e dos relacionamentos com outros conceitos. A modelagem de axiomas ontológicos, responsável por promover uma maior semântica conceitual, não é contemplada pela tecnologia RDF Schema.

Na Figura 20 temos um exemplo de um *Schema* RDF apresentado na forma de grafo (MARINO, 2001, p. 47):

FIGURA 20: Grafo de um *Schema* RDF.

FONTE: Marino (2001, p. 47).

O *Schema* RDF apresenta como Classe (tema central) “Recursos Hídricos” e subclasses “Rios” e “Oceanos”. As relações entre os recursos são estabelecidas por intermédio das propriedades, como por exemplo, a relação de associação entre o recurso “Rios” e o recurso “Espécie” pela propriedade “fauna-aquática”. Já as restrições são associadas às propriedades, como por exemplo, a propriedade “população” que tem como restrição de domínio a classe “Unidade de Federação” e como restrição de valor a classe de números “Reais” (MARINO, 2001). Na Figura 21 temos o mesmo exemplo expresso em XML (MARINO, 2001, p. 48-49):

```

<?xml version="1.0" encoding="UTF-8"?>
<!-- edited with XML Spy v3.5 (http://www.xmlspy.com) by M Teresa Marino
(private) -->
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:tipos="http://www.w3.org/2000/03/example/classes#">
  <rdf:Description ID="Recursos Hídricos">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
  </rdf:Description>
  <rdf:Description ID="Unidade de Federação">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
  </rdf:Description>
  <rdf:Description ID="Espécie">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
  </rdf:Description>
  <rdf:Description ID="Artigo">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
  </rdf:Description>
  <rdf:Description ID="Oceanos">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
    <rdfs:subClassOf rdf:resource="#Recursos Hídricos"/>
  </rdf:Description>
  <rdf:Description ID="Rios">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
    <rdfs:subClassOf rdf:resource="#Recursos Hídricos"/>
  </rdf:Description>
  <rdf:Description ID="taxa-evaporação">
    <rdfs:domain rdf:resource="#Oceanos"/>
    <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#Real"/>
  </rdf:Description>
  <rdf:Description ID="extensão">
    <rdfs:domain rdf:resource="#Recursos Hídricos"/>
    <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#Real"/>
  </rdf:Description>
  <rdf:Description ID="localização">
    <rdfs:domain rdf:resource="#Recursos Hídricos"/>
    <rdfs:range rdf:resource="#Unidade de Federação"/>
  </rdf:Description>
  <rdf:Description ID="população">
    <rdfs:domain rdf:resource="#Unidade de Federação"/>
    <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#Real"/>
  </rdf:Description>

```

```

<rdf:Description ID="fauna-aquática">
  <rdfs:domain rdf:resource="#Rios"/>
  <rdfs:range rdf:resource="#Espécie"/>
</rdf:Description>
<rdf:Description ID="nome-UF">
  <rdfs:domain rdf:resource="#Unidade de Federação"/>
  <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#String"/>
</rdf:Description>
<rdf:Description ID="descrição">
  <rdfs:domain rdf:resource="#Recursos Hídricos"/>
  <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#String"/>
</rdf:Description>
<rdf:Description ID="publicações">
  <rdfs:domain rdf:resource="#Recursos Hídricos"/>
  <rdfs:range rdf:resource="#artigo"/>
</rdf:Description>
<rdf:Description ID="mês-catalogação">
  <rdfs:domain rdf:resource="#Recursos Hídricos"/>
  <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#Integer"/>
</rdf:Description>
<rdf:Description ID="ano-catalogação">
  <rdfs:domain rdf:resource="#Recursos Hídricos"/>
  <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#Integer"/>
</rdf:Description>
<rdf:Description ID="ano">
  <rdfs:domain rdf:resource="#Artigo"/>
  <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#Integer"/>
</rdf:Description>
<rdf:Description ID="título">
  <rdfs:domain rdf:resource="#Artigo"/>
  <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#String"/>
</rdf:Description>
<rdf:Description ID="concentração-oxigênio">
  <rdfs:domain rdf:resource="#Rios"/>
  <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#Integer"/>
</rdf:Description>
<rdf:Description ID="nome-espécie">
  <rdfs:domain rdf:resource="#Espécie"/>
  <rdfs:range
rdf:resource="http://www.w3.org/2000/03/example/classes#String"/>
</rdf:Description>
</rdf:RDF>

```

FIGURA 21: Descrição de um *Schema* RDF em RDF/XML.

FONTE: Marino (2001, p. 48-49).

a) USO DO MECANISMO NAMESPACES XML

O uso dos namespaces XML pela linguagem RDF é um mecanismo de muita importância na tentativa de promover a interoperabilidade na rede. Segundo Marino (2001, p. 49) "Através deste mecanismo é possível distinguir diferentes camadas de modelagem, bem como reusar e integrar esquemas e aplicações definidos por diferentes comunidades de descrição de recursos". Marino (2001, p. 50) aponta que,

Namespaces representam os esquemas de domínios específicos sobre os quais predicados contidos no documento RDF assumem valores. Em RDF, cada predicado utilizado em um statement precisa ser identificado univocamente por um namespace ou esquema. Desta forma, é possível compor a descrição de um recurso através de um conjunto de statements cujos predicados podem vir de diversos esquemas. Conflitos como definição de termos com mesmo nome também são evitados, uma vez que os predicados estão associados a termos de um único namespace.

Não é objetivo, neste momento, tratar da estrutura da arquitetura RDF em sua totalidade, mas apenas fornecer elementos básicos para o entendimento das potencialidades desta arquitetura.

Além dessas questões de estruturas abordadas, é necessário lembrar que a URI é uma tecnologia que irá ser importante nessa arquitetura, pois é por meio das URI(s) que os recursos serão encontrados e que possibilitará encontrar as representações descritivas destes recursos.

De acordo com Faria e Girardi (2002?), o RDF utiliza as URI(s) para "[...] codificar a informação num documento e assegura que aquelas palavras não são somente palavras, mas que possuem um significado único" e que o uso de URI(s) diferentes para conceitos diferentes garante uma maior consistência nas inferências realizadas pelos mecanismos de busca.

Outro fator que merece destaque é que a linguagem XML e a arquitetura RDF podem ser consideradas complementares, pois a sintaxe RDF utiliza a XML para expressar o significado de informação (FARIA; GIRARD, 2002?). "Enquanto que o XML define uma

estrutura, o RDF permite expressar o significado associado aos dados" (FARIA; GIRARD, 2002?).

4.3 CROSSWALKS: ferramenta para o mapeamento entre formatos de metadados

Para Cromwell-Kessler (1999), a correspondência entre padrões é um dos segredos para que o usuário possa buscar as informações desejadas na Web, pois é estabelecida uma associação entre estes padrões, permitindo que o usuário possa pesquisar e acessar as informações desejadas, por meio de uma única interface de busca. Sendo assim, uma outra ferramenta que vem se destacando, juntamente com a questão do uso de arquiteturas de metadados e que pode ser utilizada para auxiliar os processos de correspondência entre formatos de metadados distintos são os *crosswalks*.

Não é objetivo, tratar em profundidade essa nova ferramenta, mais apenas abordar, de modo resumido, o que essa ferramenta pode possibilitar na tentativa de estabelecer a interoperabilidade na rede. O *crosswalk*, que também é denominado de mapeamento, é uma ferramenta que está sendo indicado para permitir o estabelecimento de uma correspondência ou mediação entre os diversos formatos de metadados, na tentativa de minimizar a falha existente entre ambientes informacionais que utilizam formatos de metadados diferentes. Sendo assim, sua função seria mapear elementos de metadados com conteúdos semelhantes para posterior intercâmbio.

Baseado no que foi dito acima podemos verificar que essa correspondência se estabelece em nível semântico, pois trata do significado dos elementos descritivos de cada formato de metadado. Se utilizarmos como exemplo os padrões de metadados Dublin Core e

MARC 21, o mapeamento ou cruzamento estaria ocorrendo entre seus elementos descritivos, ou seja, os elementos do Dublin Core com os campos descritivos do MARC 21. Dessa forma, elementos ou campos que apresentam o mesmo conteúdo tornariam-se intercambiáveis. Portanto, o *crosswalk* possibilitaria a conversão de dados de um formato de metadados para outro, e esse mapeamento permitiria o intercâmbio informacional entre diferentes padrões de metadados e também a possibilidade de manutenção de diferentes bases de dados e uma única interface de busca, capaz de recuperar documentos em qualquer uma delas.

Como já apontado, estabelecer essa correspondência entre formatos de metadados para que o usuário possa pesquisar e acessar as informações, por meio de uma única interface de busca em qualquer base de dados é uma questão que, se efetivada, trará grandes vantagens para o sistema ou ambiente que faz uso desta tecnologia, principalmente na questão da interoperabilidade. A integração de diferentes bases de dados, bem como, a possibilidade de recuperar informações que foram descritas por padrões específicos de comunidades diferentes, possibilitaria um maior acesso às informações.

Entretanto, para que haja esse mapeamento ou correspondência entre formatos de metadados é preciso que tanto a estrutura sintática quanto a semântica sejam atendidas, mas como todo o processo de interoperabilidade, quando tratamos de semântica o que parece fácil se complica.

O *crosswalk* é uma ferramenta útil para o mapeamento das informações, para o estabelecimento de interoperabilidades entre representações e sistemas, para a agilização do processo de representação descritiva e que, portanto, é uma ferramenta útil para ser adotada pelos profissionais que trabalham com a representação da informação.

4.4 INTEROPERABILIDADE NA REDE: trabalho conjunto entre aplicações e ferramentas

O surgimento de diversos formatos de metadados possibilitou uma padronização na descrição, mas ao mesmo tempo ocasionou uma incompatibilidade entre eles (os formatos). Nesse sentido, as arquiteturas de metadados surgem com a possibilidade de oferecer "[...] a flexibilidade necessária em ambientes heterogêneos, permitindo que recursos possam ser descritos seguindo diversos padrões, aproveitando assim o que cada um tem de melhor em termos de semântica descritiva" (MARINO, 2001, p. 17). Por esse motivo podemos considerar que a interoperabilidade desempenhará um papel fundamental no compartilhamento e intercâmbio de informações em qualquer ambiente em meio eletrônico.

A interoperabilidade em si é algo que parece simples, pois é a troca ou intercâmbio de recursos ou informações disponibilizadas em ambientes heterogêneos. Entretanto, sua prática é mais complexa, pois envolve outras questões e ferramentas que devem seguir normas e padrões altamente detalhados para que possa ser promovida essa interoperabilidade. Tais ferramentas ou tecnologias irão trabalhar de modo simbiótico e são as seguintes: formatos de metadados para garantir uma boa representação descritiva dos recursos; o uso de uma linguagem de marcação, hoje a XML, que garante não só o armazenamento dos dados, mas garante também a promoção da interoperabilidade por ser uma linguagem flexível e que permite a representação de relacionamentos entre dados; o uso da arquitetura de metadados para promover a interoperabilidade nos três níveis, semântico, estrutural e sintático e por fim, o uso de ontologias, que irão tratar da conceituação semântica do significado dos dados.

A arquitetura RDF é uma recomendação do W3C para o estabelecimento da interoperabilidade semântica, sintática e estrutural. No entanto, quando falamos de

interoperabilidade semântica, a arquitetura RDF se apresenta apenas como uma solução parcial, pois não oferece mecanismos suficientes para definição de axiomas (regras) mais genéricos, nem mecanismos suficientes para definir o significado de um conceito, independente do domínio ou comunidade que ele pertence (MARINO, 2001).

Nesse sentido, essa complementação é dada pelo uso de linguagens de representação do conhecimento ou das ontologias. Atuando juntamente com a arquitetura de metadados RDF, as ontologias possibilitam estender e complementar a arquitetura para que a interoperabilidade semântica seja realizada na Web.

Dentro da proposta da Web Semântica, a arquitetura RDF se apresenta como uma ferramenta importante e está sendo indicada pela W3C como necessária para o estabelecimento de relacionamentos entre recursos e suas propriedades, bem como para o estabelecimento de interoperabilidades na Web Semântica. Todavia, como já foi dito por Marino (2001, p. 63),

Propostas como o RDF se mostram adequadas no sentido de prover uma solução para a interoperabilidade semântica epistemológica. Entretanto, somente formalismos como ontologias podem lidar com problemas de interoperabilidade semântica ontológica.

Como pode ser visto, a arquitetura de metadado é fundamental para o estabelecimento de interoperabilidade na rede, mas para que seja estabelecida uma rede de conhecimento é preciso que além do trabalho conjunto das ferramentas estudadas até agora: ontologias, linguagem de marcação XML e arquitetura de metadados, é preciso haver uma base que forneça a representação informacional necessária aos recursos. Essa representação é fornecida pelos metadados que serão tratados no próximo capítulo.

5 METADADOS PARA A REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS NA WEB SEMÂNTICA

Conforme o foi apontado nos capítulos anteriores pode-se dizer que a Web Semântica propõe uma evolução não só nos processos de recuperação da informação, mas também na forma como os recursos informacionais são tratados e disponibilizados na rede.

De acordo com Berners-Lee o poder efetivo da Web Semântica seria percebido pelo processamento e troca de informações encontradas em fontes diversas. Por meio dos agentes de software, que tem a capacidade de processar e trocar informações dispersas em vários repositórios, teríamos acesso a uma grande quantidade de recursos que correspondem a nossas necessidades. (BERNERS-LEE, HENDLER, LASSILA, 2001). Para Souza e Alvarenga (2004, p. 134),

O projeto da Web Semântica, em sua essência, é a criação e implantação de padrões (*standards*) tecnológicos para permitir este panorama, que não somente facilite as trocas de informações entre agentes pessoais, mas principalmente estabeleça uma língua franca para o compartilhamento mais significativo de dados entre dispositivos e sistemas de informação de uma maneira geral.

Entretanto, para que isso aconteça, é necessário ter acesso padronizado não só às ferramentas tecnológicas, mas também uma padronização nas estruturas de representação dos dados (linguagens para armazenamento dos dados, representação dos dados e interoperabilidade). Isso já pode ser verificado nas camadas que compõe a arquitetura da Web Semântica e nos capítulos tratados até agora.

Mesmo sendo necessário estabelecer a semântica dos dados por meio das ontologias, estabelecer uma estrutura por meio de linguagens de marcação como a XML e a interoperabilidade na rede por meio das arquiteturas de metadados, é preciso que os agentes de software tenham como base uma coleção de recursos devidamente representados, para que

todas estas tecnologias realizem suas tarefas de modo complementar e assim possam estabelecer o funcionamento da Web Semântica. Na Figura 22 podemos ver ilustrado como seria este funcionamento:

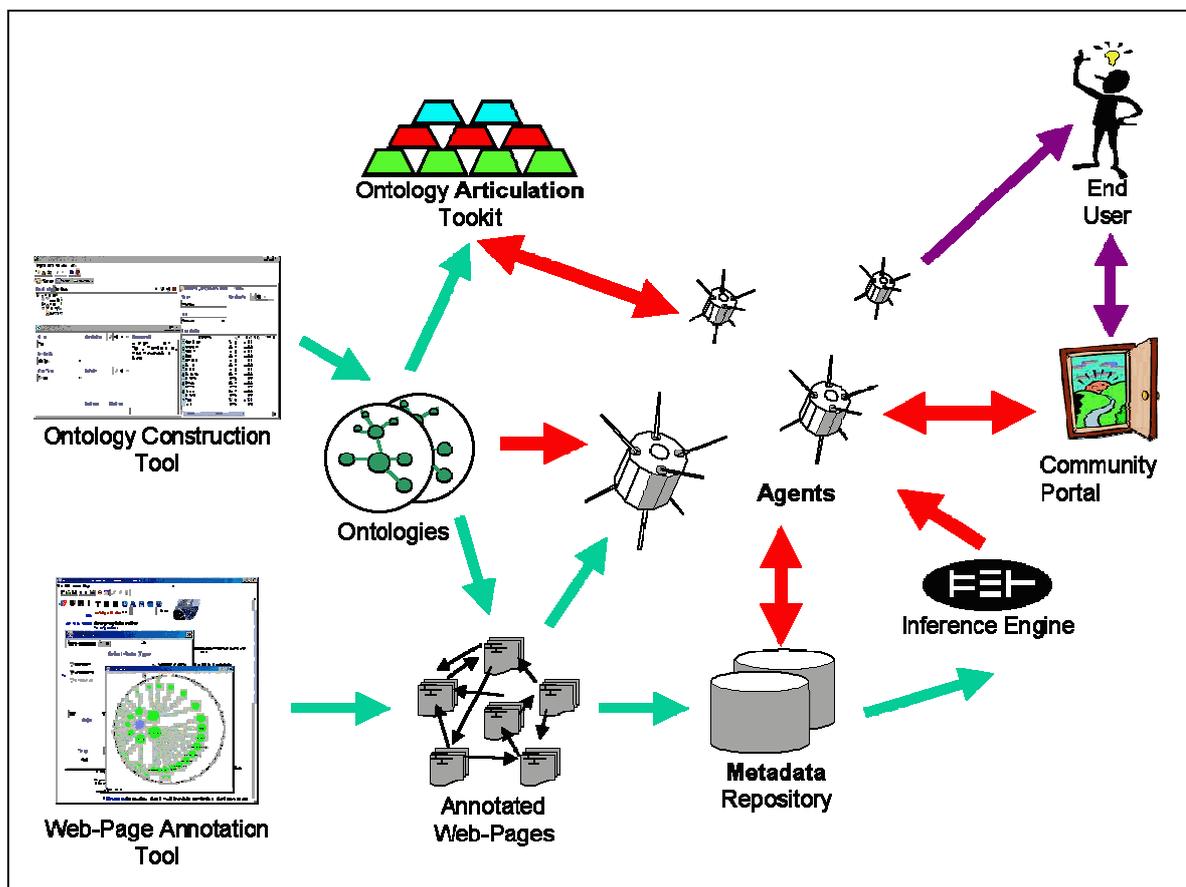


FIGURA 22: Funcionamento da Web Semântica de acordo com a SemanticWeb.org.
 FONTE: Semantic Web (2002). Disponível em: <<http://www.semanticweb.org/about.html>>.

A Figura 22 ilustra como seria o funcionamento da Web Semântica. Por esse esquema, os usuários finais teriam acesso a Web Semântica por meio de portais comunitários, ou até mesmo portais corporativos. Os agentes de softwares, presentes nos mecanismos de busca e inferência, tem acesso a um repositório de metadados constituído pelas páginas representadas por padrões de metadados e anotadas semanticamente pelas ontologias. Os agentes trabalham formando uma espécie de sinergia, ou seja, trocam entre si as informações extraídas dos repositórios de metadados e os significados definidos nas ontologias, assim, oferecem uma recuperação mais eficaz, pois são capazes de “compreender” o conteúdo dos recursos. Desse

modo, a Figura 22 apresenta duas questões que devem ser consideradas: a questão da representação e indexação dos recursos, com as ferramentas que possibilitarão o acesso aos repositórios de metadados e com a definição do significado dos dados, possibilitando o acesso mais amplo a um conteúdo semântico compartilhado em comunidades de interesse; e a questão da recuperação e uso dos recursos informacionais, que será realizada de modo mais eficiente pelo fato dos agentes de software, associados aos mecanismos de busca e inferência, serem capazes de “compreender” o conteúdo dos recursos, que estão devidamente representados e anotados semanticamente (SOUZA, ALVARENGA, 2004).

Entretanto, é preciso lembrar que as tecnologias envolvidas trabalham em conjunto: os agentes conseguem proporcionar uma recuperação mais eficiente porque encontram uma coleção de recursos devidamente representados e estruturados pelas ferramentas presentes nas camadas da Web Semântica, tais como os metadados e ontologias. Além disso, todo esse funcionamento tem como apoio a linguagem XML para a estruturação dos dados e a arquitetura de metadados RDF para estabelecer a interoperabilidade dos dados.

Podemos dizer então, que a representação do conteúdo dos recursos acontecerá por meio dos metadados; os significados dos dados representados serão definidos semanticamente nas ontologias; a interoperabilidade e a relação que um termo estabelece com outro será expressa pela arquitetura RDF e todos esses dados serão estruturados pela linguagem XML.

De acordo com Souza e Alvarenga (2004), as tecnologias necessárias para implantar a Web Semântica se articulam entre si e fazem com que a Web se assemelhe a um sistema de recuperação de informações e a base para a construção deste tipo de sistema é a representação dos recursos informacionais. Por esse motivo é apontada como fundamental para a construção da Web Semântica a representação dos recursos informacionais e isto ocorrerá por meio dos metadados presentes na camada estrutura da Web Semântica (apresentada no capítulo 2, Figura 2) (BERNERS-LEE, HENDLER, LASSILA, 2001).

Autores como Berners-Lee, Hendler e Lassila (2001) afirmam que os agentes trarão mais eficiência para a recuperação de recursos informacionais na Web Semântica, pois terão acesso aos conceitos definidos semanticamente nas ontologias. E que essas definições o ajudarão a compreender a semântica embutida nos recursos ou o contexto em que se insere o recurso. Entretanto, é preciso lembrar que a Web Semântica se estabelecerá a partir de uma coleção de recursos informacionais devidamente representados. Ou seja, somente com a representação dos recursos informacionais, por meio de formatos de metadados padronizados, será possível estabelecer a base para que se desenvolva a Web Semântica. Os metadados irão representar os recursos informacionais e a semântica dos dados será definida nas ontologias, sendo assim, podemos dizer então que as ontologias funcionam como qualificadores dos metadados e não haverá estabelecimento da semântica sem a representação por metadados.

É preciso estabelecer essa diferença, pois cada tecnologia realiza uma tarefa: os metadados representam os recursos informacionais e as ontologias definem semanticamente os conceitos dos dados que representam os recursos, ou seja, dos metadados.

Sendo assim, serão tratados neste capítulo os aspectos relacionados aos metadados, tais como: conceitos, características, formatos ou padrões de metadados e sua importância.

5.1 METADADOS: origem, definição e características

Para que os recursos informacionais sejam recuperados em um sistema de informação (seja ele digital ou não) é preciso utilizar métodos de representação da informação para que ocorra a mediação entre a informação registrada (documento) e o usuário (PEREIRA, SANTOS, 1998).

Essa tarefa não é uma novidade para os profissionais da informação que se utilizam das metodologias e processos da Biblioteconomia para fazer essa mediação entre a informação registrada e o usuário. Um desses processos seria a catalogação que busca a mediação entre conhecimento codificado (informação registrada) e o usuário, por meio de formas de representação e uso de tecnologias. Conforme apontam Pereira e Santos (1998, p. 123),

O processo de catalogação pode ser identificado como meio de comunicação, um instrumento de ligação entre o usuário e o documento, um processo de representação documental que desde a antiguidade atua como instrumento de acesso à informação e ao documento a que se utiliza dos instrumentos disponíveis, numa ação que interliga a biblioteconomia e as tecnologias disponíveis, possibilitando uma rápida recuperação e disseminação da informação, proporcionando assim condições para a agilização de conhecimento.

A forma de organização e recuperação dos recursos informacionais sempre estiveram relacionadas com as tecnologias vigentes em cada época, sendo assim, hoje estão relacionadas com as tecnologias de informática. Segundo Robredo (2004) as tecnologias oferecem atualmente ótimas soluções para organizar a crescente e variada quantidade de informações disponibilizadas em diversos meios.

No entanto, à medida que aumentam as informações disponibilizadas, é preciso que além do uso de tecnologias, novas formas ou novos métodos para organizá-las sejam criados. Isso se tornou mais evidente com o uso de tecnologias de informação e comunicação, que além de possibilitar o surgimento de novos tipos de materiais, possibilitou também, novas formas de tratamento da informação.

Entretanto, apesar das tecnologias de informação e comunicação modificarem a concepção de organização, tratamento e acesso às informações, a essência do tratamento da informação vem de métodos tradicionais já estabelecidos na área da Biblioteconomia.

Siqueira (2003, p. 14) aponta que o tratamento da informação é um “[...] processo que tem como finalidade a identificação, processamento e disponibilização do conteúdo informacional de diversificados suportes documentários”.

Temos então, a catalogação como um processo para o tratamento da informação que busca o aprimoramento de suas técnicas na tentativa de melhorar a recuperação da informação (PEREIRA; SANTOS, 1998).

Atualmente, ao nos referirmos à catalogação e ao uso das tecnologias, precisamos ter uma visão ampla e objetiva do alcance desta relação. Da máquina de escrever às redes de telecomunicações, a catalogação foi evoluindo de maneira sutil e incisiva, fazendo-se presente na história da Biblioteconomia como responsável por um processo dinâmico de discussão e propostas de mutação nas formas de representação, no armazenamento e nas estruturas de busca da informação (PEREIRA; SANTOS, 1998, p. 125).

Conforme pode ser percebido, a catalogação vem acompanhando as evoluções tecnológicas na tentativa de definir ferramentas mais eficazes para a representação dos recursos informacionais, que agora estão disponíveis também em meio eletrônico. Esses recursos necessitam de métodos de representação mais específicos que os métodos tradicionais, ou seja, necessitam de novas formas de representação que atenda a suas características.

Nesse contexto, os metadados estão sendo indicados na literatura como a ferramenta capaz de proporcionar uma forma de representação dos recursos informacionais em meio eletrônico e assim proporcionar, conseqüentemente, a mediação entre o conhecimento registrado em ambiente digital e o usuário.

Mas o que são os metadados?

De acordo com Senso e Rosa Piñero (2003, p. 97, tradução nossa) o termo metadados “[...] foi cunhado por Jack Mayers na década de 60 para descrever conjuntos de dados”.

Atualmente, a palavra metadados tem sido definida, muitas vezes, como sendo “dados sobre dados”, conforme apontam Miller (1996); Souza, Catarino e Santos (1997); Milstead e Feldman (1999); Gilliland-Swetland (1999); Souza, Vendrusculo e Melo (2000); Takahashi (2000); Senso e Rosa Piñero (2002); entre outros.

Entretanto, são encontradas na literatura várias definições para o termo, resultando em uma pluralidade semântica que varia de acordo com a área em que se insere o termo e entre os profissionais que o utilizam. Isso vem ocorrendo, pois o termo metadados está sendo usado nos últimos tempos não só na Biblioteconomia, mas em todas as áreas responsáveis pela organização, tratamento e gerenciamento de recursos de informação (ROSETTO, 2003).

Nesse sentido, o termo metadados apresenta-se como um termo neutro, comum às diversas áreas do conhecimento e igualando os bibliotecários que sempre trabalharam com metadados, aos outros profissionais envolvidos atualmente com o tratamento de informações, como por exemplo, os profissionais da área de informática (SENSO; ROSA PIÑERO, 2003).

Na realidade, apesar de ser algo aparentemente novo, o objetivo e a função dos metadados correspondem às técnicas de representação dos recursos já estabelecidas na Biblioteconomia. No entanto, “O que tem mudado é o escopo de sua atuação conforme a evolução da tecnologia da informação” (IKEMATU, 2001).

Com a necessidade de denominar o tratamento da informação em meio eletrônico o termo metadados é utilizado com intensidade como um termo que permite a comunicação entre diferentes áreas que possuem o mesmo objetivo: dar tratamento às informações tendo em vista o gerenciamento informacional.

Vários estudos estão sendo feitos sobre metadados e alguns conceitos encontrados na literatura são apresentados a seguir.

Para Takahashi (2000, p. 172), metadados são,

Dados a respeito de outros dados, ou seja, qualquer dado usado para auxiliar na identificação, descrição e localização de informações. Trata-se em outras

palavras, de dados estruturados que descrevem as características de um recurso de informação.

Metadados para Ortiz-Repiso Jiménez, (1999, p. 218) pode ser definido como,

[...] um conjunto de dados que pode ser usado para descrever e representar recursos informacionais. Contém um conjunto de elementos de dados que podem ser usados para descrever o conteúdo e a localização de um recurso informacional e facilitar sua recuperação e acesso na rede.

Senso e Rosa Piñero (2002, p. 99) consideram os metadados como,

[...] toda aquela informação descritiva sobre o contexto, qualidade, condição ou características de um recurso, dado ou objeto que tem a finalidade de facilitar sua recuperação, autentificação, evolução, preservação ou interoperabilidade.

Os autores Souza, Vendrúsculo e Melo (2000, p. 93) definem metadados como “[...] dado sobre dado. É a catalogação do dado ou descrição do recurso eletrônico”.

Grácio (2002, p. 21) aponta metadados como sendo,

Comumente chamado de dados sobre dados, o termo metadados pode ser melhor descrito como um conjunto de dados chamados de elementos, cujo número é variável de acordo com o padrão, e que descreve o conteúdo de um recurso, possibilitando a um usuário ou a um mecanismo de busca acessar e recuperar esse recurso. Esses elementos descrevem informações do tipo nome, descrição, localização, formato, entre outras, que possibilitam um número maior de campos para pesquisas.

Em outras palavras, metadados é um “Conjunto de elementos que descrevem as informações contidas em um recurso, com o objetivo de possibilitar sua busca e recuperação” (GRÁCIO, 2002, p. 23).

Pode-se perceber nas definições apresentadas que de modo geral os autores consideram metadados como sendo dados estruturados e padronizados que representam um recurso informacional, com o objetivo de facilitar sua identificação, para a sua localização e recuperação.

Entretanto, antes de levantarmos mais sobre suas características é necessário deixar claro a diferença existente entre metadados e formatos de metadados. Essa diferenciação é apontada no trabalho de Rosetto (2003) que por meio de uma revisão de literatura propõe uma revisão e reconstrução de significado dos termos metadados e formatos de metadados:

Para Rosetto (2003, p. 59), metadados são,

[...] um conjunto de dados – atributos – referenciais, metodologicamente estruturados e codificados, conforme padrões internacionais, para localizar, identificar e recuperar pontos informacionais de textos, documentos e imagens disponíveis em meios digitais ou em outros meios convencionais.

Já os formatos de metadados Rosetto (2003, p. 59) considera que,

Formatos de metadados referem-se a padrões que estabelecem regras para a definição de atributos (metadados) de recursos de informacionais, para a) obter coerência interna entre os elementos por meio de semântica e sintaxe; b) promover necessária facilidade para esses recursos serem recuperados pelos usuários; c) permitir a interoperabilidade dos recursos de informação.

Baseada nas conceituações apresentadas acima podemos considerar que os metadados são conjuntos de atributos, mais especificamente dados referenciais, que representam o conteúdo informacional de um recurso que pode estar em meio eletrônico ou não. Já os formatos de metadados, também chamados de padrões de metadados, são estruturas padronizadas para a representação do conteúdo informacional que será representado pelo conjunto de dados-atributos (metadados). Em outras palavras, os formatos ou padrões de metadados podem ser considerados como formas de representação⁶ de um item documentário.

Para os profissionais da Ciência da Informação o termo metadados está relacionado com o tratamento da informação, mais especificamente às formas de representação de um recurso informacional para fins de identificação, localização e recuperação, ou seja, dados sobre catalogação e indexação que servem para organizar e tornar a informação mais acessível (GILLILAND-SWETLAND, 1999).

Miller (1996), afirma que, mesmo sem conhecer utilizamos metadados quando anotamos informações de um livro, como por exemplo: autor, título, ano, editora etc. Sendo assim, vemos que os metadados não são necessariamente digitais, e que já são utilizados há

⁶ Siqueira (2003) aponta em seu trabalho a teoria de Formas de Representação estudadas pelas Ciências Cognitivas como sendo um fundamento para a catalogação. Essas teorias estão relacionadas com a catalogação, que se utiliza de formas de representação para “[...] individualizar os itens, de forma que não sejam confundidos entre si, reunir itens por suas semelhanças, estabelecendo relações entre si, e finalmente, permitir a localização de um item específico em acervo determinado” (MEY, 1995, p. 38). A vantagem de se utilizar formas de representação é obter estruturas capazes de representar informações relevantes; a variedade nas formas de representação proporciona mais pontos de acesso facilitando o crescimento do conhecimento. (CASA (1997); PETERSON (1996), citados por SIQUEIRA, 2003).

algum tempo, como as bibliotecas que desde os anos 60 compartilham metadados com a ajuda de normas de catalogação e padrões internacionais de estruturação de conteúdos (GILLILLAND-SWETLAND, 1999). Partindo desse ponto de vista, Milstead e Feldman (1999) afirmam que “[...] padrões de informação bibliográfica, sumários, termos de indexação, e abstracts são todos substitutos do material original, portanto metadados”.

Na realidade, catalogadores e indexadores produzem e padronizam metadados há séculos, desde as primeiras tentativas de organização da informação que se conhece na história da Biblioteconomia. O que vem acontecendo recentemente é que profissionais de diversas áreas estão buscando criar outros métodos de tratamento da informação, sem saber que isso já foi feito, gerando uma variedade de padrões que muitas vezes não atende satisfatoriamente as necessidades informacionais que já foram bem estabelecidas na área da Biblioteconomia (MILSTEAD; FELDMAN, 1999).

Seja qual for o nome que se use, catalogação, indexação ou metadados, o conceito é familiar para profissionais de informação. Agora o mundo eletrônico finalmente o descobriu. Até poucos anos atrás, somente alguns filósofos tinham ouvido a palavra “metadado”. Hoje, é difícil encontrar uma publicação que a ignore (MILSTEAD; FELDMAN, 1999).

Sendo assim, é preciso deixar claro que os metadados em si não são algo novo, e que a novidade está na variedade de padrões disponíveis e em desenvolvimento e também as formas como vêm sendo utilizados (ORTIZ-REPISO JIMÉNEZ, 1999).

Apesar de ser uma nova denominação para uma prática já estabelecida na Biblioteconomia, é preciso estabelecer algumas de suas características que são peculiares a essa “nova” forma de representação.

Senso e Rosa Piñero (2003) consideram a catalogação como um processo de geração de metadados. De acordo com Mey (1995) a catalogação é a representação de um item documentário por meio de mensagens codificadas que permite caracterizar estes itens, individualizá-los tornando-os únicos entre os demais itens de um acervo e também reuni-los por suas semelhanças.

Para Mey (1995, p. 06) “A riqueza da catalogação repousa nos relacionamentos entre os itens, estabelecidos de forma a criar alternativas de escolha para os usuários”, pois além de representá-los e localiza-los, a catalogação permite aos usuários escolher entre itens semelhantes que estão veiculados em suportes e meios variados.

Sob esse ponto de vista, vemos que os metadados possuem a mesma finalidade da catalogação que, de acordo com Siqueira (2003, p. 37), pode ser definida como,

[...] uma forma de representação sucinta e padronizada de um item documentário, que tem como objetivo ampliar as formas de acesso a um documento facilitando tarefas e os processos de localização de documentos e informações.

Considerando as características e funções da catalogação citadas acima, podemos ver que nada se difere dos metadados, pois eles permitem:

- A representação de um recurso por meio de mensagens codificadas;
- Identificar, tornar único e reunir recursos por suas semelhanças;
- Possibilita a escolha entre recursos de um determinado acervo; e
- A localização desse recurso, a escolha entre suportes e recursos semelhantes.

De modo geral, podemos dizer que o objetivo e a função dos metadados são os mesmos da catalogação: representar as características e o conteúdo de um recurso informacional de forma padronizada, facilitando a identificação, busca, localização e recuperação desses recursos. A diferença dessa forma de representação está na nova abordagem dada pelo ambiente tecnológico em que ela se insere.

Os autores Senso e Rosa Piñero (2003) afirmam que se ampliarmos a definição de “dados sobre dados”, podemos ver que o conceito de metadados abarca mais informação que o termo catalogação. Um dos motivos apontados por esses autores seria a inclusão de informações sobre o contexto em que se insere o recurso e não somente informações

referenciais. Entretanto, essa visão é um pouco equivocada, pois a função dos metadados não se difere da função da catalogação, que também possibilita a inclusão de informações sobre o contexto em que o recurso se insere (como exemplo, a área 7: Notas no código de catalogação AACR2, e o campo 700: Notas no Marc 21).

Na Ciência da Informação os metadados estão relacionados com a catalogação em meio eletrônico. Entretanto em outras áreas pode ser abordado de modo diferente, ou seja, outro contexto que não seja catalogação. Talvez por esse motivo os autores Senso e Rosa Piñero tenham feito tal afirmação pensando na utilização de metadados para designar outro tipo de tratamento da informação que não seja catalogação (como por exemplo, o gerenciamento eletrônico de documentos – GED – do setor administrativo de uma empresa).

Os metadados e formatos de metadados apresentam algumas características em comum que foram abordadas por diversos autores e que será explorada de forma resumida a seguir, com o intuito de deixar mais claras as características de um dos objetos de estudo desta pesquisa.

Características dos metadados e formatos de metadados.

Essa categorização, apontada por Gilliland-Swetland (1999) e adotada por Rosetto (2003) e Senso e Rosa Piñero (2003), destaca que os tipos de metadados estão relacionados com as características e funções que eles apresentam e podem pertencer a mais de uma das categorias descritas abaixo. Os metadados podem ser dos seguintes tipos:

- Administrativos: são metadados usados no gerenciamento e administração dos recursos de informação. Esse tipo de metadado fornece informações como: data de criação dos recursos, tipos de arquivos, formas de acesso, controle de direitos e reproduções,

informação sobre registros legais, informação sobre localização etc (GILLILAND-SWETLAND, 1999; ROSETTO, 2003; SENSO, ROSA PIÑERO, 2003);

- Descritivo: são metadados usados para descrever, identificar e representar recursos de informações. Esse tipo de metadado fornece informações como: informações relacionadas com a catalogação como título, autor, imprensa, data, resumo, palavras-chave, e ainda a relação dos hiperlinks entre os recursos, anotações de usuários etc (GILLILAND-SWETLAND, 1999; ROSETTO, 2003; SENSO, ROSA PIÑERO, 2003);
- Conservação: são metadados relacionados com a conservação e preservação dos recursos de informação. Esse tipo de metadado fornece informações como: informações sobre as condições físicas de um recurso, informações de como conservar e preservar as versões físicas e digitais de um recurso etc (GILLILAND-SWETLAND, 1999; ROSETTO, 2003; SENSO, ROSA PIÑERO, 2003);
- Técnico: são metadados relacionados com o funcionamento dos sistemas e o comportamento dos metadados. Esse tipo de metadado fornece informações como: informações sobre hardware e software, digitalização, controle do tempo de resposta dos sistemas, autenticidade e segurança dos dados etc (GILLILAND-SWETLAND, 1999; ROSETTO, 2003; SENSO, ROSA PIÑERO, 2003);
- Uso: são metadados relacionados com o nível e tipo de uso dos recursos de informação. Esse tipo de metadado fornece informações como: informações sobre os registros de exibição, controle de uso e usuários, controles de acesso, informação sobre versões múltiplas etc (GILLILAND-SWETLAND, 1999; ROSETTO, 2003; SENSO, ROSA PIÑERO, 2003).

Além de determinar os tipos de metadados Gilliland-Swetland (1999) destaca que os metadados apresentam também alguns atributos que os caracterizam. Esses atributos

destacados por Gilliland-Swetland também foram abordados por Senso e Rosa Piñero (2003), são eles:

- Fonte dos metadados: se os metadados são internos, gerados no momento da criação do recurso (exemplo: nomes de arquivos), ou são metadados externos, gerados posteriormente a criação do recurso (exemplo: fichas e registros de catalogação);
- Método para criação dos metadados: se os metadados são automáticos, gerados automaticamente por um computador (exemplo: índices de palavras-chave), ou são metadados manuais, criados por indivíduos (exemplo: descrição de um recurso pelo padrão Dublin Core);
- Caráter dos metadados: se os metadados foram criados por indivíduos que não são especialistas da área de informação (exemplo: metadados criados pelo desenvolvedor de uma página pessoal), ou são metadados criados por especialistas temáticos ou da área de informação (exemplo: registros em formato Marc elaborados por um bibliotecário);
- Status: se os metadados são estáticos, que não mudam depois de criados (exemplo: título e data de criação de um recurso), ou são metadados dinâmicos, que podem se modificar de acordo com o uso e manipulação do recurso (exemplo: registros de operações dos usuários); se os metadados são de longa duração, para assegurar a acessibilidade e usabilidade do recurso (exemplo: formatos técnicos e processamento da informação), ou são metadados de curta duração, que são principalmente do tipo operacional (exemplo: informam sobre conservação e administração dos recursos);
- Estrutura: se os metadados são estruturados, ou seja, apresentam uma estrutura previsível, pré-determinada baseada em um padrão normalizado internacionalmente (exemplo: MARC), ou são metadados não estruturados, ou seja, não possuem estrutura previsível (exemplo: formatos de banco de dados locais);

- Semântica: se os metadados são controlados, ou seja, aqueles que seguem ou são normalizados por um vocabulário controlado, formulário de autoridade etc (exemplo: AACR2), ou são metadados não controlados, ou seja, aqueles que não seguem um vocabulário controlado ou formulário de autoridade (exemplo: meta etiquetas HTML);
- Nível: se os metadados são de coleções, ou seja, estão relacionados a coleções de documentos ou recursos (exemplo: uma coleção de recursos descritos pelo formato MARC), ou são metadados individuais, ou seja, metadados relacionados com recursos individuais ou que não pertencem a nenhuma coleção (exemplo: legenda de uma imagem).

É importante estabelecer os tipos de metadados, suas características, funções e atributos para que possa ser demonstrada toda a sua potencialidade para a representação dos dados. Somente com o uso de metadados cuidadosamente elaborados será possível aproveitar as oportunidades e vantagens que esta “nova” forma de representação traz para o tratamento da informação em meio digital. Para isso é necessário antes de tudo conhecê-los (GILLILAND-SWETLAND, 1999).

Como exemplo de metadados Senso e Rosa Piñero (2003) destacam os metadados relacionados a identificação dos recursos, a descrição de seu conteúdo, a localização, acessibilidade, bem como metadados relacionados a gestão de direitos autorais, de reprodução, acesso etc., que são alvo de seu estudo e apontam os seguintes exemplos:

- O cabeçalho de um arquivo multimídia (imagem, vídeo ou áudio).
- O resumo de um documento.
- O catálogo de uma base de dados.
- Os termos controlados fazendo uso de um tesouro.
- As palavras extraídas de um texto.
- As fichas catalográficas em qualquer formato (ISBD, MARC...).
- As páginas amarelas.
- Etc. (SENSO; ROSAPIÑERO, 2003, p. 99, tradução nossa).

Senso e Rosa Piñero (2003) afirmam também que podemos encontrar os metadados na Internet sob várias formas tais como: índices de documentos contidos em uma Intranet,

cabeçalhos de mensagens de correio eletrônico; descrição dos arquivos acessíveis via FTP, termos extraídos pelos motores de busca e indexação, entre outros.

Entretanto, como pode ser visto nos exemplos acima, os metadados não são somente aqueles que seguem um padrão ou formato complexo, mas também metadados mais simples, ou seja, representações de dados simples, porém padronizadas, que pode ser extraída automaticamente por motores de ferramentas de busca da Web.

Os metadados são desenvolvidos em uma variedade de níveis, pois dependem da necessidade da comunidade e das características dos recursos a serem descritos. Sendo assim, Rosetto (2003) e Senso e Rosa Piñero (2003) destacam um esquema de classificação proposto por Dempsey e Harry (1997), que apontam três tipologias (ou níveis) de formatos de metadados. Essa categorização foi apresentada pelo quadro de Rosetto (2003, p. 54), conforme visto na Figura 23 a seguir:

	Banda um	Banda dois	Banda três
Característica do registro	- Formatos simples - Padrão proprietário - Todo texto indexado	- Formatos estruturados - Padrões emergentes - Estrutura em campos	- Formatos altamente estruturados - Padrões internacionais - Estrutura por meio de etiquetas (tags)
Formatos dos registros	- Lycos, Altavista, Yahoo, etc.	- Dublin Core, Planilha IAFA, RCF 1807, SOIF, LDIF	- MARC, TEI, CIMI, EAD, ICPSR

FIGURA 23: Tipologia de formatos de metadados.
FONTE: Rosetto (2003, p. 54).

Com base na Figura 23, Rosetto (2003, p. 54) explica o seguinte,

- 1) Na banda um, encontram-se os formatos com dados não-estruturados, tipicamente extraídos em base automática dos recursos e indexados por motores de busca existentes na Internet;
- 2) Na banda dois, entram os formatos com dados básicos estruturados, contendo descrições suficientes que permitem ao usuário verificar a potencialidade de sua utilidade ou o interesse por um recurso sem ter que recupera-lo ou conecta-lo;
- 3) Na banda três, encontram-se os formatos cujos registros são descritos mais formalmente, que podem ser usados tanto para a localização e

recuperação como para documentar os objetos, ou muito freqüentemente as coleções de objetos. (ROSETTO, 2003, p. 54).

Baseado na citação anterior de Rosetto (2003) e nas explicações de Senso e Rosa Piñero (2003) que também utilizam essa categorização podemos considerar o seguinte:

- a. Formatos Simples: são formatos de metadados com dados não-estruturados, cuja recuperação é feita de modo automático, gerados por robôs, apresenta na maioria das vezes uma semântica reduzida (SENSO; ROSA PIÑERO, 2003). Como exemplo podemos citar as Meta Tag(s) e os metadados utilizados na transferência de dados utilizando o protocolo http – hipertext transfer protocol (BARRETO, 1999).
- b. Formatos Estruturados: são formatos de metadados mais estruturados baseados em normas emergentes e que proporcionam uma descrição mais clara do recurso por proporcionar o armazenamento da informação em campos, facilitando assim a recuperação do recurso. Nessa categoria começa a ser inserido a ajuda de especialistas em informação. Como exemplo dessa categoria podemos citar o padrão Dublin Core (SENSO; ROSA PIÑERO, 2003).
- c. Formatos Ricos: são formatos de metadados mais complexos, com alto grau de descrição, baseados em normas especializadas e códigos específicos. Seu alto nível de especificidade possibilita a descrição ideal de recursos, sendo eles individuais ou pertencentes a coleções em um repositório, facilitando assim sua localização. Como exemplo dessa categoria podemos citar o formato MARC (SENSO; ROSA PIÑERO, 2003).

A partir da análise desses autores foi possível verificar então que os metadados são representações que seguem estruturas padronizadas de descrição que variam das mais simples às mais complexas, tendo uma estrutura intermediária entre estes. As mais simples são como as palavras-chave retiradas automaticamente de um recurso pelo indexador de uma ferramenta. As intermediárias apresentam um pouco mais de detalhes devido aos seus

elementos de representação, como o padrão Dublin Core, com seus quinze elementos de descrição. E por fim, os metadados com uma maior complexidade, pois exigem o uso de normas e códigos internacionais a serem seguidos como é o caso do formato de intercâmbio de dados bibliográficos MARC 21.

5.2 FORMATOS DE METADADOS

Diante da heterogeneidade de recursos informacionais disponibilizados na rede, foram sendo criados, com o intuito de melhor representá-los, vários tipos de formatos ou padrões de metadados. A adoção desses tipos de padronização tem a intenção de “[...] promover a uniformidade da descrição dos conteúdos dos documentos e da forma de sua apresentação para facilitar o acesso fácil e universal às informações [...]” (ROSETTO, 2003, p. 40) e, além disto, proporcionam também,

[...] o fornecimento de um modelo normativo que oriente a estruturação e transferência de dados, diminuindo custos e esforços e melhorando os procedimentos de catalogação descritiva. O modelo dá condições, ainda, à padronização, convenções de descrição intelectual para assegurar a consistência e a caracterização precisa dos materiais (ROSETTO, 2003, p. 40-41).

Para haver uma boa representação, os metadados precisam “obedecer” a estruturas padronizadas de descrição, ou mais precisamente, a formas de representação que são os padrões de metadados, contudo, como já comprovado anteriormente, são coisas distintas. Após analisar vários autores, Rosetto (2003, p. 50, grifo do autor) aponta as características dos formatos de metadados e conclui que,

[...] um formato de metadados é um conjunto de elementos estruturados por meio de uma semântica e uma sintaxe próprias para a codificação dos dados. Os componentes do recurso em referência devem estar formatados,

conforme regras pré-estabelecidas e devem ser autorizados por uma instituição responsável pelo seu gerenciamento.

Explicando a afirmação acima, podemos dizer então que, o formato de metadados é um conjunto de elementos estruturados ou projetados para um propósito. Para isso o formato possui segundo Rosetto (2003, p. 49),

- a) uma “semântica da estrutura”, que estabelece um conjunto de campos pré-determinados, com regras bem definidas para a descrição dos componentes;
- b) uma sintaxe, que tem regras estabelecidas para a codificação e transferência dos dados.

Além disso, os componentes que descrevem o recurso devem estar formatados conforme regras pré-definidas em normas e regras de catalogação (Exemplo: nome de autores descritos e normalizados pelas regras do AACR2) (ROSETTO, 2003).

Os formatos são utilizados para padronização dos metadados a serem descritos, para entendermos melhor serão tratadas algumas características dos formatos de metadados. Nesta pesquisa serão consideradas as duas fases ou duas eras: *a pré-Internet e Internet*, apontadas por Rosetto (2003).

Era Pré-Internet.

Nesse período, para a Biblioteconomia o Formato MARC se destaca e segundo Rosetto (2003, p. 42), é “[...] considerado um formato de metadados altamente estruturado e semanticamente complexo, com ênfase na descrição completa dos objetos bibliográficos, produzidas por bibliotecários”. Hoje atualizado para o formato MARC 21, contém o campo 856, responsável pela inclusão da URL na descrição dos recursos eletrônicos.

Era Internet.

Os formatos de metadados dessa categoria foram criados exclusivamente para a representação de recursos em meio eletrônico. Podemos citar como exemplo o padrão de

metadados Dublin Core e mais recentemente o formato MARCXML, que une as vantagens do formato MARC com a flexibilidade da linguagem de marcação XML.

Rosetto (2003) aponta que um formato de metadado deve ser normalizado por convenções que especificam a semântica, a estrutura e a sintaxe necessárias para representar e intercambiar as representações dos recursos. Para ocorrer a representação dos recursos informacionais, os responsáveis pela aplicação dos metadados devem estabelecer e declarar as características técnicas e de conteúdo que compõe o padrão ou formato utilizado, tais como: a estrutura dos metadados com seus elementos de descrição, quais desses elementos serão obrigatórios ou repetitivos, se terá a adoção de normas complementares para a representação do conteúdo dos recursos, se o formato de metadado promove a interoperabilidade, entre outras características (ROSETTO, 2003).

Tendo como base de categorização as propostas de Rosetto (2003) e Senso Rosa Piñero (2003), trataremos e estabeleceremos algumas das características dos padrões pertencentes a estas três categorias: formatos de metadados simples, estruturados e ricos (complexos ou altamente estruturados).

5.2.1 FORMATOS DE METADADOS SIMPLES

De acordo com as afirmações de Berners-Lee (1997) e Senso e Rosa Piñero (2003), os metadados também são descrições de dados mais simples como, por exemplo, cabeçalhos de arquivos multimídia, resumos de documentos e até mesmo as palavras-chave extraídas de um texto. Sendo assim, encontramos no ambiente Web vários tipos de formatos de metadados simples.

De acordo com os estudos de Barreto (1999), desde o surgimento da Web existe troca de metadados, sendo que a forma mais comum de uso de metadados ocorre na transferência do recurso de informação do cliente para o servidor pelo protocolo http.

Segundo Almeida (2002) as páginas da Web são consideradas dados semi-estruturados, pois apresentam algum tipo de estrutura. São, portanto, de caráter intermediário, por estar entre os dados estruturados, que são aqueles encontrados nos bancos de dados relacionais e os dados não estruturados, encontrados em textos livres.

Os dados semi-estruturados se caracterizam por serem descritos por uma sintaxe simples, mas não com ausência de padronização. A vantagem de trabalhar com dados semi-estruturados está na possibilidade de uma flexibilização, fazendo com que exista, mesmo que de forma simples, uma representação do conteúdo dos recursos.

Segundo Almeida (2002, p. 08) “Os dados semi-estruturados possuem a habilidade de aceitar variações na estrutura, de forma que possam se adequar melhor a situações reais”. Essa flexibilização permitirá também a inclusão de metadados mais complexos, possibilitando assim uma representação mais rica.

Dois pontos se destacam sobre a troca de formatos de metadados simples a URI – Uniform Resource Identifier, para a localização do recurso na Web e as Meta Tag(s) – ou etiquetas – para a descrição do conteúdo do recurso. A seguir veremos cada um deles.

A WWW utiliza tecnologias relativamente simples para que ocorra a troca de informações. Os recursos são localizados na rede por meio de identificadores globais denominados URI e essa identificação compõem uma base arquitetônica na rede de três itens de acordo com W3C (2004):

- 1) Identificação: são usadas as URI(s) para identificar o recurso. A URL – Uniform Resource Locator é um exemplo de URI (W3C, 2004).

2) Interação: Os agentes da Web se comunicam usando protocolos padronizados que habilitam uma comunicação entre usuário e recurso, por meio da troca de mensagens, entre cliente e servidor, que segue uma sintaxe e uma semântica definidas. Quando os usuários indicam o endereço eletrônico, o navegador envia o protocolo http aos servidores, que mandam por sua vez, o recurso (ou mensagem de erro) para o usuário (W3C, 2004). Nesse processo há uma troca de metadados simples, que são anexados ao cabeçalho do protocolo http, que podem ser dos seguintes tipos, segundo Barreto (1999, p. 40),

[...] *expires*, que informa a data em que um documento se torna obsoleto; *refresh*, que especifica um período (“delay”) em segundos antes que o navegador recarregue o documento automaticamente; e *content language*, que especifica o idioma relativo ao conteúdo do documento, etc.

3) Formatos: para a transferência de recursos entre agentes, a maioria dos protocolos usados para representar a localização do recurso fazem uso de uma sucessão de “mensagens” que contém uma representação de dados e metadados. W3C (2004) aponta que o http, por exemplo, transmite metadados simples em sua estrutura e “[...] usa ‘Content-Type’ e ‘Content-Encoding’ do cabeçalho de arquivos para futuras identificações de formatos de representações”. Esses campos presentes no cabeçalho do http indicam os dados de representação que podem ser processados e transferidos em determinadas aplicações (HTML, XML ou XHTML).

Existe uma relação entre identificador, recurso e representação do recurso, esta relação pode ser ilustrada da seguinte forma, conforme a Figura 23 apontada por W3C (2004):

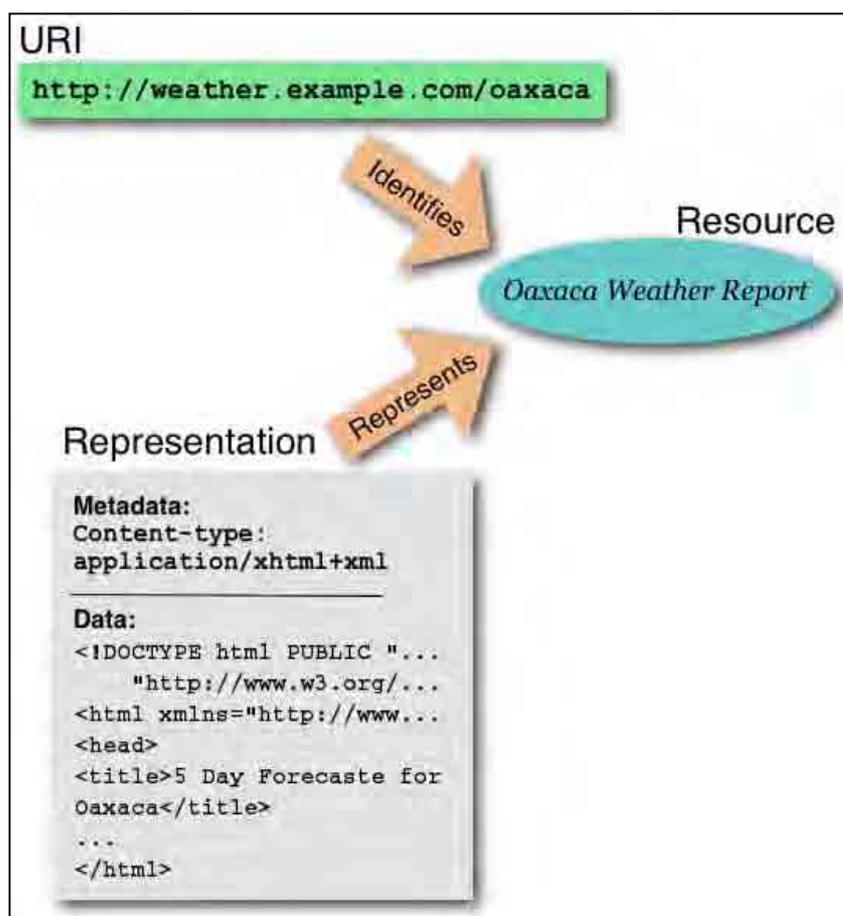


FIGURA 24: Relação entre identificador, recurso e representação.
FONTE: W3C (2004).

O exemplo da Figura 24 mostra a relação entre identificador, recurso e representação. É por meio da URI, que nesse exemplo é do tipo URL, que identificamos e localizamos um determinado recurso de informação com sua correspondente representação (metadados). Isso ocorre, pois os agentes de software se comunicam utilizando protocolos unificados que aderem a uma sintaxe e semântica bem definida. Esses protocolos utilizam uma sucessão de uma ou mais mensagens e proporcionam a transferência de representações entre agentes de softwares. O protocolo HTTP, apresentado no exemplo, transmite além da localização, informações que identificam o formato de representação de um recurso informacional. E é por meio dessa relação entre identificador, recurso e representação que os agentes de software encontram informações sobre os metadados.

Contudo, é preciso destacar que nessa relação existem diferentes níveis de metadados. Primeiramente, temos a URL para a identificação do recurso que segundo Barreto (1999, p. 41) é “[...] um metadado que expressa a sua localização em um determinado espaço de informação da Internet”. Depois temos a representação do recurso por meio das TAG(s) descritivas da linguagem de marcação. Tanto a URL como as TAG(s) descritivas podem ser consideradas como formatos de metadados simples, por se tratar de dados semi-estruturados. No entanto, é preciso lembrar que, por mínima e simples que seja essa representação possui seu valor, pois a URI é a base para a localização de recursos e as Tag(s) a base para a inclusão posterior de uma representação mais detalhada.

Para falarmos de Tag(s) temos que lembrar da linguagem de marcação HTML, que é a linguagem mais utilizada atualmente para disponibilizar informações na rede. Uma página HTML é dividida em duas partes: a) cabeçalho, delimitado pelas Tag(s) inicial e final: <HEAD> ... </HEAD>, e; b) corpo, delimitado pelas Tag(s) inicial e final: <BODY ... </BODY>. (SANTARÉM SEGUNDO, 2004).

É possível encontrar na linguagem HTML alguma forma de representação do conteúdo do recurso por meio das Tag(s) descritivas ou Meta Tag(s), que são as únicas que comportam algum tipo de representação do conteúdo do recurso e que possibilita, posteriormente a inclusão de algum tipo de metadado. Para Barreto (1999, p. 40), as Meta Tag(s) funcionam como um recipiente de metadados, pois armazenam “[...] as propriedades do documento na forma de pares (nome-atributo, valor-atributo)”.

As Meta Tag(s), que se localizam dentro do cabeçalho HEAD da página HTML, existem em dois tipos: a Meta Tag com o atributo NAME e a HTTP-EQUIV. Ambos os tipos de Meta Tag(s) podem utilizar os seguintes tipos de atributos para descrição dos dados: generator, keywords, content-type, description, author, formather, refresh, robots,

distribution, resource-type, template, copyright, classification etc (O'NEILL; LAVOIE; McCLAIN, 1998?).

O'Neill, Lavoie e McClain (1998?), que relatam um projeto desenvolvido para verificar o uso de metadados na Web destacam que, o uso de Meta Tag(s) para a descrição do conteúdo dos recursos com os atributos acima são uma realidade no momento, portanto constituem-se no emprego mais comum de formatos de metadados para a descrição das páginas em geral.

Entretanto, é possível encontrar na rede páginas descritas por outros tipos de formatos de metadados, que apresentam uma estrutura um pouco mais detalhada que a descrição por Meta Tag(s), esses formatos são denominados formatos de metadados estruturados.

5.2.2 FORMATOS DE METADADOS ESTRUTURADOS

Os formatos de metadados estruturados podem ser caracterizados como formatos que possuem elementos de descrição simples e genéricos para atingir a descrição de recursos variados disponibilizados em domínios distintos.

Esses formatos se apresentam como intermediários entre os formatos de metadados simples e os formatos de metadados ricos (complexos ou altamente estruturados). Foram criados com o intuito de promover a localização e descoberta de recursos em meio eletrônico.

Barreto (1999), apesar de classificar esses formatos em outra categoria “Padrões de metadados para a descoberta de recursos na rede”, aponta em seus estudos que esses formatos apresentam um modelo de dados simples, na forma de pares (nome-atributo, valor-atributo), e

não expressam relacionamentos e hierarquias complexas entre recursos, mas são altamente interoperáveis.

Um dos formatos de metadados que pertence a essa categoria é o formato de metadados Dublin Core (DC). Segundo os estudos de Grácio (2002) o formato de metadados DC vem sendo utilizado em diversas comunidades, pois a necessidade de representação da informação em meio eletrônico ocorre em diversas áreas do conhecimento.

Criado originalmente para promover a descoberta, identificação e localização de recursos na Web, o formato DC vem ganhando maior destaque na Web devido a sua simplicidade e flexibilidade.

Criado em 1995 por profissionais de várias áreas do conhecimento, o DC se caracteriza por ser um formato baseado na catalogação descritiva (BARRETO, 1999; GRÁCIO, 2002). Para atender as necessidades de descrição dos novos tipos de materiais seus desenvolvedores buscaram durante sua criação atender às seguintes características, conforme aponta Grácio (2002),

- Que o formato pudesse ter uma infra-estrutura padronizada internacionalmente, por isto a participação de profissionais de diversas áreas do conhecimento;
- Que o formato apresentasse uma simplicidade na representação semântica dos elementos, garantindo assim uma maior interoperabilidade nos diversos ambientes informacionais e proporcionando também um maior uso entre os desenvolvedores de sites, pois não exige conhecimento prévio ou experiência na área de descrição para usá-lo;
- Que o formato pudesse ser flexível, ou seja, que proporcionasse a adição ou exclusão de elementos descritivos necessários para a representação de determinados tipos de recursos presentes em comunidades de interesse distintas.

Com essas características, o padrão DC apresenta-se atualmente em três níveis de representação: o qualificado, com quinze elementos de descrição; o não qualificado que apresenta apenas alguns elementos básicos para a descrição; e uma versão mais completa, que apresenta um elemento adicional (*Audience*) e um grupo de elementos de refinamento ou qualificadores, que proporcionarão uma representação mais detalhada. Esses elementos podem ser divididos nas seguintes categorias presentes na Figura 25 (HILLMANN, 2003):

Content	Intellectual Property	Instantiation
<u>Coverage</u>	<u>Contributor</u>	<u>Date</u>
<u>Description</u>	<u>Creator</u>	<u>Format</u>
<u>Type</u>	<u>Publisher</u>	<u>Identifier</u>
<u>Relation</u>	<u>Rights</u>	<u>Language</u>
<u>Source</u>		
<u>Subject</u>		
<u>Title</u>		
<u>Audience</u> ⁷		

FIGURA 25: Elementos do padrão de metadados Dublin Core e sua categorização.

FONTE: Hillmann (2003). Disponível em:

<<http://dublincore.org/documents/usageguide/elements.shtml>>.

De modo resumido, cada elemento presente na Figura 25 será apresentado conforme descreve HILLMANN (2003):

- *Tag Title (Título)*: nome atribuído ao recurso e é pelo título que geralmente o recurso é conhecido;
- *Tag Creator (Criador)*: é o responsável pela criação do conteúdo do recurso, que pode ser uma pessoa, organização ou serviço;

⁷ Elemento de refinamento, não está contemplado na lista dos quinze elementos de descrição do Dublin Core Metadata Element Set, considerados básicos para a descrição dos recursos.

- *Tag Subject and Keywords (Assunto e palavra-chave)*: um tópico ou tema que representa resumidamente o conteúdo do recurso, geralmente o recurso será representado por palavras-chave, frase ou mesmo códigos de classificação e há neste elemento a recomendação do uso de vocabulário controlado ou esquemas de classificação pelo DCMI (Dublin Core Metadata Initiative);
- *Tag Description (Descrição)*: resumo do conteúdo do recurso que sintetize fidedignamente a informação contida no recurso;
- *Tag Publisher (Publicador)*: entidade responsável pela publicação do recurso, podendo ser pessoas, organizações ou serviços;
- *Tag Contributor (Contribuidor)*: entidade responsável por contribuições feitas na elaboração do conteúdo do recurso, que também podem ser representados por pessoas, organizações ou serviços;
- *Tag Date (Data)*: indica a data de algo ocorrido durante o ciclo de vida do recurso, geralmente está associada à sua criação. É recomendado o uso do seguinte formato de apresentação recomendado pela ISO 8601 (W3CDTF): AAAAMMDD (ano/ mês/dia);
- *Tag Resource Type (Tipo do recurso)*: trata da natureza ou gênero do tipo do recurso, são termos que estabelecem categorias gerais, funções, gêneros ou níveis de agregação para conteúdo (recomenda-se uso de vocabulário controlado);
- *Tag Format (Formato)*: indicação da estrutura física ou digital do recurso, por exemplo, se o recurso é em PDF, RDF, entre outros (recomenda-se o uso de vocabulário controlado);
- *Tag Resource Identifier (Identificador)*: é um número ou uma referência única ao recurso para a sua individualização e localização, como por exemplo, *Uniform Resource Identifier (URI)*, *Uniform Resource Locator (URL)*, *Digital Object Identifier (DOI)* ou *International Standard Book Number (ISBN)*;

- *Tag Source (Fonte)*: referência que indica a fonte usada para o desenvolvimento do recurso que está sendo descrito e preferencialmente deve ser indicada pelo *string* ou número do identificador;
- *Tag Language (Língua)*: o idioma no qual o recurso foi escrito. Para a padronização desse elemento é recomendado o uso da RFC 3066 [RFC3066] que está baseada na ISO 639;
- *Tag Relation (Relação)*: indica, pelo uso de identificador, a relação existente entre recursos;
- *Tag Coverage (Cobertura)*: descreve a extensão do conteúdo do recurso, incluindo a localização espacial, período ou jurisdição (também é recomendado o uso de vocabulário controlado como, por exemplo, o *Thesaurus of Geographic Names* [TGN]);
- *Tag Rights Management (Direitos autorais)*: descreve informações a respeito dos direitos autorais do recurso, propriedade intelectual e informações sobre gerenciamento do recurso;

O elemento adicional *Audience*, inserido no quadro da Figura 25, não está contemplado no *template*⁸, esse elemento está relacionado com o público alvo do recurso descrito, sua importância está na determinação das comunidades de interesse na qual o recurso informacional se destina. Além desses elementos de descrição, o padrão DC ainda apresenta elementos de refinamento que também não estão inseridos no *template*, mas que é possível encontrá-los em DCMI (Dublin Core Metadata Initiative) Metadata Terms, disponível em: <<http://dublincore.org/documents/dcmi-terms/>>, com seus respectivos significados e incluí-los posteriormente na descrição.

⁸ Template está relacionado a modelo, no caso do Template do padrão Dublin Core é uma estrutura pré-definida que contém os elementos de descrição desse padrão na forma de campos que ao serem preenchidos geram uma representação com Tags características desse padrão.

O *template* do padrão Dublin Core disponibiliza duas versões para a descrição: a página principal do *template* com o DC qualificado, onde são utilizados todos os quinze elementos de descrição; e um *link* para o DC não qualificado, que oferece um número mínimo de elementos necessários para a descrição do recurso. Na Figura 26 estão os elementos pertencentes a cada nível (DUBLIN, 1997a; DUBLIN, 1997b):

DC QUALIFICADO	DC NÃO QUALIFICADO
1 TITLE	1 TITLE
2 CREATOR	2 CREATOR
3 SUBJECT: Keywords	3 SUBJECT: Keywords
3 SUBJECT: Controlled vocabulary	8 TYPE
3 SUBJECT: Classification	10 IDENTIFIER: URL
4 DESCRIPTION	10 IDENTIFIER (string ou ISBN)
5 PUBLISHER	12 LANGUAGE
6 CONTRIBUTOR	
7 DATE	
8 TYPE	
9 FORMAT	
10 IDENTIFIER: URL	
10 IDENTIFIER	
11 SOURCE	
12 LANGUAGE	
13 RELATION	
14 COVERAGE	
15 RIGHTS	

FIGURA 26: Comparação entre elementos: Dublin Core Qualificado e Dublin Core não Qualificado.
 FONTE: Dublin Core Metadata Template. Disponível em: <<http://www.lub.lu.se/cgi-bin/nmdc.pl>>.

Nas Figuras 27 e 28, podemos ver o mesmo recurso descrito pelas duas versões do Dublin Core: DC Qualificado, Figura 27 e DC não Qualificado, Figura 28:

```

<META NAME="DC.Title" CONTENT=" Necessidades e expectativas dos usuários na educação a
distância">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#title">
<META NAME="DC.Title.Alternative" CONTENT="estudo preliminar junto ao programa de pos
graduaca em engenharia de producao da universidade de Santa Catarina">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#title">
<META NAME="DC.Creator" CONTENT="Eliane Maria Stuart Garcez">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.Creator.Address" CONTENT="garcez@newsite.com.br">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.Creator.2" CONTENT="Greg&oacute;rio J. Varvakis Rados">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.Creator.Address.2" CONTENT="grego@eps.ufsc.br">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.Subject" CONTENT="Necessidade do usu&aacute;rio">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" CONTENT="Educa&ccedil;&atilde;o a Dist&acirc;ncia">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" CONTENT="Expectativa do usu&aacute;rio">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" SCHEME="ERIC" CONTENT="Educacao a distancia">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" SCHEME="ERIC" CONTENT="Necessidade do usuario">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" SCHEME="ERIC" CONTENT="Necessidade educacional">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" SCHEME="ERIC" CONTENT="Educacao mediada por computador">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" SCHEME="DDC" CONTENT="374">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Description" CONTENT="Objetiva identificar necessidades e expectativas
informacionais de usuarios de bibliotecas academicas realizada junto aos mestrandos e
professores de curso a distancia do Programa de Pos Graduacao em Engenharia de Producao da
Universidade Federal de Santa Catarina. Sugere-se que as bibliotecas academicas atuem
centradas nessas necessidades identificadas, disponibilizando produtos tanto em formato
tradicional quando digital.">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#description">
<META NAME="DC.Publisher" CONTENT="ibict">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#publisher">
<META NAME="DC.Date" SCHEME="ISO8601" CONTENT="2002-01">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#date">
<META NAME="DC.Type" CONTENT="Text.Article">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#type">
<META NAME="DC.Format" SCHEME="IMT" CONTENT="application/pdf">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#format">
<LINK REL=SCHEMA.imt HREF="http://sunsite.auc.dk/RFC/rfc/rfc2046.html">
<METANAME="DC.Identifier" CONTENT="http://www.ibict.br/cionline/310102/31102030">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#identifier">
<META NAME="DC.Identifier" SCHEME="ISSN" CONTENT="0100-1965">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#identifier">
<META NAME="DC.Language" SCHEME="ISO639-1" CONTENT="pt">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#language">
<META NAME="DC.Coverage" CONTENT="Permanente">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#coverage">
<META NAME="DC.Rights" CONTENT="Revista Ciencia da Informacao">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#rights">
<METANAME="DC.Date.X-MetadataLastModified"SCHEME="ISO8601" CONTENT="2005-08-08">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#date">

```

FIGURA 27: DC Qualificado.

FONTE: Representação descrita no Template do Dublin Core.

Disponível em: <<http://www.lub.lu.se/cgi-bin/nmdc.pl>>.

```

<META NAME="DC.Title" CONTENT="Necessidades e expectativas dos usuários na educação a
distância: estudo preliminar junto ao programa de pos graduação em engenharia de produção da
universidade de Santa Catarina">
<META NAME="DC.Creator" CONTENT="Eliane Maria Stuart Garcez">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.Creator.Address" CONTENT="garcez@newsite.com.br">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.Creator.2" CONTENT="Greg&oacute;rio J. Varvakis Rados">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.Creator.Address.2" CONTENT="grego@eps.ufsc.br">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.Subject" CONTENT="Necessidade do usu&aacute;rio">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" CONTENT="Educa&ccedil;&atilde;o a Dist&acirc;ncia">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" CONTENT="Expectativa do usu&aacute;rio">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Type" CONTENT="Text.Article">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#type">
<METANAME="DC.Identifier" CONTENT="http://www.ibict.br/cionline/310102/31102030">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#identifier">
<META NAME="DC.Identifier" SCHEME="ISSN" CONTENT="0100-1965">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#identifier">
<META NAME="DC.Language" SCHEME="ISO639-1" CONTENT="pt">
<LINKREL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#language">

```

FIGURA 28: DC não Qualificado.

FONTE: Representação descrita no Template do Dublin Core. Disponível em:
<http://www.lub.lu.se/cgi-bin/nmdc.pl?lang=en&save-info=on&simple=1>.

É interessante destacar que o DC qualificado apresenta os quinze elementos de descrição sendo que os elementos *subject* e *identifier* aparecem repetidos para designar tipos de informações diferentes, conforme pode ser visto na Figura 26. Outro fator que deve ser destacado é que na página do DC não qualificado, é possível encontrar também os outros elementos que não foram incluídos na descrição, e é possível incluir qualquer um dos elementos de descrição que se encontram no final da página do *template*. Esses níveis de descrição que o padrão DC proporciona podem ser explicados da seguinte forma:

- 1º nível de representação: apresenta um número mínimo de elementos de descrição presentes no DC não qualificado (sete elementos).
- 2º nível de representação: apresenta os quinze elementos de descrição presentes no DC qualificado.

- 3º nível de representação: apresenta, além dos quinze elementos de descrição, o elemento adicional Audience e a possibilidade de inclusão de outros elementos de refinamento (qualificadores).

Considerando esses níveis de descrição, podemos verificar que a estrutura do padrão DC apresenta certa semelhança com o código de catalogação AACR2 por também estabelecer níveis diferentes de descrição: no primeiro nível que apresenta uma descrição básica com o mínimo de elementos descritivos (áreas); no segundo nível um pouco mais de detalhes; e no terceiro nível, é incluído o maior número de elementos de descrição do recurso.

Isso serve para mostra que os idealizadores do DC se preocuparam em disponibilizar níveis diferentes de descrição permitindo que se fizesse a opção por uma descrição mínima apenas para a localização do recurso. A preocupação dos desenvolvedores não era a de descrição como no AACR2, mas sim a de localização.

Uma característica do formato DC que merece ser destacada aqui é que apesar de não apresentar nenhum conjunto de regras para sua utilização, é possível encontrar, em quase todos seus elementos, a indicação do uso de vocabulário controlado. Para que possa ser estabelecida uma maior padronização na representação por esse formato, atualmente o DC encontra-se em análise pela NISO (*North American Standard Organization*) para que possa ser reconhecido como uma estrutura padrão e possa, assim como o formato MARC, estar vinculado diretamente a uma norma de intercâmbio, descrição e comunicação.

Por ser um formato de metadado estruturado, porém simples para ser utilizado, é possível sua utilização por qualquer pessoa que queira estabelecer uma representação mínima de um recurso. Entretanto, isso pode trazer certos problemas, pois não há uma garantia de que os documentos serão descritos de forma que os tornem únicos, identificáveis e recuperáveis, o que possibilitaria o aumento da dispersão de informações na Web, ao contrário de uma

representação elaborada por um catalogador que pode garantir a qualidade ao registro bibliográfico de maneira a maximizar sua disseminação, recuperação e uso.

Diante do que foi exposto é preciso dizer que a necessidade da construção das formas de representação para os recursos na Web é uma necessidade marcante no cenário atual e a descrição de modo inadequado dificultaria a recuperação dos recursos.

Entretanto, pelo fato do formato DC não exigir experiência na área de representação, e pelo fato de estar sendo usado em diversas áreas do conhecimento, o que ocorrerá é que teremos diferentes níveis de descrição com qualidades diferentes de representação, que dependerá da comunidade onde o formato estará sendo usado. A necessidade de uma representação de qualidade será fundamental para a recuperação dos recursos, exigindo assim, uma qualificação do profissional envolvido nesse trabalho e como o bibliotecário já apresenta experiência nesse campo, ele será um profissional de importância fundamental para a realização dessa tarefa.

5.2.3 FORMATOS DE METADADOS RICOS

Os formatos de metadados ricos, complexos ou altamente estruturados, proporcionam uma representação completa e detalhada de um recurso informacional. De forma geral, foram projetados para atender as necessidades de comunidades específicas e distintas, por este motivo, exige certa experiência entre os profissionais envolvidos na representação com esse tipo de formato.

Na área de Ciência da Informação, em especial na comunidade biblioteconômica, o formato que se destaca, por ser ainda hoje o que melhor representa descritivamente os itens

documentários é o formato MARC - Machine Readable Cataloging. Criado na década de 60 pela Library of Congress nos Estados Unidos, o MARC foi o primeiro formato de intercâmbio de dados criado para a catalogação automatizada.

Não vamos tratar aqui das diferenças entre versões do MARC, pois não é objetivo deste trabalho tratar de sua evolução, mas podemos dizer que esse formato passou por atualizações até chegar à versão atual, denominada MARC 21.

Sua estrutura está baseada na semântica estrutural do AACR2, pois foi desenvolvido nos princípios da catalogação tradicional e o seu uso é regido por regras de pontuação e de pontos de acesso, apresentadas no código AACR₂, e os campos de assunto são descritos de acordo com listas de cabeçalho de assunto ou tesouros.

Esse formato é destinado à criação, armazenamento, gerenciamento e intercâmbio de registros catalográficos e bibliográficos. O objetivo geral do MARC 21 é fornecer aos usuários a localização, o acesso e a recuperação dos recursos informacionais por meio de uma representação padronizada, e possibilitar a comunidade que adota esse formato intercambiar registros bibliográficos, ou seja, intercambiar os registros com a representação dos recursos informacionais. O registro de intercâmbio do formato MARC, que contém a representação de um item documentário, pode ser visualizado na Figura 28, disponível na página do Catálogo da Library Of Congress:

```

01488cam2200337a45000010008000000050017000080080041000250350021000
669060045000879550188001320100017003200200025003370400018003620420
009003800500026003890820016004151000032004312450086004632500012005
492600052005613000049006135000040006625200228007026500033009306500
033009636500024009966500021010206500023010417000021010649910065010
85-2502929-19930521155141.9-920219s1993 caua j 000 0 eng
- 9(DLC) 92005291- a7bcbccorignewdleocipf19gy-gencatlg- apcl6
to br00 02-19-92; br02 to SCD 02-21-92; fd11 02-24-92 (PS3537.A618
A...); fa00 02-26-92; fa05 03-02-92; fm31 03-06-92; CIP ver. pv08
04-16-93; pv01 to CLT 04-20-93; lb10 05-21-93- a 92005291 - -
a0152038655 :c$15.95- aDLCdDLCdDLC- alcac-00aPS3537.A618bA88
1993-00a811/.52220-1 aSandburg, Carl,d1878-1967.-10aArithmetic /-
cCarl Sandburg ; illustrated as an anamorphic adventure by Ted
Rand.- alst ed.- aSan Diego :bHarcourt Brace Jovanovich,cc1993.-
al v. (unpaged) :bill. (some col.) ;c26 cm.- aOne Mylar sheet
included in pocket.- aA poem about numbers and their
characteristics. Features anamorphic, or distorted, drawings which
can be restored to normal by viewing from a particular angle or by
viewing the image's reflection in the provided Mylar cone.- 0-
aArithmeticxJuvenile poetry.- 0aChildren's poetry, American.- 1-
aArithmeticxPoetry.- 1aAmerican poetry.- 1aVisual perception.-1 -
aRand, Ted,eill.- bc-GenCollhPS3537.A618iA88 1993p00013852947-
tCopy 1wBOOKS-_

```

FIGURA 29: Registro bibliográfico em formato de intercâmbio MARC 21.

FONTE: Library of Congress Online Catalog (2005a).

Disponível em: <<http://catalog.loc.gov/cgi-bin/Pwebrecon.cgi>>.

A Figura 29 ilustra o exemplo de um registro bibliográfico em formato de intercâmbio de importação e exportação de dados e apesar do arquivo de texto estar em várias linhas o registro de intercâmbio é composto de uma única linha com todos os caracteres que representam o recurso descrito (título, autor, assunto etc). De acordo com Ferreira (2002, p. iii),

O registro MARC é composto de 3 elementos: estrutura do registro, indicação de conteúdo e conteúdo dos elementos que compõem o registro. A **estrutura do registro**, é uma implementação da *American National Standard* para o Intercâmbio de Informação Bibliográfica (ANSI/NISO Z39.2) e sua ISO equivalente ISO 2709. A **indicação de conteúdo** – os códigos e convenções estabelecidos explicitamente para identificar e caracterizar os dados dentro do registro e permitir sua manipulação – são definidos para cada um dos formatos MARC. O **conteúdo dos elementos que compõem o registro** MARC é geralmente definido por padrões externos aos formatos, tais como *International Standard Bibliographic Description* (ISBD), *Anglo American Cataloguing Rules*, segunda edição (AACR2), *Library of Congress Subjects Headings* (LCSH) ou outras convenções e códigos usados pela organização criadora do registro. Os conteúdos de certos elementos de dados codificados, como por exemplo, o Líder, os campos 007 e 008, são definidos no formato MARC.

O formato MARC 21 abrange vários tipos de materiais em diferentes suportes, por isto, pode ser usado por qualquer organização que necessite representar recursos informacionais. Entretanto, seu maior uso se dá em bibliotecas que buscam no MARC 21 um meio de descrever e intercambiar registros bibliográficos de modo padronizado.

Também não é intenção deste capítulo tratar de toda a estrutura do MARC 21, pois isto já foi feito por muitos autores tais como: Ferreira (2000) que trata somente do MARC 21 e sua estrutura; Brene (2004) que traz em seu trabalho uma análise comparada entre AACR2, MARC 21, MARCXML e Dublin Core; Siqueira (2003) que trata das formas de representação da informação, em especial do uso do formato MARC 21 e MARCXML para a elaboração de registros bibliográficos.

O que nos interessa é destacar o MARC 21 como um formato de metadado complexo e que é aceito internacionalmente pela comunidade bibliotecária para a descrição, representação e intercâmbio de informações bibliográficas em formato legível por máquina.

A estrutura do formato MARC 21 para a representação de dados bibliográficos permite a integração entre bibliotecas e a transferência de registros através de redes, sendo que este intercâmbio de informações ocorre com o uso de protocolos, e não abrange somente a descrição de dados bibliográficos, mas também a representação sobre controle (gerenciamento) dos dados, dados relacionados à autoridade ou responsabilidade das obras, a classificação de documentos descritos e dados locais que podem ser incluídos nos catálogos para a comunidade informacional onde o material está localizado.

Estruturalmente o MARC apresenta um conjunto de elementos descritivos denominados campos, que pode ser categorizado de acordo com sua função, como segue:

- **Formato MARC 21 de dados bibliográficos:** contém especificações para a representação dos mais variados tipos de materiais bibliográficos (livros, publicação

seriada, mapas, etc.). Estabelece uma estrutura básica que garantirá a descrição e a recuperação dos diferentes suportes informacionais;

- **Formato MARC 21 para controle de dados (*Holding*):** contém especificações para a codificação dos elementos referentes ao controle e a localização de todas as formas de materiais;
- **Formato MARC 21 para dados de autoridade:** possui especificações para a codificação de elementos que identificam autoridade (responsabilidade pela obra) em um registro bibliográfico que necessita de um controle de autoridade;
- **Formato MARC 21 para dados de classificação:** contém especificações para a codificação e controle dos elementos de dados relacionados a classificação do conteúdo de um recurso informacional, possibilitando o desenvolvimento de esquemas de classificação;
- **Formato MARC 21 para comunidade informacional:** contém especificações para a codificação de registros que contenham informações sobre eventos, programas, serviços etc. Permite a inclusão e disseminação de informações relacionadas a isso e que podem ser integrados ao catálogo, tornando a informação acessível ao público (LIBRARY OF CONGRESS, 1996).

Cada categoria descrita acima exerce uma função e a representação de um recurso informacional é feita por meio de campos contidos em cada categoria. Entretanto, será enfocada nesta pesquisa somente a primeira categoria, que trata do formato MARC para dados bibliográficos. Na Figura 30 temos o exemplo de um livro descrito pelo MARC 21, no qual podemos visualizar os campos e subcampos utilizados para a representação:

```

01142cam 2200301 a 4500
000
001 92005291
003 DLC
005 19930521155141.9
008 920219s1993 caua j 000 0 eng
010 |a 92005291
020 |a0152038655 : |c$15.95
040 |aDLC|cDLC|dDLC
042 |alcac
050 00|aPS3537.A618|bA88 1993
082 00|a811/.52|220
100 1 |aSandburg, Carl, |d1878-1967.
245 10|aArithmetic / |cCarl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.
250 |a1st ed.
260 |aSan Diego : |bHarcourt Brace Jovanovich, |cc1993.
300 |a1 v. (unpaged) : |bill. (some col.) ; |c26 cm.
500 |aOne Mylar sheet included in pocket.
520 |aA poem about numbers and their characteristics. Features anamorphic, or distorted,
drawings which can be restored to normal by viewing from a particular angle or by viewing the
image's reflection in the provided Mylar cone.
650 0|aArithmetic|xJuvenile poetry.
650 0|aChildren's poetry, American.
650 1|aArithmetic|xPoetry.
650 1|aAmerican poetry.
650 1|aVisual perception.
700 1 |aRand, Ted, |eill.

```

FIGURA 30: Representação com MARC 21.

FONTE: Library of Congress Online Catalog (2005b).

Disponível em: <<http://www.loc.gov/standards/marcxml/Sandburg/sandburg.html>>.

No exemplo da Figura 30 pode-se verificar que cada informação é precedida por um número formado por três caracteres, são as denominadas etiquetas do MARC. Cada etiqueta apresenta uma estrutura: indicadores, identificadores de subcampo e o subcampo propriamente dito. Cada uma das etiquetas representam uma informação ou metadado. As etiquetas que compõem o conjunto de elementos descritores do formato MARC pode ser dividida em três partes: Líder, Diretório e Campos de dados variáveis. Segundo Ferreira (2002, p. iii-iv),

Líder – Dados que fornecem informações para o processamento do registro. Estes dados contêm números ou códigos e são identificados pela sua posição relativa. O Líder possui o tamanho de 24 caracteres e é o primeiro campo de um registro MARC.

Diretório – Uma série de entradas que contém a posição inicial e o tamanho de cada etiqueta (TAG) dentro do registro bibliográfico. Cada notação possui a extensão de 12 caracteres. No Diretório, as notações para campos de controle variável aparecem primeiro, seguidas pela etiquetas em ordem numérica crescente. Em seguida entram os campos de dados variáveis, arrumados em ordem crescente, de acordo com o primeiro caractere da etiqueta. A seqüência de armazenamento dos campos de dados variáveis, não

corresponde necessariamente à ordem das entradas correspondentes no Diretório. Etiquetas duplicadas são diferenciadas apenas pela localização dos respectivos campos dentro do registro. O Diretório termina com um caractere finalizador (hex 1E).

Campos variáveis – Os dados em um registro bibliográfico MARC 21, estão organizados em campos variáveis, cada um identificado por uma etiqueta de 3 caracteres numéricos, que estão registrados na entrada do diretório, referente a cada campo.

Há dois tipos de **campos variáveis**:

Campos de controle variável – São os campos 00X. Estes campos são identificados por uma etiqueta no Diretório, mas eles não contêm posições nem para indicadores nem para códigos de subcampo. Os campos de controle variável, são estruturalmente diferentes dos campos de dados variáveis. Eles podem conter um único dado ou uma série de dados de tamanho fixo, identificados pela posição relativa do caractere.

Campos de dados variáveis – São os restantes campos variáveis definidos no formato. Além de serem identificados por uma etiqueta no Diretório, os campos de dados variáveis contém duas posições para indicadores, localizadas no começo de cada campo, e dois caracteres para código de subcampo, precedendo cada dado dentro do campo. Os campos de dados variáveis, são agrupados em blocos, de acordo com o primeiro caractere da etiqueta, o qual, com algumas exceções, identifica a função do dado dentro do registro. O tipo de informação no campo, é identificada pelo restante da etiqueta.

Diante do surgimento de recursos informacionais em meio eletrônico, era preciso que o formato MARC passasse também por adequações para que pudesse atender as exigências de representação destes novos materiais. Na tentativa de acompanhar as mudanças de suportes, o MARC 21 apresenta em sua última atualização a inclusão do campo 856 para a descrição do endereço eletrônico dos recursos, o que possibilitou “hiperlinkar” as URL(s) proporcionando outra possibilidade de acesso aos recursos.

O MARC apresenta uma estrutura rígida, o que acaba sendo uma vantagem para as bibliotecas que fazem intercâmbio de registros bibliográficos. Entretanto, para a representação de recursos na rede, o formato MARC necessita de uma estrutura mais flexível que somada a sua estrutura descritiva completa, possibilitará uma representação de um recurso de modo detalhado e com a flexibilidade exigida no ambiente Web.

Essa flexibilidade será suprida pela linguagem XML, sendo assim, a tendência é que o MARC coexista com outros formatos e em outra versão. Essa nova versão refere-se ao MARC 21 em XML, ou MARCXML.

MARCXML

O MARCXML une todas as vantagens do formato MARC 21 (padronização, representação detalhada e rica, possibilidade de intercambiar informações etc) com as vantagens da linguagem de marcação XML (flexibilidade, extensibilidade etc) na tentativa de proporcionar uma melhor representação dos recursos informacionais em meio eletrônico. A proposta do MARCXML é oferecer solução para a representação de recursos informacionais da rede que exigem certa flexibilidade e extensibilidade.

De acordo com Esteves, Santos e Guimarães (2001),

Na área das bibliotecas, surgem os formatos MARC (Machine Readable Catalogue) para responder às necessidades de informatização de catálogos. No entanto, estes formatos não conseguem responder no contexto da Internet, pois não possuem uma linguagem de fácil aplicação por qualquer utilizador e que, ao mesmo tempo, possa ser interpretada pelos browsers. O XML aparece, no contexto da internet, para facilitar a difusão da informação documental. Como possui uma semântica própria, descreve a estrutura e conteúdo do documento, não a sua formatação, tornando-se, por isso, "revolucionário" em relação ao HTML que apenas possibilita a formatação dos dados no que respeita à sua apresentação gráfica, não fornecendo nenhum conteúdo semântico.

Siqueira (2003) aponta que a XML sozinha não é a solução, mas uma ferramenta que unida a um formato como o MARC permitirá a solução para o problema de representação de recursos eletrônicos em uma comunidade específica como a Biblioteconomia.

O MARCXML não perdeu as características advindas do MARC 21, na realidade, não houve alterações na estrutura do formato, o que ocorreu foi uma conversão para a linguagem XML. Sendo assim, continua sendo um formato indicado para a representação de qualquer tipo de material bibliográfico. Essa abrangência e a estrutura de descrição rica em detalhes fazem do MARCXML um exemplo de formato de metadados complexo na comunidade biblioteconômica.

Quanto à estrutura do MARCXML, Siqueira (2003, p. 82) nos mostra que,

A grande diferença entre o formato tradicional do MARC 21 e sua versão em XML está na estrutura usada para organizar os dados bibliográficos e catalográficos. Tradicionalmente marcado por uma seqüência de caracteres contidos em um arquivo de texto, a versão em XML do MARC 21 apresenta

uma estrutura mais organizada, hierárquica, exatamente como a principal característica da linguagem XML.

Podemos visualizar essa característica no exemplo da Figura 31, pois trata-se do mesmo documento representado anteriormente pelo MARC 21, só que agora em MARCXML:

```
<?xml version="1.0" encoding="UTF-8" ?>
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <leader>01142cam 2200301 a 4500</leader>
    <controlfield tag="001">92005291</controlfield>
    <controlfield tag="003">DLC</controlfield>
    <controlfield tag="005">19930521155141.9</controlfield>
    <controlfield tag="008">920219s1993 caua j 000 0 eng</controlfield>
    <datafield tag="010" ind1="" ind2="">
      <subfield code="a">92005291</subfield>
    </datafield>
    <datafield tag="020" ind1="" ind2="">
      <subfield code="a">0152038655 :</subfield>
      <subfield code="c">$15.95</subfield>
    </datafield>
    <datafield tag="040" ind1="" ind2="">
      <subfield code="a">DLC</subfield>
      <subfield code="c">DLC</subfield>
      <subfield code="d">DLC</subfield>
    </datafield>
    <datafield tag="042" ind1="" ind2="">
      <subfield code="a">lcac</subfield>
    </datafield>
    <datafield tag="050" ind1="0" ind2="0">
      <subfield code="a">PS3537.A618</subfield>
      <subfield code="b">A88 1993</subfield>
    </datafield>
    <datafield tag="082" ind1="0" ind2="0">
      <subfield code="a">811/.52</subfield>
      <subfield code="2">20</subfield>
    </datafield>
    <datafield tag="100" ind1="1" ind2="">
      <subfield code="a">Sandburg, Carl,</subfield>
      <subfield code="d">1878-1967.</subfield>
    </datafield>
    <datafield tag="245" ind1="1" ind2="0">
      <subfield code="a">Arithmetic /</subfield>
      <subfield code="c">Carl Sandburg ; illustrated as an anamorphic
      adventure by Ted Rand. </subfield>
    </datafield>
```

```

<datafield tag="250" ind1="" ind2="">
  <subfield code="a">1st ed.</subfield>
</datafield>
<datafield tag="260" ind1="" ind2="">
  <subfield code="a">San Diego :</subfield>
  <subfield code="b">Harcourt Brace Jovanovich,</subfield>
  <subfield code="c">c1993.</subfield>
</datafield>
<datafield tag="300" ind1="" ind2="">
  <subfield code="a">1 v. (unpaged) :</subfield>
  <subfield code="b">ill. (some col.) ;</subfield>
  <subfield code="c">26 cm.</subfield>
</datafield>
<datafield tag="500" ind1="" ind2="">
  <subfield code="a">One Mylar sheet included in pocket.</subfield>
</datafield>
<datafield tag="520" ind1="" ind2="">
  <subfield code="a">A poem about numbers and their characteristics.
  Features anamorphic, or distorted, drawings which can be restored to
  normal by viewing from a particular angle or by viewing the image's
  reflection in the provided Mylar cone.</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="0">
  <subfield code="a">Arithmetic</subfield>
  <subfield code="x">Juvenile poetry.</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="0">
  <subfield code="a">Children's poetry, American.</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="1">
  <subfield code="a">Arithmetic</subfield>
  <subfield code="x">Poetry.</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="1">
  <subfield code="a">American poetry.</subfield>
</datafield>
<datafield tag="650" ind1="" ind2="1">
  <subfield code="a">Visual perception.</subfield>
</datafield>
<datafield tag="700" ind1="1" ind2="">
  <subfield code="a">Rand, Ted,</subfield>
  <subfield code="e">ill.</subfield>
</datafield>
</record>
</collection>

```

FIGURA 31: Representação de um recurso pelo formato MARCXML.

FONTE: The Library of Congress (2005). Disponível em:

<<http://www.loc.gov/standards/marcxml/Sandburg/sandburg.xml>>.

Na Figura 31 podemos visualizar os mesmos elementos descritivos do exemplo anterior, com a diferença que neste, os elementos descritivos são marcados pelas Tag(s) iniciais <...> e finais </...> dos seguintes tipos: <collection>, <record>, <leader>,

<controlfield>, <datafield> e <subfield>. Siqueira (2003, p. 89) aponta as seguintes características presentes nas Tag(s) do MARCXML,

- <collection>: Início do registro, descrevendo informações sobre a coleção de documentos que será apresentada, por exemplo, a localização;
- <record>: Informa o tipo de registro que será apresentado, livro, mapa, etc;
- <leader>: O Líder do registro MARC 21;
- <controlfield>: Informações que se enquadram nas etiquetas 00X, por exemplo, 006, 007 e 008;
- <datafield>: Informações que se enquadram nas demais etiquetas, as que utilizam subcampos, por exemplo, 010, 100 e 245;
- <subfield>: Os subcampos.

De acordo com os estudos da Library of Congress (2002), as principais características do formato MARCXML, podem ser resumidas em oito tópicos, conforme descritos abaixo:

1. **MARCXML Schema simples e flexível:** a estrutura do MARCXML combina simplicidade e flexibilidade pelo fato de usar a linguagem XML para estruturação dos dados que serão representados pelo formato de descrição MARC 21. Portanto, o MARCXML Schema contém a semântica do MARC. A diferença é que seus campos e subcampos passaram a ser tratados como elementos-atributos;
2. **Conversão do MARC para XML:** todos os dados considerados essenciais em um registro MARC foram convertidos e expressos em XML sem perda de informações, a diferença é que os campos relacionados à estrutura, como posição de entrada (líder), não são necessários no XML e são deixados em branco ou são retirados;
3. **Reciprocidade entre XML e MARC:** o registro criado em XML, pode ser passado para o MARC sem perda de informações e vice-versa;
4. **Apresentação dos dados:** uma vez criado ou convertido para XML, é possível apresentar dados descritos em formato MARC por meio de uma folha de estilo;
5. **MARC editing:** é possível atualizar e alterar dados com uma simples alteração na estrutura em XML;

6. **Conversão dos dados:** muitos dados podem ser escritos e convertidos em XML, sendo assim, vários softwares podem ler MARCXML;
7. **Validação dos dados MARC:** é dada através de uma ferramenta de software externo ao esquema e que apresenta três níveis para o processo:
 - Básica: de acordo com o esquema XML;
 - Validação da tag MARC 21: que opera nos campos e subcampos;
 - Validação de conteúdo do registro: ex. valores de códigos, datas e tempos.
8. **Extensibilidade:** o uso da XML como estrutura dos registros MARC possibilita aos usuários a construção de suas próprias ferramentas para consumir, manipular e converter dados MARC e, além disto, usá-los de acordo com suas necessidades informacionais.

Por comportar toda a estrutura do MARC 21, o MARCXML permite a conversão de um registro tradicional MARC 21 para MARCXML sem perda de informações e ainda possibilita trabalhar com outros tipos de estruturas ou formatos. Contudo, por se tratar de um formato complexo e altamente estruturado, se for feita a conversão de outros formatos (simples ou estruturados) para o MARCXML é provável que se tenha uma perda de informações, já que se tratam de formatos menos ricos em sua representação.

Podemos destacar algumas vantagens para o uso do formato MARCXML:

- Pode ser usado para a descrição de qualquer recurso, independente do suporte (LIBRARY OF CONGRESS, 2002);
- Permite a conversão de um registro MARC para MARCXML sem perda de informações (LIBRARY OF CONGRESS, 2002);
- Permite que diferentes softwares trabalhem com a versão desse formato (LIBRARY OF CONGRESS, 2002);
- Permite diferentes formas de representação documentárias a partir das necessidades do usuário (SIQUEIRA, 2003);

- Apresenta-se adequado para a representação de recursos em meio eletrônico, pois combina as vantagens do formato MARC (representação detalhada e rica) com as vantagens da linguagem de marcação XML (flexibilidade e extensibilidade);

5.3 ALGUMAS CONSIDERAÇÕES SOBRE OS FORMATOS DE METADADOS SIMPLES, ESTRUTURADOS E RICOS

A representação de um recurso informacional tem como objetivo facilitar e simplificar sua busca e recuperação; e intermediar a comunicação entre usuários e o conhecimento registrado disponível em um determinado ambiente informacional (MEY, 1995; PEREIRA; SANTOS, 1998). Entretanto, é preciso que essa representação seja padronizada, por meio do uso de normas, códigos, formatos e padrões de metadados que estabeleçam regras para fornecer assim, a base para uma recuperação de qualidade. Nesse sentido, faz-se necessário neste momento uma pequena reflexão sobre os formatos de metadados tratados nesta pesquisa.

As META TAG(s), o uso de URI e os dados trocados na transferência do protocolo HTTP apresentam-se como sendo os formatos de metadado mais simples, contudo, possuem valores significativos, pois são iniciativas importantes na tentativa estabelecer a localização e recuperação dos recursos informacionais na rede.

Entretanto, o uso de formatos de metadados mais estruturados ou ricos para a representação de recursos informacionais em geral na Web é pouco frequente. Hoje em dia o modo mais comum de representação dos recursos informacionais ocorre com o uso de META TAG(s). Elas proporcionam uma representação mínima do recurso e são as únicas que

comportam algum tipo de representação e descrição do conteúdo dos recursos informacionais. De acordo com Weibel (2000?) atualmente o modo mais fácil de encontrar metadados na rede está relacionado aos metadados embutidos no cabeçalho do código HTML, mais especificamente nas META TAG(s). A inclusão de metadados simples nas TAG(s) META permite que o recurso apresente algum tipo de dado que o represente seu conteúdo, estes metadados embutidos no cabeçalho auxiliam na recuperação dos recursos, pois são indexados pelos agentes das ferramentas de busca. A desvantagem segundo Weibel (2000?) seria a falta de controle formal sobre a representação, pois muitas vezes não são informadas nessas TAG(s), por exemplo, as atualizações feitas no conteúdo do recurso, o que conduz a uma inconsistência na descrição.

Por esse motivo, foram sendo criados formatos de metadados que proporcionassem uma representação mais detalhada do recurso, dando origem aos formatos estruturados e ricos. A tendência é que formatos dessas categorias passem a ser mais utilizados. Os formatos que se destacam em cada categoria são: o Dublin Core como formato estruturado e o MARC em sua versão atual em XML, como um formato rico.

É importante lembrar que cada formato apresenta sua importância, entretanto, é preciso considerar também que cada um fornecerá um tipo de representação e que quanto mais específica, completa e detalhada for a representação, melhor será a recuperação dos recursos informacionais (MILLER, 1996).

Partindo desse ponto de vista é preciso estabelecer algumas diferenças entre o formato DC e o MARC, pois irão proporcionar representações diferentes de um mesmo recurso.

O formato de metadados Dublin Core vem sendo indicado como uma opção para a representação de recursos informacionais na rede, entretanto, apesar de ser um formato de metadados estruturado, pode não ser eficiente na descrição de um recurso para uma

comunidade específica, que necessita de representações detalhadas de um recurso informacional para ampliar suas formas de acesso e uso.

O Dublin Core é um formato ou padrão que apresenta um conjunto de elementos de metadados com o objetivo de promover a descoberta de recursos. (WOODLEY, CLEMENT, WINN, 2003).

O Dublin Core é um padrão criado para possibilitar a localização de recursos informacionais em geral disponibilizados em meio eletrônico, em outras palavras, o DC proporciona por meio de seus elementos de descrição uma representação para identificação e localização do recurso informacional na rede.

Já o MARC pode ser definido como sendo um “[...] formato padronizado para o armazenamento e intercâmbio de registros bibliográficos e informações relacionadas em formato legível por máquina” (BRITISH LIBRARY, 2003). Em outras palavras, é um formato padronizado para a representação, armazenamento e comunicação ou intercâmbio de registros bibliográficos legíveis por máquina (BRITISH LIBRARY, 2003).

Sendo específico da comunidade biblioteconômica, o MARC foi desenvolvido com uma estrutura para representar detalhadamente os recursos informacionais, com o intuito de promover com seu produto, o registro bibliográfico, o intercâmbio das informações bibliográficas e catalográficas entre bibliotecas, além de auxiliar os usuários na localização dos recursos desejados.

A grande diferença entre esses formatos está no objetivo pelo qual foram desenvolvidos. O Dublin Core tem como objetivo a identificação para a localização do recurso informacional, ou seja a descoberta dos recursos na rede. E o MARC tem como objetivo construir um registro bibliográfico de modo que o mesmo represente um recurso no que diz respeito ao seu conteúdo e sua forma permitindo o intercâmbio deste registro e a

localização do documento que representa independente do ambiente em que este recurso se encontra.

Contudo, a evolução nas versões do MARC, possibilitou a inclusão do campo 856 para a indicação da URL e mais recentemente a versão do MARC em XML, o que facilitou ainda mais representação dos recursos informacionais em meio eletrônico neste formato.

Partindo desse ponto de vista e sabendo que o MARC é um formato de metadado que representa qualquer item documentário, independente do tipo de suporte, consideramos o MARC em sua versão atual em XML, como o formato de metadado ideal para fornecer uma representação detalhada e completa dos recursos informacionais na Web, pois quanto mais detalhada for a representação, melhor será a recuperação. Além disso, os métodos tradicionais de representação, já estabelecidos na Ciência da Informação, trarão maior credibilidade e segurança no estabelecimento da Web Semântica.

Atualmente há uma grande variedade de formatos de metadados criados na tentativa de organizar os recursos informacionais disponíveis na Internet, e é necessário pensar que a variedade de formatos de metadados requer uma padronização na representação dos recursos, mas não se pode pensar na adoção de um formato único para a descrição dos recursos da rede, pela variedade de interesses e de recursos disponíveis. Entretanto, é necessário perceber que os metadados utilizados deverão obedecer a padrões de descrição definidos por interesses e objetivos específicos das áreas que representam, pois é por meio deles que será garantida uma busca e recuperação de qualidade dos recursos. É a partir deles que o estabelecimento da Web Semântica será possível.

5.4 METADADOS: a chave para a representação de recursos informacionais na Web Semântica

Uma característica marcante da Web é a heterogeneidade de recursos informacionais disponibilizados. Essa característica também se estenderá a Web Semântica, com a diferença que nesta os recursos estarão marcados semanticamente, proporcionando uma melhor recuperação. A heterogeneidade dos recursos informacionais não deixará de existir, pelo contrário, a tendência é que surjam novos tipos de recursos, assim, a questão da representação dos recursos é um dos principais fatores a serem solucionados na Web Semântica.

É nítido que o problema que enfrentamos atualmente com a questão da recuperação da informação na rede está relacionado com a falta de representação adequada dos recursos. Mas independentemente do ambiente informacional que nos referimos, os problemas para recuperar essa crescente quantidade de informação são os mesmos. A diferença está na amplitude alcançada pela Web, pois nele estamos trabalhando com uma quantidade de informação maior do que em qualquer outro ambiente informacional. Nesse sentido, os métodos de representação são ferramentas indispensáveis para individualização dos itens disponíveis, bem como para a recuperação da informação, pois é responsável por multidimensionar a forma de acesso possibilitando uma gama variada para a recuperação precisa e eficiente é valorizada e necessária. Portanto, a recuperação de qualquer recurso informacional, esteja ele em meio digital ou não, só é possível mediante uma representação.

De acordo com Milstead e Feldman (1999), qualquer ferramenta que torne mais fácil o processo de busca e recuperação dos recursos informacionais na Web é importante. Afirmam também que os metadados são cruciais para melhorar o processo de busca e

recuperação, pois proporcionam uma padronização na estruturação e representação do conteúdo dos recursos.

Ortiz-Repiso Jiménez (1999, p. 219, tradução nossa), afirma que os metadados são importantes para a recuperação da informação na Internet por diversas razões entre elas destaca que os metadados,

- Permitem indexar grandes quantidades de dados de diferentes tipos [...].
- Ajuda a descobrir e recuperar recursos na rede, pois analisam o conteúdo do objeto em profundidade.
- Compartilham e integram recursos de informação heterogêneos e localizados em sites diversos.
- Podem controlar o acesso a informações restritas.

De acordo com Gilliland-Swetland (1999) podemos considerar os metadados como sendo algo mais que simplesmente elementos descritivos, pois são de significativa importância para proporcionar pontos de acesso adicionais para o conteúdo dos recursos, trazendo benefícios para os sistemas digitais de recuperação da informação.

Conforme afirma Robredo (2004), estamos presenciando uma nova mudança de paradigma, o foco era o documento, passa a ser seu conteúdo, o que torna necessário o desenvolvimento de novas técnicas para a representação não só do documento, mas também de seu conteúdo informacional.

Baseada nessa afirmação de Robredo e nas características dos metadados estudadas neste trabalho pode-se dizer que os metadados são chaves para o estabelecimento de uma organização e representação dos recursos informacionais não só na web atual, como principalmente na Web Semântica e em qualquer outro sistema de recuperação da informação.

A representação dos recursos informacionais por metadados irá formar a base necessária para se constituir a Web Semântica e unidos as outras tecnologias estudadas: ontologias, linguagem XML e arquitetura RDF, irão proporcionar uma rede de conhecimento interligados.

Cada tipo de formato ou padrão metadado fornecerá um tipo de representação, e quanto mais específica, completa e detalhada for esta representação, mais pontos de acesso ao recurso ela fornecerá e melhor será a recuperação. Uma representação mais detalhada do recurso proporcionará sua maior identificação, individualizando-os, tornando-os únicos entre muitos, multiplicando as formas de acesso a eles e possibilitando assim, uma recuperação mais precisa. Assim, entre os formatos de metadados estudados neste trabalho, aponto o MARC em sua versão em XML - MARC XML - como o formato de metadado ideal para proporcionar uma representação detalhada dos recursos informacionais, no que diz respeito ao seu conteúdo e sua forma na área de Ciência da Informação.

6 CONSIDERAÇÕES FINAIS

À medida que as tecnologias de informação e comunicação foram sendo desenvolvidas e aperfeiçoadas, os sistemas de recuperação de informação se depararam com uma nova realidade, a quantidade cada vez maior de recursos informacionais disponíveis em ambientes tecnológicos.

Alvarenga (2001) aponta que o meio digital se constitui em um espaço sem precedentes para armazenamento, disponibilização e recuperação da informação e que necessita de novos elementos que facilitem a recuperação dos variados tipos de recursos, pois houve uma mudança nos meios em que a informação passou a ser veiculada, produzida e registrada.

A parte substancial dos documentos que se refere a seu conteúdo, à sua atenção, ao seu significado, aos enunciados que compõem os conceitos neles contidos, tudo isso continua invariável; tudo isso é uma contingência com a qual as máquinas têm que conviver e daí decorre a dificuldade primordial do processo de tratamento da informação, antes em ambientes tradicionais e hoje na Web (ALVARENGA, 2001).

É nesse cenário de surge a Web Semântica, um projeto a ser desenvolvido a longo prazo pela equipe de pesquisadores da W3C, que visa uma melhor utilização do vasto repositório de informações disponíveis na Web, ou seja, um uso mais produtivo e significativo dos recursos informacionais.

A transição da World Wide Web para a Web Semântica busca, por meio de uma estrutura tecnológica e métodos de representação do conhecimento, possibilitar o acesso a um sistema de recuperação mais eficiente. De acordo com Souza e Alvarenga (2004, p. 134),

O projeto da Web Semântica, em sua essência, é a criação e implantação de padrões (standards) tecnológicos para permitir este panorama, que não somente facilite as trocas de informações entre agentes pessoais, mas principalmente estabeleça uma língua franca para o compartilhamento mais significativo de dados entre dispositivos e sistemas de informação de uma maneira geral.

Para atingir tal propósito, o projeto da Web Semântica conta com a implementação de recursos tecnológicos e técnicas de representação da informação presentes em várias áreas do conhecimento. Na realidade o estabelecimento da Web Semântica requer um esforço interdisciplinar entre diversas áreas, dentre elas destacamos: a área da Ciência da Computação, que fornece ferramentas e a estrutura tecnológica; áreas como a Inteligência Artificial, que fornece meios para se estabelecer o raciocínio sobre os dados; e a Ciência da Informação, que proporciona os métodos e técnicas para a representação da informação, e consequentemente áreas como a lingüística e semiótica.

O crescimento do número de recursos na rede e a falta de eficiência dos mecanismos de busca atuais, exigiram por parte da Ciência da Computação, novas ferramentas e metodologias para que os recursos informacionais pudessem ser recuperados de modo mais eficiente.

Nesse contexto, a questão da recuperação da informação em ambiente eletrônico acaba por envolver a Ciência da Informação, pois é necessário perceber que, a Internet, com seu grande volume de informações, e os usuários cada vez mais exigentes por buscas precisas e rápidas, tem se apresentado como ambiente propício para o desenvolvimento de processos e métodos de representação, armazenamento e busca de informações mais eficientes.

Sendo assim, para estabelecermos a relação da Web Semântica com a Ciência da Informação é preciso, antes de tudo, conceitua-la.

Segundo Carvalho (1999, p. 51), a Ciência da Informação é uma ciência recente e,

[...] surgiu da demanda social pela otimização dos processos de coleta, armazenamento, recuperação e disseminação da informação científica e tecnológica, cuja produção apresentava um crescimento exponencial ao final da década de 50 - a chamada "crise da informação".

De modo geral, a Ciência da Informação é uma área voltada para as questões científicas e práticas profissionais relacionadas aos registros, comunicação e uso do

conhecimento registrado tendo como base o uso das tecnologias informacionais vigentes (SARACEVIC, 1996). Sendo assim, podemos dizer que a,

Ciência da Informação é a disciplina que investiga as propriedades e o comportamento da informação, as forças que governam seu fluxo, e os meios de processá-la para otimizar sua acessibilidade e uso. A CI está ligada ao corpo de conhecimentos relativos à origem, coleta, organização, estocagem, recuperação, interpretação, transmissão, transformação e uso da informação[...] Ela tem tanto um componente de ciência pura, através da pesquisa dos fundamentos, sem atentar para sua aplicação, quanto um componente de ciência aplicada, ao desencadear produtos e serviços (BORKO apud SARACEVIC, 1996, p. 45).

Sabendo que a Ciência da Informação é uma área voltada para a otimização do fluxo informacional, dos processos de coleta, tratamento, armazenamento, recuperação, disseminação da informação científica e tecnológica, da comunicação e uso do conhecimento registrado tendo como base o uso das tecnologias informação e comunicação vigente, podemos dizer que ela estabelece uma estreita relação com a Web Semântica.

A Ciência da Informação sempre se preocupou com questões que envolvem a representação da informação e a disponibilização de informações em grandes repositórios de informação. Com os avanços nas tecnologias de informação e comunicação, a Ciência da Informação passou a ter um novo ambiente de atuação, o meio eletrônico. Entretanto, as técnicas e metodologias para a representação dos recursos informacionais, apesar de terem sido adaptadas para este novo meio, mantém a essência das técnicas e metodologias tradicionais. O que ocorre nesse novo cenário não é uma novidade para os profissionais da Ciência da Informação, que sempre se depararam com a tarefa de organizar, representar e disponibilizar grandes quantidades de informações em ambientes variados. A diferença agora é que isso vem ocorrendo em meio eletrônico, numa escala gigantesca por se tratar da Web.

Sendo assim, a Web Semântica estabelece uma relação com a Ciência da Informação no seguinte âmbito: por se tratar de uma iniciativa tecnológica que estabelece uma melhor estruturação dos dados e representação do conteúdo dos recursos para uma posterior recuperação, há na Web Semântica uma semelhança com os sistemas de recuperação da

informação, nos quais a Ciência da Informação estabelece regras, técnicas, padrões e metodologias para proporcionar a necessária representação das informações ou do conhecimento registrado e assim, proporcionar uma eficiente recuperação da informações nesses sistemas. Desse modo, podemos dizer que a Web Semântica se assemelha com os Sistemas de Recuperação da Informação, porque apresenta em sua proposta, as tarefas de representação, armazenamento, organização e acesso aos recursos informacionais.

De acordo com Codina (2003) a proposta da Web Semântica pode ser comparada com a estrutura de uma base de dados que apresenta dados “etiquetados”, ou seja, marcados em um “campo” que possui um atributo correspondente bem definido. Partindo desse ponto de vista, podemos verificar que não há nada de novo para o profissional da Ciência da Informação que trabalha há algum tempo com base de dados. O que muda é o novo modo de organizar os recursos informacionais com o emprego das ferramentas tecnológicas que compõe a Web Semântica.

É difícil prever o futuro da Web Semântica, entretanto, autores como Souza e Alvarenga (2004) apontam que a Web semântica fornecerá alternativas para qualquer sistema de recuperação da informação e que provavelmente proporcionará mudanças nas atividades dos profissionais da área da Ciência da Informação. Como novas formas de trabalho os autores Souza e Alvarenga (2004) apontam que há possibilidades para as seguintes atividades:

a) Projetos de novos e melhorados motores de busca: é possível criar melhores motores de busca utilizando as técnicas de representação e recuperação da informação estabelecida na Web Semântica. Com a marcação semântica nos recursos informacionais disponibilizados na Web é possível que as ferramentas de busca usem técnicas automáticas para o "entendimento" do conteúdo dos recursos e assim, possibilitar uma recuperação mais eficiente. Mas para que isso ocorra efetivamente é preciso que os motores de busca também se modifiquem, pois necessitam de tecnologias mais específicas como o uso de agentes inteligentes.

- b) Construção de novas interfaces com o usuário para sistemas de informação: as interfaces dos sistemas de recuperação de informação também poderão sofrer alterações e poderão se tornar mais parecidas com o funcionamento cognitivo dos seres humanos, pois o uso de agentes inteligentes possibilita a utilização dos perfis dos usuários e uma interação mais significativa com o sistema.
- c) Construção automática de tesouros e vocabulários controlados: os autores Souza e Alvarenga (2004) apontam que poderá surgir novas metodologias para a construção automática de tesouros e vocabulários controlados, a partir das marcações semânticas existentes nos recursos informacionais, pelas ontologias disponibilizadas nas diversas áreas do conhecimento e pelas declarações de relação entre recurso e representação estabelecidas na arquitetura RDF.
- d) Indexação automática de documentos: com o uso efetivo de ontologias e metadados nas diversas comunidades de interesse acredita-se que futuramente será possível desenvolver novas metodologias para analisar automaticamente os recursos representados e classificá-los automaticamente.
- e) Gestão do conhecimento organizacional: as tecnologias disponíveis na Web Semântica atuarão significativamente na área da gestão do conhecimento organizacional, os portais corporativos (símbolos da gestão do conhecimento), apresentarão maior funcionalidade devido as melhores possibilidades de recuperação e interoperabilidade proporcionada pelas ferramentas da Web Semântica, bem como melhor representação do conhecimento ou do capital intelectual da organização, proporcionado pelos metadados e ontologias.
- f) Gestão da informação estratégica e da inteligência competitiva: As ferramentas da Web Semântica também afetarão a gestão da informação estratégica e inteligência competitiva, pois aqui, os agentes que irão automatizar e agilizar a colheita de informações estratégicas que auxiliarão as tomadas de decisões em um ambiente ou empresa.

Diante do que foi apresentado, podemos verificar que a Web Semântica estabelece uma relação com a Ciência da Informação, pois envolve a aplicação de tecnologias e a representação do conhecimento para proporcionar uma boa recuperação da informação. A Ciência da Informação oferece, por meio das técnicas e metodologias para a representação do conhecimento, a base necessária para estabelecer uma representação mais adequada dos recursos informacionais no ambiente tecnológico proporcionado pela Web Semântica.

Embora tenha sido apontada como um caminho para solucionar os problemas de recuperação da informação é preciso destacar o exagero por parte de alguns autores, quando afirmam, por exemplo, que a Web Semântica é indicada como “promessa” para solucionar os problemas de recuperação da informação na Web e que permitirá que “computadores entendam” o significado dos dados. Na realidade, a Web Semântica se apresenta com um caminho na busca por uma solução mais adequada para as questões de recuperação da informação na Web, por meio da construção de formas de representação das informações onde a capacidade de compreender o significado dos dados, ficará a cargo dos agentes de softwares. Entretanto, promover esse “entendimento” por parte dos agentes requer a implementação de estruturas complexas, que ainda estão em desenvolvimento nos estudos de Inteligência Artificial.

Os sistemas computacionais não assimilam a informação como nos ambientes operacionalizados por humanos, mas poderão manipular dados com mais eficiência de maneira até mesmo, mais significativa, mas para que isto ocorra, a construção adequada de formas de representação dos recursos informacionais se faz necessária, de modo que seu conteúdo semântico fique devidamente marcado para que os agentes de software possam utilizá-los com mais eficiência.

Diante disso, é preciso lembrar também que os estudos sobre a Web Semântica ainda estão no início e embora seu futuro pareça ser promissor, não se tem um consenso sobre a

direção provável que ela tomará, pois algumas das tecnologias e ferramentas que a compõe ainda estão sendo avaliadas.

A Web Semântica será completamente estabelecida, conforme aponta seus idealizadores, a partir da organização dos recursos. Essa organização se realizará por meio da implementação de ferramentas tecnológicas, vindas da Ciência da Computação e de ferramentas que tratam da representação do conhecimento, baseadas na área da Ciência da Informação. A infraestrutura tecnológica para a implementação da Web Semântica, já está estabelecida, apesar de ainda estar em estudo e aperfeiçoamento. Ela só funcionará efetivamente com o uso dos métodos de representação da informação, principalmente com o uso de metadados. Nesse sentido, é importante destacar que os metadados se apresentam como um fator chave para promover a representação dos recursos informacionais na Web Semântica.

É difícil dizer se toda a Web se transformará em Web Semântica, devido sua extensão. Por enquanto, o que vemos na literatura é que isso será difícil de ocorrer, pois temos acesso apenas a uma parcela de recursos disponibilizados atualmente. O que podemos dizer é que futuramente o conhecimento registrado e disponibilizado na Web será dividido em comunidades de interesse e a Web Semântica se estabelecerá nessas comunidades.

A Web Semântica se efetivará em três escalas: pequena, média e grande.

a) Em pequena escala teremos comunidades específicas fazendo uso das tecnologias e de ferramentas que a Web Semântica propõe. Como exemplo podemos citar uma biblioteca digital que utiliza as tecnologias e ferramentas da Web Semântica para estruturar o Sistema de Recuperação da Informação de sua base de dados.

b) Em média escala teremos comunidades distintas, mas que compartilham e tem em comum áreas afins, como um exemplo dessa aplicação podemos citar duas ou mais bibliotecas digitais

de instituições diferentes que utilizam as tecnologias da Web Semântica e que, unidas, irão compor uma rede maior de conhecimentos interligados.

c) Em grande escala teremos a união de várias comunidades de interesses distintos, que compartilharão recursos e formação uma rede de conhecimentos, baseada na estrutura da Web Semântica, como por exemplo, várias bibliotecas digitais e portais de pesquisa de assuntos diversos, que utilizam a estrutura da Web Semântica para compartilhar informações.

A tendência é haver cada vez mais compartilhamento de informações na Web, mesmo entre comunidades de interesses distintos e a Web Semântica oferece a estrutura necessária para proporcionar esse compartilhamento. Assim, questões como interoperabilidade, representação de recursos e estabelecimento de conceitos semânticos deverão ser adotados e solucionados nessas comunidades.

Conforme estabelecido na introdução desse trabalho, a Web Semântica tem sido indicada como um caminho para solucionar e melhorar a busca e a recuperação das informações na rede, pois visa a proporcionar o acesso automatizado aos recursos informacionais, por meio da representação da informação. Ela será estabelecida com o trabalho conjunto de várias outras ferramentas tecnológicas, entretanto a base para sua construção está no uso de metadados para a representação dos recursos informacionais.

Com o desenvolvimento dessa pesquisa foi possível atingir os objetivos propostos: estabelecer a relação da Web Semântica com a Ciência da Informação, bem como, abordar as ferramentas tecnológicas necessárias para o estabelecimento da Web Semântica (linguagem de marcação XML, arquitetura de metadados RDF, ontologias e em especial os metadados para a representação de recursos informacionais).

Diante dos objetivos alcançados, podemos considerar mais especificamente como resultados que:

- 1) A Web Semântica estabelece uma relação com a Ciência da Informação principalmente pelos métodos e técnicas necessários para estabelecer a representação dos recursos informacionais. Dentre eles se destaca principalmente o uso de metadados.
- 2) O desenvolvimento e estabelecimento da Web Semântica possibilitará uma melhor recuperação dos recursos informacionais, pois além do uso de agentes de softwares responsáveis por "entender" significados e pela recuperação mais eficiente, há a implementação de ontologias que garantirão uma melhor definição do significado ou semântica dos dados estabelecidos pelos metadados;
- 3) O uso da linguagem de marcação XML na estrutura da Web Semântica é essencial, pois irá garantir uma maior flexibilidade e extensibilidade, além de possibilitar maior enfoque ao conteúdo dos recursos informacionais e não somente a sua forma de apresentação;
- 4) O estabelecimento de interoperabilidade pelo uso de arquiteturas de metadados, em especial a arquitetura RDF também se apresenta como fundamental, pois além de ser uma recomendação da W3C a arquitetura RDF, unida a linguagem XML e as ontologias, proporcionam a interoperabilidade nos três níveis necessários: nível sintático, estrutural e semântico;
- 5) O uso e aplicação de metadados apresentam-se como base fundamental para o estabelecimento de uma representação dos recursos informacionais no desenvolvimento da Web Semântica.

Portanto, podemos considerar então, que as tecnologias da Web Semântica convergem para a área de Ciência da Informação, estabelecendo uma estreita relação na questão da representação da informação, principalmente no que diz respeito ao uso de metadados que são considerados essenciais para se estabelecer a representação dos recursos informacionais na Web. Sendo assim, para o futuro estabelecimento de uma rede de conhecimentos interligados,

o desenvolvimento e implantação da Web Semântica irá ocorrer a partir do uso intensivo de metadados para a representação das informações ou do conhecimento registrado.

"Mudam-se os meios, sofisticam-se os instrumentos e surgem nomes novos para designar coisas velhas. Entretanto, a essência das coisas permanece". Essa citação de Alvarenga (2001) reflete algumas questões tratadas nessa pesquisa e demonstra que mesmo com os avanços tecnológicos a essência do tratamento da informação, a necessidade de compartilhamento e a construção de formas de representação do conhecimento existe e sempre existiu em qualquer ambiente informacional, seja ele eletrônico ou não. E nesse cenário, ressaltamos que as técnicas de representação da informação estabelecidas na área da Ciência da Informação serão essenciais para criar a base para o estabelecimento da Web Semântica.

Espera-se que essa pesquisa tenha contribuído para com os profissionais da Ciência da Informação por proporcionar um referencial teórico sobre o tema Web Semântica e o uso de metadados, pois apresentam-se como caminhos para atingir o objetivo de proporcionar uma melhor recuperação dos recursos informacionais na Web.

Além disso, é importante destacar que o profissional da área da Ciência da Informação exercerá um papel muito importante nesse novo cenário, onde o aumento de informações disponíveis é uma constante. Sua atuação se responsabilizará pela organização, tratamento, armazenamento, recuperação e disseminação das informações, e será fundamental para proporcionar uma otimização do ambiente no sentido de facilitar a identificação, a localização e a recuperação de recursos informacionais.

REFERÊNCIAS

AFONSO, M. M. R. *Semantic Web*. [S. l.: S. n.], 2001. Disponível em: <<http://paginas.fe.up.pt/~mgi00014/ari/SW.doc>>. Acesso em: 26 jun. 2005.

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. *Data Gram Zero – Revista de Ciência da Informação*, v. 2, n. 6, dez. 2001. Disponível em: <http://www.dgzero.org/dez01/Art_05.htm>. Acesso em: 31 jan. 2003.

BAGGIO, R. A sociedade da informação e a infoexclusão. *Ciência da Informação*, Brasília, v. 29, n. 2, p. 16-21, ago. 2000.

BARRETO, C. M. *Modelo de metadados para a descrição de documentos eletrônicos na web*. 1999. 189 f. Dissertação (Mestrado em Ciências em Sistemas de Computação)–Instituto Militar de Engenharia, Rio de Janeiro, 1999. Disponível em: <<http://ipanema.ime.eb.br/~de9/teses/1999/>> . Arquivo: cássia.zip. Acesso em: 13 set. 2001.

BAX, M. P. Introdução às linguagens de marcas. *Ciência da Informação*, Brasília, v. 30, n. 1, p. 32-38, jan./abr. 2001.

BERNERS-LEE, T. *Axioms of Web Architecture: Metadata*. [S. l.: S. n.], 1997. Disponível em: <<http://www.w3.org/DesignIssues/Metadata.html>>. Acesso em: 09 abr. 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*. [S. l.: S. n.], 2001?. Disponível em: <<http://www.scientificamerican.com>>. Acesso em: 09 jan. 2005.

BÉZIVIN, J. *Who's afraid of ontologies?* [S. l.: S. n.], 1998. Disponível em: <<http://www.metamodel.com/oopsla98-cdif-workshop/bezivin1/>>. Acesso em: 11 maio 2005.

BORGES, M. A. G. A compreensão da sociedade da informação. *Ciência da Informação*, Brasília, v. 29, n. 3, p. 25-32, dez. 2000.

BRANSKI, R. M. Localização de informações na Internet; características e formas de funcionamento dos mecanismos de busca. *Transinformação*, Campinas, v. 12, n. 01, p. 11-19, jan./jun. 2000.

BRENE, D. C. G. *Padrões de metadados para a representação descritiva de documentos eletrônicos: uma análise comparada entre AACR2, MARC 21, MARCXML, e Dublin Core*. 2004. 210 f. Trabalho de Conclusão de Curso, (Graduação em Biblioteconomia)-Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2004.

BRITISH LIBRARY. *Exchange Formats*. [S. l.: S. n.], 2003. Disponível em: <<http://www.bl.uk/services/bibliographic/exchange.html>>. Acesso em: 17 set. 2005.

CARVALHO, E. C. A natureza social da ciência da informação. In: PINHEIRO, L. V. R. (Org.). *Ciência da informação, ciências sociais e interdisciplinaridade*. Brasília : IBICT, 1999, p. 51-63.

CASTRO, E. *XML para a World Wide Web*. Tradução de Hugo de Souza Melo. Rio de Janeiro: Campus, 2001. 269p. (Visual quickstar guide).

CENDÓN, B. V. Ferramentas de busca na Web. *Ciência da Informação*, Brasília, v. 30, n. 1, p. 39-49, jan./abr. 2001.

CENDÓN, B. V.; KREMER, J. M. (Org.). *Fontes de informação para pesquisadores e profissionais*. Belo Horizonte: UFMG, 2000. p. 191-198.

CODINA, L. Internet invisible y web semántica: el futuro de los sistemas de información em línea? *Tradumática*, [S. l.], n. 2, nov. 2003. Disponível em: <<http://www.fti.uab.es/tradumatica/revista>>. Acesso em: 09 jan. 2005.

CUNHA, L. M. S. Web Semântica: um estudo preliminar. *Documentos*, Campinas, v. 18, out. 2002.

DODEBEI, V. L. D. *Tesouro: linguagem de representação e memória documentária*. Niterói: Intertexto. 2002.

DUBLIN Core Metadata Template [DC Qualificado]. [S. l.: S. n.], 1997a. Disponível em: <<http://www.lub.lu.se/cgi-bin/nmdc.pl>>. Acesso em: 13 mar. 2005. em: 18 jun. 2004.

DUBLIN Core Metadata Template [DC não Qualificado]. [S. l.: S. n.], 1997b. Disponível em: <<http://www.lub.lu.se/cgi-bin/nmdc.pl?lang=en&save-info=on&simple=1>>. Acesso em: 13 mar. 2005.

ESTEBAN VILLAMIZAR, L. A. *Un punto intermedio entre la actual web y la futura web semántica*. Madrid: Universidad Carlos II de Madrid. [2002?]. Disponível em: <http://www.cidlisuis.org/aedo/RGTIN2V1/RGTI_02.pdf>. Acesso em: 28 nov. 2004.

ESTEVEVES, A.; SANTOS, L.; GUIMARÃES, P. *XML nas bibliotecas digitais: Standard*. [S. l.: S. n.], 2001. Disponível em: <http://www.bibliosoft.pt/projectoxml/_standard.htm>. Acesso em: 10 abr. 2005.

FARIA, C. G. de; GIRARDI, R. *Uma análise da Web Semântica e suas implicações no acesso à informação*. [2002?]. Disponível em: <<http://maae.deinf.ufma.br/Ensino/IA/Uma%20An%C3%A1lise%20da%20Web%20Sem%C3%A2ntica%20e%20suas%20Implica%C3%A7%C3%B5es%20no%20Acesso%20%C3%A0%20Informa%C3%A7%C3%A3o.PDF>>. Acesso em: 07 fev. 2003.

FERNEDA, E. *Recuperação de Informação: análise sobre a contribuição da Ciência da Informação para a Ciência da Computação*. 2003, 137 f. Tese (Doutorado em Ciência da Comunicação)—Escola de Comunicações e Artes da Universidade de São Paulo, São Paulo, 2003.

FERREIRA, M. M. (Trad. e Adap.). *MARC 21: formato condensado para dados bibliográficos*. 2.ed. Marília: Universidade Estadual Paulista, 2002. v. 1.

GILLILAND-SWETLAND, A. J. La definición de los metadatos. In: INTRODUCCIÓN a los metadatos: vías a la información digital. [S. l.]: GETTY, 1999. p. 1-9.

GRÁCIO, J. C. A. *Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade*. 2002. 127 f. Dissertação (Mestrado em Ciência da Informação)—Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2002.

GUARINO, N. *Understanding, building, and using ontologies*. [S. l.: S. n.], 1996. Disponível em: <<http://ksi.cpsc.ucalgary.ca/KAW/KAW96/guarino/guarino.html>>. Acesso em: 11 maio 2005.

GRUBER, T. *What is na ontology?* [S. l.: S. n.], 1996. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 11 maio 2005.

GUIMARÃES, C. *Introdução a linguagem de marcação: HTML, XHTML, SGML, XML*. [S. l.: S. n.], 2004. Disponível em: <<http://www.dcc.unicamp.br/~celio/inf533/docs/markup.html>>. Acesso em: 13 mar. 2005.

HILLMANN, D. *Using Dublin Core: the elements*. [S. l.: S. n.], 2003. Disponível em: <<http://dublincore.org/documents/usageguide/elements.shtml>>. Acesso em: 13 mar. 2003.

IANNELLA, R. *Mostly metadata: a bit smarter technology*. [S. l.: S. n.], 1998. Disponível em: <<http://archive.dstc.edu.au/RDU/reports/VALA1998/>>. Acesso em: 03 jul. 2005.

IKEMATU, R. S. *Gestão de Metadados: sua evolução na Tecnologia da Informação*. *Data Gramma Zero – Revista de Ciência da Informação*, v. 2, n. 6, dez. 2001. Disponível em: <http://www.dgz.org.br/Atual/Art_02.htm>. Acesso em: 22 jan. 2002.

INTEROPERABILITY. In: INSTITUTE FOR TELECOMMUNICATION SCIENCES. *Glossary of telecommunication terms*. Colorado: [S. n.], 2000. Disponível em: <http://www.its.bldrdoc.gov/fs-1037/dir-019/_2838.htm>. Acesso em: 05 jun. 2004.

W3C. *Architecture of the World Wide Web: W3C Recommendation*, 2004. v. 1. Disponível em: <<http://www.w3c.org/TR/2004/REC-webarch-20041215/>>. Acesso em: 09 jan. 2005.

JONES, D.; BENCH-CAPON, T.; VISSER, P. *Methodologies for ontology development* [S. l.: S. n.], 1998. Disponível em: <<http://cweb.inria.fr/Resources/ONTOLOGIES/methodo-for-dev.pdf>>. Acesso em: 18 jun. 2005.

JOTA, Z. dos S. *Dicionário de lingüística*. Rio de Janeiro: Presença, 1976. (Coleção Linguagem).

LASSILA, O.; SWICK, R. R. *Resource description framework (RDF) model and syntax specification*. [S. l.: S. n.], 1999. Disponível em: <<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>>. Acesso em: 11 maio 2005.

LIBRARY OF CONGRESS ONLINE CATALOG. [S. l.: S. n.], 2005a. Disponível em: <<http://catalog.loc.gov/cgi-bin/Pwebrecon.cgi>>. Acesso em: 10 ago. 2005.

LIBRARY OF CONGRESS ONLINE CATALOG. [S. l.: S. n.], 2005b. Disponível em: <<http://www.loc.gov/standards/marcxml/Sandburg/sandburg.html>>. Acesso em: 10 ago. 2005.

LIBRARY OF CONGRESS. *MARC and XML design considerations*. [S. l.: S. n.], 2002. Disponível em: <<http://www.loc.gov/standards/marcxml/marcxml-design.html>>. Acesso em: 10 abr. 2005.

LIBRARY OF CONGRESS. *The MARC 21 formats: background and principles*. [S. l., S. n.], 1996. Disponível em: <<http://www.loc.gov/marc/96principl.html>>. Acesso em: 10 abr. 2005.

MARCONDES, C. H.; SAYÃO, L. F. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. *Ciência da Informação*, Brasília, v. 30, n. 3, p. 24-33, set./dez. 2001. Disponível em: <<http://www.ibict.br/cionline/300301/3030401.pdf>>. Acesso em: 26 mar. 2002.

MARINO, M. T. *Integração de informações em ambientes científicos na web: uma abordagem baseada na arquitetura RDF*. 2001. 122 f. Dissertação (Mestrado em Informática)–Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001. Disponível em: <http://genesis.ncl.ufrj.br/dataware/Metadados/Teses/Teresa/pagina_tese.htm>. Acesso em: 13 set. 2001.

MARTÍNEZ GONZALES, M. M. Extended Markup Language (XML): uma solución para modelar documentos y sus interrelaciones basada em la semântica de la información y organización del conocimiento. *SCIRE*, [S. l.], v. 6, n. 2, p. 121-151, jul./dic. 2000.

MARTINS JÚNIOR, J. Classificação das páginas na Internet. Dissertação (Mestrado em Ciência da Computação e Matemática Computacional)–Instituto de Ciências e Matemáticas e de Computação, Universidade de São Paulo, São Paulo, 2003.

MEY, E. S. A. *Introdução à catalogação*. Brasília: Briquet de Lemos, 1995.

MILLER, E. *W3C Semantic web activity*. [S. l.: S. n.], 2001. Disponível em: <<http://www.w3.org/2001/12/semweb-fin/w3csw>>. Acesso em: 23 jun. 2005.

- MILLER, P. *Metadata for the masses*. [S. l.: S. n.], 1996. Disponível em: <<http://www.ariadne.ac.uk/issue5/metadata-masses/>>. Acesso em: 16 abril 2004.
- MILSTEAD, J.; FELDMAN, S. *Metadata: cataloging by any other name*. *Online*, [S. l.], january 1999. Disponível em: <<http://www.online.com/online/o11999/milstead1.html>>. Acesso em: 22 jun. 2004.
- MOURA, A. M.^a de C. *A Web Semântica: fundamentos e tecnologias*. [S. l.: S. n.], 2002a. Disponível em: <<http://ipanema.ime.eb.br/~anamoura/public/WebSemantica.zip>>. Acesso em: 14 fev. 2003.
- MOURA, A. M.^a de C. *A Web Semântica: fundamentos, tecnologias e tendências*. [S. l.: S. n.], 2002b. Disponível em: <http://genesis.nce.ufrj.br/dataware/TESI_2002_3/unidades/tutorial_sbbd2002.pdf>. Acesso em: 28 nov. 2004.
- NOVELLO, T. C. *Ontologias: sistemas baseados em conhecimento e modelos de banco de dados*. [S. l.: S. n.], 2002? Disponível em: <http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_taisa.pdf>. Acesso em: 15 maio 2003.
- O'NEILL, E.T.; LAVOIE, B.F.; McCLAIN, P.D. *Web Characterization Project: an analysis of metadata usage on the web*. [S. l.: S. n.], 1998. Disponível em: <http://www.oclc.org/oclc/research/publications/review98/oneill_etal/metadata.htm>. Acesso em: 14 mar. 2001.
- ORTIZ-REPISO JIMÉNEZ, V. Qué enseñamos después del MARC? *Organizacion del Conocimiento en Sistemas de Información y Documentación*, Zaragoza, v. 03, p. 217-225, 1999.
- PALMER, S. B. *The semantic web: an introduction*. [S. l.: S. n.], 2001. Disponível em: <<http://infomesh.net/2001/swintro/>>. Acesso em: 08 jul. 2005.
- PEREIRA, A. M., SANTOS, P. L. V. A da C. O uso estratégico das tecnologias em catalogação. *Cadernos da F.F.C.*, Marília, v. 7, n. 1/2, p. 121- 131, 1998.
- ROBREDO, J. Organização dos documentos ou organização da informação: uma questão de escolha. *Data Gram Zero – Revista de Ciência da Informação*, v. 5, n. 1, fev. 2004. Disponível em: <http://www.dgzero.org/fev.04/Art_05.htm>. Acesso em: 17 abr. 2004.

ROSA, P. A. *Web Semântica*. [S. l.: S. n.], 2002. Disponível em: <http://www.ime.usp.br/~yw/ano2002/mac5701/sem2/rosa_final.ps>. Acesso em: 26 jun. 2005.

ROSETTO, M. *Metadados e formatos de metadados em sistemas de informação: caracterização e definição*. 2003, 95 f. Dissertação (Mestrado em Ciências da Comunicação)–Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2003.

SANTAREM SEGUNDO, J. E. *Recursos tecno-metodológicos para a descrição e recuperação de informações na web*. 2004, 157 f. Dissertação (Mestrado em Ciência da Informação)–Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2004.

SANTARÉM SEGUNDO, J. E.; VIDOTTI, S. A. B. G. Organização da informação na web: a busca na qualidade do armazenamento e da recuperação com a utilização de XML e RDF. In: SIMPÓSIO EM FILOSOFIA E CIÊNCIAS, 5., 2003, Marília. *Anais...* Marília: Unesp Marília Publicações, 2003.

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.

SEMANTIC WEB. *About SemanticWeb.org*. 2002. [S. l.: S. n.]. Disponível em: <<http://www.semanticweb.org/about.html>>. Acesso em: 10 ago. 2005.

SENSO, J. A.; ROSA PIÑERO, A. de la. El concepto de metadato. Algo más que descripción de recursos electrónicos. *Ciência da Informação*, Brasília, v. 32, n. 2, p. 95-106, maio/ago. 2003.

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.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ciência da Informação*, Brasília, v.29, n.1, p.93-102, jan./abr. 2000. Disponível em: <<http://www.ibict.br/cionline/290100/29010010.pdf>>. Acesso em: 06 jul. 2000.

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.

SOUZA, T. B. de, CATARINO, M. E., SANTOS, P. C. dos. Metadados: catalogando dados na Internet. *Transinformação*, Campinas, v. 9, n. 2, maio/ago. 1997. Disponível em: <<http://www.puccamp.br/~biblio/tbsouza92.html>>. Acesso em: 12 jul. 2000.

SOWA, J. F. *Building, sharing, and merging ontologies*. [S. l.: S. n.], 1999. Disponível em: <<http://users.bestweb.net/~sowa/ontology/ontoshar.htm>>. Acesso em: 11 maio 2005.

TAKAHASHI, T. (Org.). *Sociedade da Informação no Brasil: o livro verde*. Brasília: Ministério da Ciência e Tecnologia, 2000.

TELLO, A. L. *Ontologías em la Web Semântica*. In: JORNADAS DE INGENIERÍA WEB'01, [S. l.: S. n.], 2002?. Disponível em: <<http://www.informandote.com/jornadasIngWEB/articulos/jiw02.pdf>>. Acesso em: 12 nov. 2004.

THE LIBRARY OF CONGRESS. *MARCXML: MARC 21 XML Schema official web site*. [S. l.: S. n.]. 2005. Disponível em: <<http://www.loc.gov/standards/marcxml/>>. Acesso em: 10 ago. 2005.

WEIBEL, S. *The evolving metadata architecture for the World Wide Web: bringing together the semantics, structure and syntax of resources description*. [S. l.: S. n.], 2000? Disponível em: <<http://purl.org/net/weibel>>. Acesso em: 09 jan. 2005.

WOODLEY, M. S.; CLEMENT, G.; WINN, P. *DCMI Glossary*. [S. l.: S. n.], 2005. Disponível em: <<http://dublincore.org/documents/usageguide/glossary.shtml>>. Acesso em: 07 jul. 2005.

XML Schema: processamento estruturado de documentos 2001. [S. l.: S. n.], 2001. Disponível em: <<http://www.di.uminho.pt/~jcr/AULAS/micei2002/Slides/XML%20Schema.ppt>>. Acesso em: 06 ago. 2005.

BIBLIOGRAFIA CONSULTADA

BACA, M. (Ed.). *Introducción a los metadatos vías a la información digital*. Traducido al español por Marisol Jacas-Santoll. Los Angeles, CA: J. Paul Getty Trust, 1998. 43 p.

BARRETO, A. de A. Mudança estrutural no fluxo do conhecimento: a comunicação eletrônica. *Ciência da Informação*, Brasília, v. 27, n. 02, p. 122-127, maio/ago. 1998.

BASTOS, F. M.; FUJITA, M. S. L. Representação de assuntos em padrões de metadados. In: SIMPÓSIO DE FOLOSOFIA E CIÊNCIAS, 5., 2003, Marília. *Anais...* Marília: Unesp Marília Publicações, 2003.

BAX, M. P. As bibliotecas na web e vice-versa. *Perspect. Ciênc. Inf.*, Belo Horizonte, v. 3, n. 1, p. 5-20, jan./jun. 1998.

BONIFÁCIO, A. S.; HEUSER, C. A. *Metadados semânticos para buscas em Bibliotecas Digitais*. [S. l.: S, n. , S. d]. Disponível em: <<http://www.uel.br/pessoal/ailton/trabalhos/semanaacad-ailton.html>>. Acesso em: 30 jan. 2003.

CASTELLS, P. *Aplicación de técnicas de la web semântica*. Madrid: Escuela Politécnica Superior Universidad Autónoma de Madrid. Disponível em: <<http://www.ii.uam.es/~castells/publications/coline02.pdf>>. Acesso em: 28 nov. 2004.

CENDÓN, B.V. A Internet. In: CAMPELLO, B. S.; et al. (Org). *Fontes de informação para pesquisadores e profissionais*. Belo Horizonte: UFMG, 2000. p. 275-300.

CÓDIGO de catalogação anglo – americano. São Paulo: FEBAB, 1983.

CROMWELL-KESSLER, W. Correspondencias entre metadatos e interoperabilidade: Qué significa todo esto?. In : INTRODUCCIÓN a los metadatos: vías a la información digital. [S. l.]: GETTY, 1999. p. 21-24.

DELGADO DOMINGUEZ, A. M. *Mecanismos de Recuperación de Información en la WWW*. 1998. 84f. Tese (Doutorado em Informática)–Universitat de les Illes Balears, Palma, 1998. Disponível em: <<http://servidorti.uib.es:8000/adelaida/tice/modul6/memfin.pdf>>. Acesso em: 04 jan. 2004.

DEMPSEY, L.; POWELL, A. *Dublin Core and metadata: a tutorial*. [S. l.: S. n., S. d]. Disponível em: <<http://hosted.ukoln.ac.uk/ec/metadata-1997/tutorial/presentation/>>. Acesso em: 21 mar. 2001.

DÜRSTELER, J. C. *La telaraña semântica*. [S. l.: S. n., S. d]. Disponível em: <http://dialógica.com.ar/unr/postitulo/madialab/archives/cat_web_semantica>. Acesso em: 16 abr. 2004.

EÍTO BRUN, R. *Tema 5: XML en la descripción de recursos*. Sevilla: [S. n.], 2002. Disponível em: <<http://www.forpas.us.es/aula/xml/doc/09.XML%20en%20la%20descripci%C3%B3n%20de%20recursos%20MARC.ppt>>. Acesso em: 22 jun. 2004.

FERREIRA, M. M. (Trad. e Adap.). *MARC 21: formato condensado para dados bibliográficos*. Marília: UNESP Marília Publicações, 2000. v. 1.

FLAMINO, A. N. *MARC21 e XML como ferramentas para a consolidação da catalogação cooperativa automatizada: uma revisão de literatura*. Marília, 2003, 142 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003.

FONSECA, F.; ENGENHOFER, M.; BORGES, K. A. V. *Ontologias e interoperabilidade Semântica entre SIGs*. [S. l.: S. n., S. d.]. Disponível em: <<http://www.geoinfo.info/geoinfo2000/papers/011.pdf>>. Acesso em: 12 nov. 2004.

FREITAS, F. L. G. de. *Ontologias e a Web Semântica*. Santos: Programa de Pós-Graduação em Informática da Universidade Católica de Santos – UniSantos. [S. d.]. Disponível em: <<http://ftp.inf.pucpcaldas.br/CDs/SBC2003/pdf/arq0018.pdf>>. Acesso em: 28 nov. 2004.

GILL, T. Los metadatos y la World Wide Web. In: INTRODUCCIÓN a los metadatos: vías a la información digital. [S. l.]: GETTY, 1999. p. 10-20.

MARCONDES, C. H. Tecnologias da informação e impacto na formação do profissional da informação. *Transinformação*, Campinas, v. 11, n. 3, p. 189-193, set./dez., 1999.

MÉNDEZ RODRÍGUEZ, E. M. *Metadatos y tesauros: aplicación de XML/RDF a los sistemas de organización del conocimiento en Intranets*. [S. l.: S. n.], 2000. Disponível em: <<http://rayuela.uc3m.es/~mendez/publicaciones/fesabid00/fesabid002.pdf>>. Acesso em: 16 mar. 2002.

MÉNDEZ RODRÍGUEZ, E. M. *RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio*. [S. l.: S. n.], 1999. Disponível em: <<http://rayuela.uc3m.es/~mendez/publicaciones/7jc99/rdf.htm>>. Acesso em: 16 mar. 2002.

MÉNDEZ RODRÍGUEZ, E. M.; MERLO VEGA, J. A. *Localización, identificación y descripción de documentos web: tentativas hacia la normalización*. [S. l.: S. n., S. d.]. Disponível em: <<http://rayuela.uc3m.es/~mendez/publicaciones/fesabid00/fesabid001.pdf>>. Acesso em: 16 abr. 2002.

MILLER, E. An introduction to the Resource Description Framework. *D-Lib Magazine*, maio, 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 18 jul. 2004.

MILSTEAD, J.; FELDMAN, S. Metadata projects and standards. *Online*, [S. l.], january, 1999. Disponível em: <<http://www.online.com/online/ol1999/milstead1.html>>. Acesso em: 22 jun. 2004.

NAFRÍA, I. *El futuro de Internet tiene nombre: la Web Semântica*. [S. l.: S. n., S. d.]. Disponível em: <<http://www.baquia.com/com/20010523/art00008.html>>. Acesso em: 18 junho 2004.

OLIVEIRA, R. M. V. B. *Web semântica: novo desafio para os profissionais da Informação*. [S. l.: S. n., S. d.]. Disponível em: <<http://www.sibi.ufrj.br/snbu/snbu2002/oralpdf/124.a.pdf>>. Acesso em: 28 nov. 2004.

OLIVEIRA, C. E. T. de; MARTINS, R. M^a. *Web Semântica: uma visão geral*. [S. l.: S. n., S. d.]. Disponível em: <http://www.eng.uerj.br/~rodane/survey_generico.pdf>. Acesso em: 14 fev. 2003.

RIBEIRO JÚNIOR, D. I. *Agentes inteligentes como mediadores na recuperação de informação*. 2001, 113 f. Dissertação (Mestrado em Ciência da Informação)–Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2001.

ROBREDO, J. *Da ciência da informação revisitada aos sistemas humanos de informação*. Brasília: Thesaurus, 2003.

ROWLEY, J. *A biblioteca eletrônica*. 2. ed. Brasília: Briquet de Lemos/Livros, 2002.

SAMTAMARIA GONZÁLES, F. *XML (Extensible Markup Language):* nuevo estándar para la descripción de documentos em la Word Wide Web. In: FESABID 98 – JORNADAS ESPAÑOLAS DE DOCUMENTACIÓN, 5., 1998. Disponível em: <http://fesabid98.florida-uni.es/comunicaciones/f_santamaria/f_santamaria.htm>. Acesso em: 13 abr. 2004.

SAN SEGUNDO MANUEL, R. *Organización del conicimiento em Internet:* metadatos bibliotecários Dublin Core. In: FESABID 98 – JORNADAS ESPAÑOLAS DE DOCUMENTACIÓN, 5., 1998. Disponível em: <http://fesabid98.florida-uni.es/comunicaciones/r_sansegundo.htm>. Acesso em: 16 abr. 2004.

SANTARÉM SEGUNDO, J. E. et al. Linguagem XML como base na busca da interoperabilidade e organização da informação. In: SIMPÓSIO EM FILOSOFIA E CIÊNCIAS, 5., 2003, Marília. *Anais...* Marília: Unesp Marília Publicações, 2003.

SILVA MUÑOZ, L. *Representacion de ontologias em la web semântica.* Rio Grande do Sul: Instituto de Informática–Universidade Federal do Rio Grande do Sul (UFRGS). [S. d]. Disponível em: <http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_lydia.pdf>. Acesso em: 28 nov. 2004.

SILVA, L. A. E. da. *Geração dinâmica de interfaces de bibliotecas digitais baseada em metadados.* 2000, 130 f. Dissertação (Mestrado em Sistemas e Computação) - Instituto Militar de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2000. Disponível em: <<http://ipanema.ime.eb.br/~de9/teses/2000/>>. Arquivo: luis.zip . Acesso em: 13 set. 2001.

TARAPANOFF, K. O profissional da informação e a sociedade do conhecimento: desafios e oportunidades. *Transinformação*, Campinas, v. 11, n. 1, p. 27-38, jan./abr. 1999.

WAYNE, J.; AHRONHEIM, J. R.; CRAWFORD, J. *Cataloging the web: metadata, AACR, and MARC21.* [S. l.]: The Scarecrow Press, 2002.