

Cíntia Braga Ferreira Pinheiro

**A construção do conhecimento científico: a
Web Semântica como objeto de estudo**

**Marília
2008**

Cíntia Braga Ferreira Pinheiro

**A construção do conhecimento científico: a
Web Semântica como objeto de estudo**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista, UNESP – Campus de Marília, como requisito parcial para obtenção do título de Mestre.

Área de concentração: Informação, Tecnologia e Conhecimento.

Linha de pesquisa: Informação e Tecnologia

Orientador: Prof. Dr. Edberto Farneda

**Marília
2008**

FICHA CATALOGRÁFICA

P655c Pinheiro, Cintia Braga Ferreira
A construção do conhecimento científico: a web semântica como objeto de estudo / Cintia Braga Ferreira Pinheiro. – Marília, 2008.
63 f.; 30cm

Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2008.

Orientador: Prof. Dr.Edberto Ferneda.

Bibliografia: f. 58-63

1.Produção Científica - Brasil. 2. Web Semântica. 3. Ciência da Informação. 4. Ciência da Computação. I Autor. II. Título

Cíntia Braga Ferreira Pinheiro

**A construção do conhecimento científico:
a Web Semântica como objeto de estudo**

BANCA EXAMINADORA:

Dr. Edberto Ferneda
Programa de Pós-Graduação em Ciência da Informação
Universidade Estadual Paulista – UNESP – Campus de Marília

Dr. Raimundo Nonato dos Santos
Programa de Pós-Graduação em Ciência da Computação
Centro de Ciências da Educação - UFSC

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

Marília, 25 de fevereiro de 2008.

À pessoa mais importante da
minha vida, minha Mãe. Sem
sua coragem Eu não existiria.

Agradecimentos

Primeiramente ao amigo e colega Edberto, pelo incentivo e apoio nas tentativas de conciliar família, estudos e trabalho. Serei sempre grata pelo apoio em meus momentos de completo desespero e pelas terapias disfarçadas de caronas para Marília.

Ao meu orientador Prof. Dr. Edberto Ferneda pelo auxílio e empenho na concretização deste projeto.

Especialmente à professora Silvana Vidotti pelo apoio e pela acolhida tanto no Programa de Pós-Graduação quanto em seu cotidiano familiar. Profissional dedicada, um exemplo no gerenciamento da vida pessoal e profissional. Agradeço seu empenho, paciência e toda colaboração em minha formação.

Aos professores da linha Informação e Tecnologia representados aqui pelas professoras Silvana Vidotti e Plácida L. V. A Santos, pela organização das disciplinas que permitiram a conclusão dos créditos, pelas caronas para a rodoviária, pelos almoços, pelas conversas e opiniões proferidas, pelo carinho demonstrado e com qual me senti acolhida.

Ao professor Raimundo Nonato dos Santos pelas contribuições e apontamentos feitos no Exame de Qualificação.

Ao Departamento de Física e Matemática da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto que autorizou as ausências necessárias para a realização deste projeto de vida.

Aos professores do Programa de Pós-Graduação em Ciência da Informação.

Aos colegas do Programa de Pós-Graduação em Ciência da Informação que compartilharam todas as minhas insanidades.

PINHEIRO, Cíntia Braga Ferreira. *A construção do conhecimento científico: a web semântica como objeto de estudo*. 2008. f. Dissertação (Mestrado em Ciência da Informação). Faculdade de Filosofia e Ciências – Universidade Estadual Paulista, Marília, 2008.

RESUMO

A Sociologia da Ciência ou os Estudos Sociais da Ciência têm procurado analisar a estrutura das disciplinas científicas em relação às suas práticas sociais, procurando oferecer novas perspectivas sobre a construção do saber e o desenvolvimento científico e tecnológico. Um conceito importante para o desenvolvimento deste trabalho é o de comunidade científica caracterizada pela prática de uma especialidade, por uma formação teórica comum, pela circulação abundante de informação no interior do grupo. Este estudo tem por objetivo caracterizar a pesquisa brasileira em Web Semântica, considerado um tema na fronteira de pesquisa de duas matrizes disciplinares: a Ciência da Computação e a Ciência da Informação. Optou-se por um estudo cienciométrico, tendo como fonte para coleta dos dados o Curriculum Lattes de pesquisadores doutores e nas dissertações e teses defendidas sobre Web Semântica e suas tecnologias. A integração entre comunidades científicas formadas em matrizes disciplinares diferentes encontra um ambiente favorável nos espaços de fronteira que demarcam as matrizes disciplinares. Comprova-se isso ao observar-se a comunidade pesquisadora sobre Web Semântica no Brasil, composta por pesquisadores de diversas matrizes como a Lingüística, a Ciência da Informação e a Ciência da Computação.

Palavras-chave: Web Semântica. Cienciométrica. Comunidade Científica

PINHEIRO, Cíntia Braga Ferreira. *A construção do conhecimento científico: a web semântica como objeto de estudo*. 2008. f. Dissertação (Mestrado em Ciência da Informação). Faculdade de Filosofia e Ciências – Universidade Estadual Paulista, Marília, 2008.

ABSTRACT

Science is a critical and dynamic social activity whose objective is the knowledge production on different aspects of the nature. An important concept for the development of this work is of scholarly community characterized by the practical one of a specialty, for a common theoretical formation, the abundant circulation of information in the interior of the group. This study it has for objective to characterize the Brazilian research in Semantic Web, considered a subject in the border of research of two matrices discipline matrices the Computer Science and the Information Science. It was opted to a scientometrics study, having as source of the data the Curriculum Lattes of doctors. The integration between scholarly communities formed in matrices of different disciplines finds a favorable environment in the scientific boundaries that demarcate the discipline matrices Observing that researching community on Semantic Web in Brazil, composed for researchers of diverse matrices as Linguistics, Information Science and Computer Science.

Keywords: Semantic Web. Scientometrics. Scientific Community.

Sumário

AGRADECIMENTOS	6
1 INTRODUÇÃO	10
1.1 O ESTUDO DA CIÊNCIA: COLABORAÇÃO CIENTÍFICA E CIENCIOMETRIA	11
1.2 CIÊNCIA DA INFORMAÇÃO, CIÊNCIA DA COMPUTAÇÃO E A WEB SEMÂNTICA	12
1.3 OBJETIVOS DA PESQUISA.....	14
1.4 METODOLOGIA	14
1.5 ORGANIZAÇÃO DA DISSERTAÇÃO.....	17
2 A CIÊNCIA DA INFORMAÇÃO E A CIÊNCIA DA COMPUTAÇÃO COMO MATRIZES DISCIPLINARES.....	18
2.1 CIÊNCIA DA COMPUTAÇÃO	22
2.2 CIÊNCIA DA INFORMAÇÃO	27
2.3 CIÊNCIA DA INFORMAÇÃO, CIÊNCIA DA COMPUTAÇÃO E O COMPARTILHAMENTO DE OBJETOS DE ESTUDO.....	30
3 DESENVOLVIMENTO CIENTÍFICO DA WEB.....	34
3.1 RECUPERAÇÃO DA INFORMAÇÃO NA WEB	35
3.2 WEB SEMÂNTICA	37
3.2.1 <i>Infra-estrutura da Web Semântica</i>	38
3.2.2 <i>Ontologias</i>	41
4 UM RECORTE DA PESQUISA BRASILEIRA EM WEB SEMÂNTICA	44
4.1 O CONHECIMENTO CIENTÍFICO E SUA COMUNICAÇÃO	45
4.2 A WEB SEMÂNTICA NA PLATAFORMA LATTES	46
4.3 QUEM CONTRIBUI PARA A PESQUISA EM WEB SEMÂNTICA NO BRASIL?	48
4.4 QUEM INFLUENCIA A PESQUISA EM WEB SEMÂNTICA NO BRASIL?	51
5 CONSIDERAÇÕES FINAIS.....	55
REFERÊNCIAS BIBLIOGRÁFICAS.....	58

1 INTRODUÇÃO

A ciência de uma maneira geral é considerada como uma atividade social crítica e dinâmica cujo objetivo é a produção de conhecimento sobre diferentes aspectos da natureza. Trata-se de um empreendimento que envolve colaboração e competição entre seus atores. Geralmente é composta por: um núcleo de conhecimento, que constitui uma disciplina específica, um método próprio de investigação e um campo experimental para verificação das proposições. Essa atividade social possui objetivos diversos, tais como: a descrição, o controle, a predição e a explicação dos aspectos naturais e sociais formadores da natureza. Os métodos científicos variam conforme o objeto de estudo e a própria concepção dos pesquisadores sobre o fenômeno ou objeto de estudo.

As relações sociais e a estruturação das disciplinas científicas em relação às suas práticas metodológicas e o desenvolvimento científico são objetos de estudo da Sociologia da Ciência.

Whitley (1974 apud KOBASHI; SANTOS, 2006) propõe que Ciência seja analisada sociologicamente como a institucionalização tanto cognitiva quanto social de uma disciplina científica ou matriz disciplinar. O ponto de vista de institucionalização cognitiva envolve os aspectos epistemológicos, teóricos e metodológicos das matrizes disciplinares. São os consensos alcançados em relação aos conceitos da matriz, a legitimidade dos problemas e suas soluções, métodos e técnicas adotados pela matriz disciplinar para a observação dos fenômenos.

A institucionalização social de uma matriz envolve as estruturas formais que sustentam uma comunidade científica. É a análise do seu grau de organização: a capacidade de formar de novos pesquisadores, a constituição de canais de comunicação, a estruturação de instituições legitimadoras como Institutos de Pesquisa, Sociedades Científicas, etc e as condições de acesso aos programas de fomento.

Independente das formas como a analisar a institucionalização de uma matriz disciplinar, as interações políticas e sociais dentro da comunidade científica desempenham um importante papel para sua consolidação. Analisando a profissionalização e a institucionalização da ciência, Brown (1993) afirma que conforme seus conhecimentos e sua organização crescem, os cientistas se tornam capazes de restringir o acesso aos privilégios aos novos pesquisadores. Demonstrando assim a capacidade de associação e articulação que estes

profissionais adquirem a partir de suas relações sociais. Este aspecto evidencia a importância da comunidade científica para a institucionalização de uma matriz disciplinar.

Para Brown (1993) a institucionalização de uma nova disciplina ou matriz disciplinar é um processo retórico, cognitivo e político no qual as alianças profissionais são formalizadas estabelecendo influências na constituição desta nova disciplina.

Kuhn (2003) também destaca a importância da comunidade científica na socialização dos paradigmas de uma disciplina objetivando a adesão as crenças e valores compartilhados dentro de determinada matriz disciplinar. Portanto todo fazer científico envolve interações sociais, defesas de interesses políticos e cognitivos corroborando a afirmação que toda ciência tem uma estrutura que estimula a colaboração e a competição entre pesquisadores.

1.1 O estudo da Ciência: colaboração científica e Cienciometria

Cada vez mais as tecnologias de informação e comunicação têm estimulado a cooperação das ciências, derrubando fronteiras paradigmáticas, institucionais e continentais. “As atividades informatizadas parecem destinadas a melhorar a produtividade nos anos que virão, permitindo assim que continue a crescer a quantidade de informações científicas em circulação”. (MEADOWS, 1999, p. 246) O conhecimento gerado é registrado em periódicos cujo acesso está geralmente disponível eletronicamente em todo o mundo, os pesquisadores se encontram em reuniões internacionais, estabelecem seus contatos através do correio eletrônico, listas de discussão, viajam para trabalhar em laboratórios de outros países e frequentemente envolvem colegas estrangeiros em seus projetos tendo como suporte as Tecnologias de Informação e Comunicação.

Uma das formas de se estudar as colaborações entre pesquisadores é através dos estudos métricos da informação, dentre eles a Cienciometria ou Cientometria conforme a opção terminológica dos pesquisadores da área. Neste trabalho optou-se pelo uso do termo Cienciometria e estudos cienciométricos para denominar os estudos quantitativos da ciência como disciplina ou atividade econômica. É de interesse da Cienciometria o crescimento quantitativo das ciências, o desenvolvimento das disciplinas ou sub-disciplinas, a obsolescência de paradigmas científicos, etc... (MACIAS-CHAPULA, 1998)

Lara (2006) define a Cienciometria como uma área que objetiva estudos dos aspectos quantificáveis da ciência como atividade humana e do conhecimento científico, seja através do uso de instrumentos para mensuração da produção e atividade científica de um país ou de

uma comunidade científica específica. Os estudos cienciométricos permitem a análise do aspecto social e colaborativo das ciências através das relações sociais e paradigmáticas explicitadas nas publicações científicas. Merton (1979) destaca como uma norma fundamental da ciência a divulgação dos resultados de uma pesquisa aos pares, pois permite a continuidade da ciência. “A literatura de um assunto científico é tão importante para ele quanto a própria pesquisa, pois esta não estaria completa se seus resultados não fossem divulgados” (MUELLER, 1995, p. 64) Sendo imperativo da ciência a prática da comunicação científica, as publicações geradas pela prática da pesquisa são reflexo das interações entre pesquisadores e instrumentalizam o estudo da ciência.

Dentre os diversos estudos cienciométricos destaca-se a análise de co-autoria que permite identificar a colaboração entre pesquisadores, grupos de pesquisa, matrizes disciplinares, países, instituições de pesquisa, etc. A co-autoria analisa as parcerias presentes na ciência moderna. Para Lima, Velho, Faria (2007) a co-autoria estabelece indicadores de ligação que permitem a observação da dinâmica da produção científica. Também se analisa as ligações entre documentos, autores, e matrizes disciplinares através das citações presentes nas publicações científicas, esse estudo quantitativo permite mapear o desenvolvimento científico em determinado frente de pesquisa, por exemplo, as influências paradigmáticas que constroem o conhecimento científico e tecnológico de uma comunidade.

1.2 Ciência da Informação, Ciência da Computação e a Web Semântica

Os avanços tecnológicos na área da informação têm despertado os interesses de diversas áreas do conhecimento na análise de suas aplicações e impactos no mundo pós-moderno. Dentre as disciplinas que têm como objeto de estudo a informação podemos destacar a Ciência da Informação e a Ciência da Computação, cada uma delas influenciadas por seus paradigmas. (PINHEIRO; VIDOTTI, 2005, RAMALHO; VIDOTTI; FUJITA, 2007) Ferneda (2003, p. 1) destaca que “a informação, objeto de comum interesse de ambas as ciências, é paradoxalmente o que mais as distancia”.

Para a Ciência da Informação, “a informação comporta um elemento de sentido. É um significado transmitido a um ser consciente por meio de uma mensagem inscrita em um suporte espaço-temporal: impresso, sinal elétrico, onda sonora, etc.” (LE COADIC, 1996 p.5)

Para a Ciência da Computação o conceito de informação está relacionada à Teoria Matemática da Informação que envolve a representação matemática para “codificação da

informação e sua medida, entropia de código, transmissão da informação e modelagem do sistema de transmissão, maximização do fluxo de informação por canal, processamento digital de sinais” (RAMALHO; VIDOTTI; FUJITA, 2007, p. 7)

Ramalho; Vidotti; Fujita, (2007, p. 8) destacam que

Utilizando como exemplo um processo de recuperação de informação no ambiente Web, pode-se verificar que no âmbito da área da Ciência da Informação é necessário levar em consideração os componentes semânticos inerentes a tal processo, no entanto, de acordo com o enfoque da área da Ciência da Computação, observa-se que os tradicionais ‘motores de busca’, baseiam-se exclusivamente na recuperação de dados, não levando em consideração as semânticas contidas nas páginas da Web, recuperando apenas seqüências de caracteres que satisfaçam determinadas condições de busca

Alves; Santos (2005, p.1) destacam que

As tecnologias de informação e comunicação estão cada vez mais presente em nosso cotidiano, aliada a essa característica temos também a crescente valorização da informação e o crescimento exponencial dos recursos informacionais disponibilizados em diversos ambientes, principalmente na Web. Esse novo cenário, caracterizado pela disponibilização de grandes quantidades de informação, também apresenta problemas, que estão relacionados principalmente com a busca e recuperação de recursos informacionais digitais na Web.

A crise provocada pelo aumento exponencial dos documentos Web, prejudicando a recuperação da informação relevante, estimulou o desenvolvimento de um projeto denominado Web Semântica que tem por objetivo desenvolver tecnologias capazes de agregar valor semântico aos dados disponíveis na Web possibilitando que os sistemas computacionais possam naturalmente compreendê-los de forma direta ou indireta, facilitando assim a interação Homem-computador.

Este trabalho se enquadra na linha de pesquisa Informação e Tecnologia do Programa de Pós-Graduação em Ciência da Informação da Unesp de Marília, ao analisar e refletir sobre a dinâmica das comunidades científicas brasileiras da Ciência da Informação e da Ciência da Computação na construção e desenvolvimento de uma nova tecnologia que pretende aprimorar a estrutura da Web agregando valor semântico a ela, e em conseqüentemente aprimorando as formas de Recuperação da Informação na Web.

Qual o papel destas Ciências dentro do desenvolvimento científico e tecnológico que os ambientes Web demandam? Existe a percepção pela comunidade científica da Ciência da Computação dos paradigmas da Ciência da Informação sobre organização da informação

como instrumentos que podem auxiliar o desenvolvimento da Web Semântica? Existe interação entre pesquisadores brasileiros da Ciência da Informação e da Ciência da Computação nas publicações científicas sobre Web Semântica e suas tecnologias? Existe espaço para grupos de pesquisa inter ou transdisciplinares ¹em Web Semântica? Todos estes questionamentos motivaram a realização deste trabalho.

1.3 Objetivos da pesquisa

Este estudo tem por objetivo geral refletir sobre a participação e integração da Ciência da Informação e da Ciência da Computação na construção e desenvolvimento do conhecimento científico sobre Web Semântica, para alcançá-lo pretende:

- Identificar a produção científica em Web Semântica disponível nas principais iniciativas brasileiras de bancos e bases de dados científicos em relação as matrizes disciplinares, canais de comunicação, colaboração científica;
- Identificar os autores mais citados nas dissertações e teses sobre o tema a fim de mapear as influências recebidas no seu desenvolvimento científico;
- Identificar a presença da Ciência da Informação nos canais de comunicação da Ciência da Computação e vice-versa;
- Identificar a colaboração científica da Ciência da Informação e da Computação através de co-autorias.

1.4 Metodologia

O trabalho fundamenta-se na perspectiva cienciométrica de análise das publicações científicas produzidas pelos pesquisadores brasileiros e no conceito de que as publicações

^{1 1} Interdisciplinaridade para Pombo (1993) indica além da pluralidade e uma justaposição a coesão entre saberes, trata-se da complementariedade/interação entre disciplinas; é o intercâmbio mútuo e integração recíproca entre várias ciências, é o uso de conhecimentos de várias disciplinas para resolver um problema concreto ou compreender um determinado fenômeno sob diferentes pontos de vista. A transdisciplinaridade é a integração global das várias ciências numa etapa posterior e superior à interdisciplinaridade, situando-a as relações no interior de um sistema total, sem fronteiras estáveis entre as disciplinas; visando a articulação de diferentes áreas do conhecimento tendo em vista a construção de um meta-conhecimento.

científicas são as expressões de pessoas ou grupo trabalhando em uma frente de pesquisa, portanto é possível dizer alguma coisa sobre as relações entre os pesquisadores a partir destas publicações. (MACIAS-CHAPULA, 1998)

A Cienciometria é um segmento da Sociologia da Ciência que tem como objetos de estudo as disciplinas, assuntos, áreas e campos científicos, analisando os fatores que diferenciam as sub-disciplinas (revistas, autores, documentos, como os cientistas se comunicam) utilizando métodos de análise de conjuntos e de correspondência, objetivando identificar os domínios de interesses, concentração de assuntos e compreender a comunicação entre os cientistas (MACIAS-CHAPULA, 1998). É o estudo dos aspectos quantitativos da ciência como disciplina ou atividade econômica.

Para a Cienciometria, a dinâmica da atividade científica pode ser observada através das publicações científicas, pois “se o documento é a expressão de uma pessoa ou grupo trabalhando em uma frente de pesquisa, pode-se dizer alguma coisa sobre as relações entre as pessoas a partir dos próprios documentos”. (MACIAS-CHAPULA, 1998, p. 134).

Como destaca Santos (2003, p.134) são premissas da Cienciometria alguns postulados:

uma obra científica é o produto objetivo da atividade intelectual criativa. Num contexto científico, uma publicação é uma representação da atividade de pesquisa de seu autor. O maior esforço deste autor é de persuadir os pares de que suas descobertas, seus métodos e técnicas são particularmente pertinentes. O modo de comunicação escrita fornecerá, portanto, todos os elementos técnicos, conceituais, sociais e econômicos que o autor busca afirmar ao longo de sua argumentação. A atividade de publicação científica é uma eterna confrontação entre as reflexões intrínsecas do autor e os conhecimentos que ele adquiriu pela leitura dos trabalhos originários dos outros autores.

Dentre os diversos estudos cienciométricos podemos destacar a análise de co-autoria e os estudos de citação. Os estudos de citação têm-se constituído instrumento para avaliação de diversos aspectos do fazer científico como: estabelecer o fator de impacto de publicações periódicas, verificar a produção e visibilidade de autores, identificar o impacto dos autores através da quantificação de citações, determinar e/ou identificar frentes de pesquisa, avaliar instituições de pesquisa, etc. Trata-se de análises quantitativas que permitem identificar a disseminação da produção científica (LÓPEZ YEPES, 2003). Geram indicadores que Lima, Velho, Faria (2007, p. 154) caracterizam como de atividade, pois são “criados a partir da contagem de publicações e visam à elaboração de listas de frequência de produção ou *ranking* de grupos de pesquisa, instituições, empresas e países”.

Através do uso de indicadores gerados nos estudos quantitativos de citações “é possível saber como se dá a comunicação científica de uma área do conhecimento, obtendo-se assim um mapeamento da mesma” (VANZ; CAREGNATO, 2003, p. 251).

Os estudos envolvendo a autoria das publicações científicas procuram identificar as redes de cooperação entre pesquisadores, instituições, matrizes disciplinares, etc gerando indicadores de ligação que permitem a identificação de parceiras entre autores, instituições e países, por exemplo. Lima, Velho, Faria (2007, p. 154) consideram estes indicadores “importantes na compreensão da dinâmica de produção do conhecimento científico, pois identificam e quantificam as relações entre os diversos campos”.

A colaboração entre matrizes disciplinares tem sido mais comum nos temas de pesquisa que exploram a sinergia de múltiplas disciplinas e domínios de conhecimento (HAYTHORNTHWAITE, 2006). A seguir apresenta-se a metodologia e os resultados de um estudo de co-autoria envolvendo doutores brasileiros, que publicam sobre Web Semântica e suas tecnologias, a fim de identificar a existência desta sinergia nas pesquisas brasileiras sobre o tema.

Os dados analisados foram coletados dos seguintes banco e bases de dados: Plataforma Lattes desenvolvida pelo Conselho Nacional de Pesquisa (CNPq) tendo como fonte Currículo Lattes dos pesquisadores doutores. Esta base de dados registra as atividades profissionais de pesquisadores nacionais e estrangeiros vinculados a Instituições de pesquisa. Pretende-se identificar nesta fonte as Instituições de pesquisa, matrizes disciplinares, e a cooperação científica na comunidade pesquisadora da Web Semântica através de um recorte nesta comunidade: os cientistas doutores. Estabeleceu-se uma estratégia de busca nos Currículos Lattes dos pesquisadores doutores através do formulário de Busca Avançada por Assunto com o termo Web Semântica visando a localização dos doutores que tivessem trabalhos publicados sobre o tema.

O Banco de Teses da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a Biblioteca Digital de Teses e Dissertações do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) também serviram como fonte para a coleta de dados. Tendo como meta identificar as influências científicas que a pesquisa na pós-graduação brasileira em Web Semântica tem recebido.

Estas duas iniciativas de Banco de Teses e Dissertações abrangem senão a totalidade das Instituições de Ensino Superior, as mais significativas sob o ponto de vista da avaliação institucional dos seus programas de pós-graduação.

A iniciativa da CAPES é o Banco de Teses que “recentemente registrava 285 mil trabalhos defendidos no período 1987-2004, foi atualizado com a inclusão de 81.341 teses e dissertações publicadas em 2005 e 2006, representando um acréscimo de 28,5 % na base de dados”. (COORDENAÇÃO) Este Banco inclui a coleção de textos completos das dissertações e teses mantidas pelas Instituições de Ensino Superior (IES) do Brasil.

A Biblioteca Digital de Teses e Dissertações é iniciativa do IBICT “que busca integrar os sistemas de informação de teses e dissertações existentes nas IES brasileiras, bem como estimular o registro e a publicação de teses e dissertações em meio eletrônico.” (INSTITUTO)

A estratégia de busca adotada nos dois bancos de teses e dissertações foi a mesma tendo como palavras-chaves os termos Web Semântica e suas tecnologias: Ontologia e/ou Ontologias, RDF.

1.5 Organização da dissertação

No capítulo 2 procura-se apresentar a fundamentação teórica da pesquisa e caracterizar a Ciência da Informação e a Ciência da Computação como matrizes disciplinares, destacando principalmente a teoria relativista de Kuhn.

O capítulo 3 trata do desenvolvimento científico da Web, as questões envolvendo a recuperação da informação em ambientes Web e a apresentação da frente de pesquisa comum das duas matrizes: o projeto Web Semântica e suas tecnologias. Nele pretende-se salientar os aspectos considerados de interesse para as duas matrizes.

E finalmente no capítulo 4 procura-se caracterizar a pesquisa brasileira em Web Semântica com o propósito de identificar a participação das diversas matrizes disciplinares, principalmente a Ciência da Informação e a Ciência da Computação na construção e desenvolvimento desta tecnologia.

2 A CIÊNCIA DA INFORMAÇÃO E A CIÊNCIA DA COMPUTAÇÃO COMO MATRIZES DISCIPLINARES

A Sociologia da Ciência ou os Estudos Sociais da Ciência têm procurado analisar a estrutura das disciplinas científicas em relação às suas práticas sociais, procurando oferecer novas perspectivas sobre a construção do saber e o desenvolvimento científico e tecnológico. Thomas Kuhn, com formação acadêmica em física e matemática, tornou-se um dos autores de maior repercussão dentro da Sociologia da Ciência ao publicar livro *A Estrutura das Revoluções Científicas*, em 1962, no qual propunha uma nova problematização da ciência. A obra de Kuhn provocou um impacto na Sociologia da Ciência ao privilegiar os aspectos históricos e sociológicos na análise da prática científica.

Kuhn (2003) ao identificar, descrever e analisar a dinâmica interna da ciência observou ciclos de ciência normal e revoluções, caracterizando o desenvolvimento científico como transições sucessivas de paradigmas. A ciência normal é uma atividade, de caráter cumulativo, que objetiva a solução de problemas e consiste numa “pesquisa firmemente baseada em uma ou mais realizações científicas passadas (...) reconhecidas durante algum tempo por alguma comunidade científica específica como proporcionando fundamentos para sua prática.” (KUHN, 2003, p. 29). Ela procura expandir o conhecimento de fatos identificados como importantes pelo paradigma. Este, por sua vez, é um conjunto de leis, teorias, aplicações e métodos experimentais aceitos amplamente por uma comunidade científica cuja existência condiciona o aparecimento e o desenvolvimento da atividade científica normal. Cientistas cujas pesquisas fundamentam-se num mesmo paradigma obedecem às mesmas regras e às mesmas normas da prática científica; o consenso aparente num mesmo paradigma é pré-requisito da ciência normal, ou seja, o surgimento e continuação de uma tradição de pesquisa.

Na noção kuhniana a ciência normal não tem como objetivo descobrir novos fatos nem inventar novas teorias, trata-se da resolução de enigmas (*puzzles*). Os enigmas devem satisfazer as seguintes condições: ter uma solução e respeitar determinadas regras quanto aos métodos para solucioná-los. As revoluções científicas ocorrem a partir de crises, quando um paradigma deixar de explicar adequadamente os fenômenos, e os enigmas resistem às soluções disponíveis no paradigma vigente. A crise é a condição fundamental para a emergência de uma nova teoria. No entanto uma nova teoria não surge facilmente, pois os cientistas defendem profundamente seus paradigmas negando as anomalias existentes e

nenhum paradigma é abandonado enquanto não houver outro que o substitua com sucesso. Para Kuhn (2003) o abandono de um paradigma é simultâneo à adoção de outro, pois “rejeitar um paradigma sem simultaneamente substituí-lo por outro é rejeitar a própria ciência” (p. 109). O surgimento de anomalias dentro do paradigma vigente gera sucessivas crises que conduzirão ao surgimento de novos paradigmas, reorganizando um domínio científico específico, num processo radicalmente novo sem vínculo com o paradigma anterior, portanto perdendo o caráter acumulativo da ciência normal. Num processo revolucionário substituem-se paradigmas e surge um novo período de normalidade para a ciência.

Como afirmam Machado e Teixeira (2007), Kuhn estabelece uma postura relativista no estudo da ciência, o pesquisador passa a ser considerado um negociador que produz diversas formas de interpretação do real que precisam ser legitimadas pelo maior número de pessoas, consolidando-se assim o paradigma defendido. A postura relativista rejeita a idéia de critérios únicos, estáveis e atemporais da ciência. “O acordo é puramente social, ele é fruto das interações e das negociações entre os pesquisadores que compartilham de um sistema de crença que, em sua época, lhes parece objetivo, ao passo que esse mesmo sistema foi objeto de negociação.” (MACHADO; TEIXEIRA, 2007, p. 5)

Brown (1993) também defende o forte caráter social e político presente no processo de institucionalização das Ciências, do século XIX até a atualidade. Para o autor as negociações dentro da ciência são políticas com a necessidade de formalizar alianças estratégicas, derrotar os opositores, recrutar novos cientistas, formular programas de conduta para defesa de seus interesses. Richard Whitley considera que a institucionalização da ciência envolve processos cognitivos relacionados à epistemologia, à formulação de paradigmas, ao estabelecimento de métodos e práticas de pesquisa e um processo social relacionado à estrutura organizacional da matriz disciplinar. (KOBASHI; SANTOS, 2006) Destaca-se que mesmo com a separação destes aspectos em categorias distintas, ambas as institucionalizações ocorrem permeadas pelos interesses políticos e sociais presentes na comunidade científica.

Sucintamente foi apresentada a teoria kuhniana para o processo de desenvolvimento da ciência, no entanto os conceitos de paradigma presentes em “A estrutura das revoluções científicas” foram objetos de críticas de diversos historiadores, filósofos e sociólogos da ciência. Masterman (1974) identifica mais de vinte diferentes acepções para o termo paradigma, tais como mito, filosofia, manual, tradição, realização científica, analogia, dispositivo geralmente aceito, fonte de instrumentos, ilustração normal, expediente, fábrica de ferramentas, conjunto de instituições políticas, modelo, princípio organizador, ponto de vista

epistemológico, novo modo de ver, algo que define uma ampla extensão da realidade. A autora agrupa estes diversos significados em três grupos distintos: a) paradigma metafísico ou metaparadigma no qual o termo possui um significado de conjunto de crenças e de mitos, novas formas de ver as coisas; b) paradigma sociológico no qual o termo assume um significado de um resultado científico reconhecido universalmente; c) paradigma de construção no qual o termo é utilizado como fornecedor de instrumentos, uma obra, um trabalho clássico.

Kuhn reconhece a pertinência desta e de outras críticas recebidas ao acrescentar um 'posfácio' à obra no qual atribui dois sentidos ao conceito de paradigma: um sentido sociológico de crenças comuns e outro de soluções concretas dos enigmas. No sentido sociológico um paradigma indica uma constelação de crenças, valores e técnicas compartilhadas pelos membros de uma comunidade científica determinada. O autor considera o sentido sociológico muito mais amplo do que a noção de teoria e propõe que o termo seja substituído pela noção de 'matriz disciplinar', pois implica num posicionamento comum por parte de pesquisadores de uma determinada disciplina, acentuando o caráter grupal e socializador da atividade científica. Trata-se de um 'estilo de pensamento' evidenciando as normas coletivas de funcionamento as quais orientam a direção da pesquisa numa tradição específica de fazer ciência e dá visibilidade ao conjunto de soluções de problemas que os membros de uma comunidade científica partilham. Estas comunidades compartilham não só um mesmo estilo de pensamento como também recorrem aos mesmos canais de comunicação para sua disseminação.

Outro conceito importante na teoria kuhniana e na sociologia da ciência para o desenvolvimento deste trabalho é o de comunidade científica, pois a socialização entre pesquisadores permite uma adesão aos valores e as crenças do grupo. Uma comunidade científica caracteriza-se pela prática de uma especialidade científica, por uma formação teórica comum, pela circulação abundante de informação no interior do grupo e pela unanimidade de juízo em assuntos profissionais.

Dencker (2001) concorda que "Kuhn identificou, na ciência, uma estrutura comunitária em que os participantes são submetidos a uma iniciação profissional e educação similares, o que leva a compartilhar dos mesmos paradigmas."

Uma matriz disciplinar forma seus pesquisadores principalmente através do ensino superior, pois nele se consolida e dissemina paradigmas através de generalizações simbólicas, fortalecimento de crenças em determinados modelos e compartilhamento de valores. Estes

elementos fornecem a uma determinada comunidade científica analogias ou metáforas preferidas, embora possam ocorrer divergências quanto ao seu uso. A estrutura da comunidade científica se estabelece através do compartilhamento de exemplos e de possíveis soluções. Para Kuhn (2003) o pesquisador é formado através da prática e convivência dentro de determinada matriz disciplinar, consolidando um conhecimento tácito gerado através do ‘aprender fazendo’ ao identificar problemas semelhantes e aplicar as soluções aceitas por aquela comunidade específica. Portanto o saber científico é adquirido através desta formação recebida dentro de uma comunidade científica no âmbito de sistemas de convenções que têm origem nos processos cognitivos que incluem os juízos e acordos socialmente aceitos pelo grupo.

Ludwick Fleck afirma que “os pesquisadores científicos pertencem a coletivos de pensamento distintos e são socializados em estilos de pensamento incomensuráveis.” (LOWY, 1994, p. 11) O autor foi um médico que desenvolveu na primeira metade do século passado uma abordagem original para o estudo das ciências e influenciou as teorias consolidadas por Kuhn. Para ele os objetos científicos são construídos por uma comunidade científica ou ‘coletivos de pensamento’ e cada coletivo elabora seu próprio ‘estilo de pensamento’ ou matriz disciplinar. Portanto seguindo esta concepção o projeto da Web Semântica é de interesse de diversas comunidades científicas que se estruturam em estilos de pensamentos incomensuráveis como por exemplo o conceito de informação para a Ciência da Informação (registrada e portadora de sentido) e para a Ciência da Computação (passível de representação matemática).

Para Kuhn (2003) uma comunidade científica pode ser definida como um grupo produtor e legitimador do conhecimento científico e apresenta algumas características: seus integrantes são treinados de forma semelhante; compartilham e absorvem a mesma literatura técnica/científica; aprendem as mesmas lições; o grupo possui um objeto científico próprio, embora este possa ser abordado por pontos de vistas de diferentes escolas; o grupo possui um sistema de comunicação amplo com avaliação pelos pares; e constituem uma única audiência para essa comunicação. De maneira resumida pode-se definir comunidade científica como grupo de pesquisadores praticantes de determinada especialidade científica.

Diante dos fundamentos teóricos apresentados, este capítulo pretende caracterizar duas ciências que tiveram origem na revolução científica e técnica provocada pela Segunda Guerra Mundial: a Ciência da Informação e a Ciência da Computação, como matrizes disciplinares distintas que compartilham objetos de estudo como a recuperação da informação, a Web e

suas estruturas. Ambas as matrizes são consideradas campos interdisciplinares e que desenvolvem pesquisa em ciência aplicada.

2.1 Ciência da Computação

A Ciência da Computação é um campo interdisciplinar desenvolvido num núcleo teórico das ciências exatas ou ciências duras. Para Denning et al (1989) esta disciplina trata apenas de processos que possam ser executados através de algoritmos. Fonseca Filho (1999, p.13) define a Ciência da Computação "como um corpo de conhecimento formado por uma infra-estrutura conceitual e um edifício tecnológico onde se materializam o hardware e o software. A primeira fundamenta a segunda e a precedeu." Os paradigmas desta matriz disciplinar, direta ou indiretamente, estão materializados em programas (software) ou dispositivos (hardware) comprovando seu caráter de ciência aplicada.

Sob uma perspectiva histórica, a Ciência da Computação é marcada por sucessões de revoluções científicas, com interrupções repentinas de paradigmas ou sua alternância de forma inesperada. Fonseca Filho (1999) afirma que se torna impossível visualizar a evolução da área como uma linha do tempo através de invenções-pesquisadores-data, sendo necessária uma análise dos fatos conceituais que fundamentaram esta matriz disciplinar que procura formas e métodos de mecanização do raciocínio.

A Segunda Guerra Mundial é um marco para a Ciência da Computação quando efetivamente foram construídos os primeiros computadores digitais e ocorreu um desenvolvimento científico e tecnológico exponencial com o surgimento de diversas sub-especialidades. Estudar o estado da arte na Ciência da Computação atualmente exige a seleção de perspectivas, campos ou áreas específicas como: linguagem de programação, teoria da computação, robótica, inteligência artificial, computação gráfica, etc.

Pode-se afirmar, no entanto, que a Ciência da Computação é uma matriz disciplinar com fundamentos históricos vinculados no conceito abstrato de número, portanto com forte vínculo com a Matemática e sua característica de síntese, porém uma ciência com objetivo aplicado visto que se desenvolve em função de necessidades dos usuários para a solução de problemas do mundo real. Abaixo a Figura 1 mostra o caráter aplicado da Ciência da Computação:

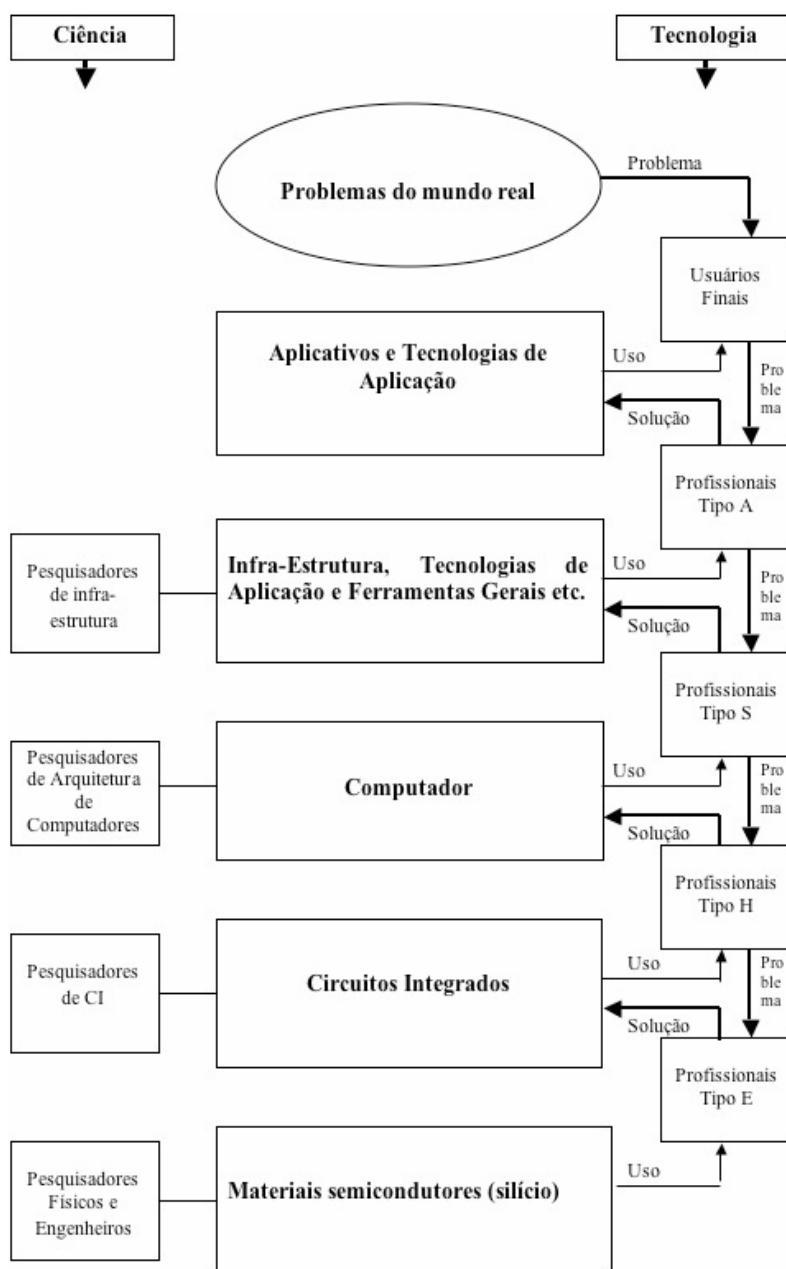


Figura 1. Caráter aplicado da Ciência da Computação. (NUNES, 2004) * CI = Circuitos integrados

A elipse superior representa os problemas do mundo real cujas soluções podem ser encontradas na Ciência da Computação: problemas nas áreas da engenharia, administração, lazer, entretenimento, educação, comunicação, artes, automação, etc. As caixas centrais representam as ferramentas que oferecem suporte na solução dos problemas. O lado direito da figura representa o desenvolvimento tecnológico e os profissionais especializados na solução dos problemas, estabelecendo uma relação entre estes e as ferramentas disponíveis. O lado esquerdo da figura representa o desenvolvimento científico e os pesquisadores envolvidos no

processo. O desenvolvimento científico dentro da ciência aplicada proporciona o aprimoramento das ferramentas disponíveis para a solução dos problemas, facilitando seu uso e/ou removendo suas limitações.

Sob o ponto de vista dos paradigmas que fundamentam as pesquisas em Ciência da Computação, Denning et al.(1989) destacam três processos como principais: a teoria, a abstração e o projeto. Esses três processos fundamentam-se em paradigmas de diversas disciplinas que contribuem para a formação da Ciência da Computação como matriz disciplinar. O primeiro processo denominado teoria fundamenta-se nas disciplinas do campo matemático e orientam as pesquisas sob o ponto de vista teórico na caracterização dos objetos de estudo, no levantamento de hipóteses de relacionamento entre estes, na determinação da veracidade das relações e na interpretação de resultados. O processo denominado de abstração tem suas raízes na ciência experimental e seus paradigmas orientam as pesquisas sobre fenômenos através do levantamento de hipóteses, construção de modelos, aplicação do modelo, coleta de dados e análise dos resultados. O terceiro processo tem seus fundamentos na Engenharia e proporcionam a construção de sistemas para a solução de problemas. No grupo projeto os paradigmas orientam na análise do problema através do estudo das necessidades e especificações da situação, desenvolvimento de um sistema e teste deste sistema.

Theory is the bedrock of the mathematical sciences: applied mathematicians share the notion that science advances only on a foundation of sound mathematics. Abstraction (modeling) is the bedrock of the natural sciences: scientists share the notion that scientific progress is achieved primarily by formulating hypotheses and systematically following the modeling process to verify and validate them. Likewise, design is the bedrock of engineering: engineers share the notion that progress is achieved primarily by posing problems and systematically following the design process to construct systems the solve them.[...] Closer examination reveals that in computing the three process are so intricately intertwined that it is irrational to say that any one is fundamental. (DENNING et al., 1989, p.10.)²

² Teoria é a base das ciências matemáticas: matemáticos compartilham a noção de que a ciência avança apenas sobre uma sólida base matemática. Abstração (modelagem) é a base das ciências naturais: cientistas compartilham a idéia de que os avanços científicos são obtidos primariamente pela formulação de hipóteses e sistematicamente seguindo a modelagem de processos para verificá-las e validá-las. Do mesmo modo, o projeto é a base da engenharia: engenheiros compartilham a idéia de que o progresso é obtido primariamente pela proposição de problemas e seguindo sistematicamente o projeto para construção de sistemas para solucioná-los. [...]Uma observação minuciosa revela que na computação os três processos estão tão intrinsecamente entrelaçados que é irracional dizer que qualquer um é fundamental. (tradução nossa)

A Matemática e a Engenharia são bases fundamentais para o desenvolvimento da Ciência da Computação a partir da segunda metade do século passado. O desenvolvimento da Matemática desde a criação abstrata do conceito de número, passando pelo surgimento de suas sub-especialidades como a Álgebra, Lógica Matemática até o desenvolvimento no início do século XX de noções de computabilidade dos processos permitem a criação de uma nova matriz disciplinar preocupada em mecanizar diversas formas de raciocínio. É a união de matemáticos e engenheiros através deste interesse em comum que forma uma nova comunidade científica preocupada em pesquisar e desenvolver dispositivos e programas capazes de executar operações computacionais que substituem a noção de intuitivo na execução de procedimentos por uma execução formalizada através de uma estrutura matemática: os algoritmos.

A Figura 2 apresenta o relacionamento entre os mundos formal, matemático e computacional de Alan Turing, matemático inglês que em 1936 demonstrou ser possível executar operações computacionais sobre a teoria dos números através de uma máquina que tenha embutida as regras de um sistema formal matemático (FONSECA FILHO, 1999) É este um dos marcos da Computação ao demonstrar ser possível a transformação de formas de raciocínio do mundo real em representações, tanto no mundo matemático quanto no computacional.

No mundo real através do uso de alfabeto, sintaxe, axiomas e determinadas regras de inferência produz-se teoremas que podem ser transformados em representações no mundo matemático através de uma semântica, da aritmética, geometria, etc. O mundo computacional também tem seus recursos para a representação destes teoremas do mundo real, como dados de entrada, instruções do programa etc. Destaca-se aqui a questão: qual o alcance e quais os limites de um computador? Quanto do raciocínio presente no mundo real pode ser transformado em algoritmos inteligíveis pela máquina?

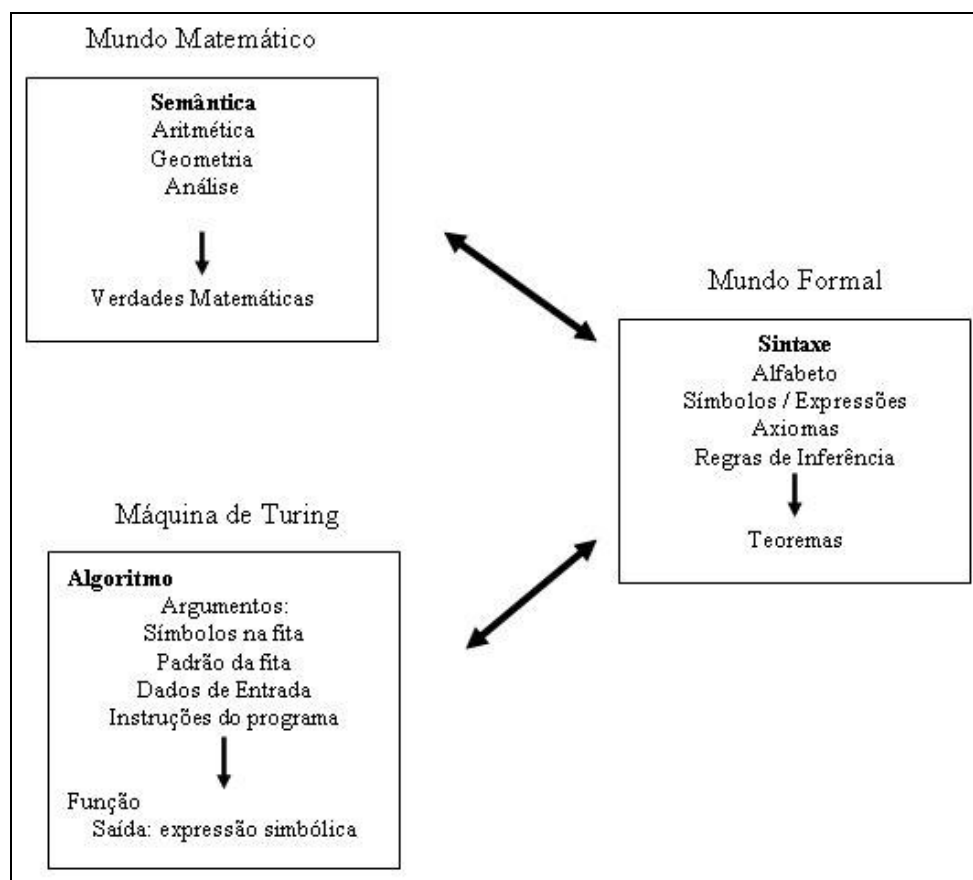


Figura 2. Relacionamento entre os mundos formal, matemático e computacional de Alan Turing (FONSECA FILHO, 1999, p. 88)

Um fator para o surgimento de uma nova comunidade científica e a consolidação da Ciência da Computação como matriz disciplinar foi, portanto, o interesse comum entre matemáticos e engenheiros em desenvolver máquinas (hardware) e sistemas (software) capazes de executar raciocínios presentes no mundo real. A evolução de hardwares e softwares permitiu à Ciência da Computação expandir seu leque de pesquisa como por exemplo: a Recuperação da Informação, Inteligência Artificial, Teoria de Bancos de Dados, etc. A partir de 1960 a Ciência da Computação se consolida definitivamente como uma matriz disciplinar com a formação não só de grupos de pesquisa como de pesquisadores doutores na área. Vale aqui relembrar a teoria kuhniana que considera a educação como um fator primordial na formação de seguidores de uma matriz disciplinar. É na prática da atividade científica que pesquisadores são formados, visto que os fatos tendem a ser interpretados dentro dos paradigmas dominantes da matriz disciplinar. Os futuros pesquisadores precisam “conhecer os paradigmas que orientam a interpretação de modo a fornecer as respostas consideradas corretas pela comunidade científica”(DENCKER, 2001)

Desde a década de 50 do século passado vários pesquisadores buscam soluções para o tratamento computacional da informação. Uma das soluções bem sucedidas é a indexação e elaboração automática de resumos, resultado da pesquisa de um engenheiro da IBM, Hans Peter Luhn, desenvolvedor de vários projetos que modificaram radicalmente os métodos tradicionais de armazenamento, tratamento e recuperação da informação (FERNEDA, 2003). A partir daí estabelece-se uma sub-disciplina da Ciência da Computação: a Recuperação da Informação que “trata os aspectos intelectuais da descrição da informação e sua especificação de busca, e também de qualquer sistema, técnicas ou máquinas que são empregadas nesta operação.” (MOOERS, 1951 apud FERNEDA, 2003, p.11)

A Recuperação da Informação consolidou-se como área de pesquisa de duas matrizes disciplinares: a Ciência da Computação e a Ciência da Informação, matriz disciplinar que também tem origem na revolução científica e técnica provocada pela Segunda Guerra Mundial.

A seguir procura-se caracterizar a Ciência da Informação como uma matriz disciplinar.

2.2 Ciência da Informação

Assim como a Ciência da Computação tem seu desenvolvimento histórico estruturado no conceito abstrato de número e no desenvolvimento no campo da Matemática, a Ciência da Informação está vinculada historicamente ao desenvolvimento da escrita e a possibilidade de se registrar fisicamente o conhecimento através dela. O desenvolvimento das técnicas de escrita até o surgimento da prensa de tipo móvel permitiu o registro, estoque e a recuperação do conhecimento. Assim como outras formas de registro além do texto escrito, como as imagens fotográficas, a microfilmagem, o registro de sons e de imagens em movimento, também colaboraram historicamente para a estruturação da Ciência da Informação.

Ainda num rápido percurso histórico da Ciência da Informação podemos destacar a figura de Paul Otlet, que no início do século XX cunhou do termo Documentação e sistematizou e previu tecnologias para sua operacionalização. Outro destaque histórico dentro da Ciência da Informação é Vannevar Bush que no período pós Segunda Guerra Mundial publicou um artigo considerado um marco na busca de soluções para os problemas da explosão informacional gerada no período pós-guerra. No artigo “As We may think” Bush propõe como solução para os problemas de gerenciamento de informações o desenvolvimento

de uma máquina capaz de agregar diversas tecnologias de informação da época. (FERNEDA, 2003)

Sob o ponto de vista de desenvolvimento paradigmático da Ciência da Informação, diversos autores têm procurado estabelecer fronteiras, objetos de estudo, definições, identificar uma estrutura disciplinar ou colaborações disciplinares. (SARACEVIC, 1996; LE COADIC, 2004; NEHMY et al., 1996; EUGÊNIO, FRANÇA E PEREZ, 1996; ARAÚJO, 2003; PINHEIRO, LOURENÇO, 1995; SARACEVIC, 1995; SAYÃO, 2001)

Para Saracevic (1996) a Ciência da Informação tem a Biblioteconomia, a própria Ciência da Computação, a Comunicação e a Ciência Cognitiva como disciplinas que estabeleceram relações mais profundas dentro de seu escopo. Le Coadic (2004) salienta também outras disciplinas como principais colaboradoras da Ciência da Informação como a Psicologia, Linguística, Sociologia, Informática, Matemática, Lógica, Estatística, Eletrônica, Economia, Direito, Filosofia, Política e Telecomunicações. O destaque é que a interdisciplinaridade da Ciência da Informação está ligada a um núcleo teórico onde predominam as ciências sociais.

A análise da relação Ciência da Informação e adoção de paradigmas está presente no texto de diversos autores. Nehmy et al. (1996), sob a perspectiva kuhniana, analisaram a Ciência da Informação e sugeriram que esta matriz disciplinar “teria outra natureza, diversa das ciências naturais, embora [...] seu desenvolvimento pareça também diferenciado do trajeto percorrido pelas ciências sociais” (p. 22). Os autores constatam claramente a existências de “duas correntes em oposição [...] a da defesa da Biblioteconomia como núcleo duro da Ciência da Informação [...] e uma outra que é a da proposta de autonomia dessa disciplina”.(p. 19). Para Wersig apud Nehmy et al (1996) a Ciência da Informação é uma ciência pós-moderna por tratar de problemas de nova complexidade. Afirma que o campo da Ciência da Informação tem sido objeto de muitas disciplinas fragmentadas e enfatiza uma construção teórica diferente da ciência normal que possa sustentar esta nova complexidade.

Eugênio, França e Perez (1996) recuperam algumas das fronteiras da Ciência da Informação identificando algumas definições das quais se destacam:

A Ciência da Informação como uma ciência que investiga as propriedades e comportamento da informação, os processos de fluxo e os meios de processar a informação visando sua acessibilidade e utilidade. Estes processos envolvem a origem, disseminação, coleção, organização, armazenamento, recuperação, interpretação e uso da informação. Seus

seguidores consideram a Ciência da Informação derivada ou relacionada com a Matemática, a Lógica, a Lingüística, a Psicologia, a Ciência da Computação, a Comunicação, a Biblioteconomia, a Administração, etc. Diversas disciplinas das áreas de exatas e sociais.

Outro grupo de pesquisadores considera a Ciência da Informação como uma ciência cuja origem está não só na expansão e metamorfose da Documentação e da Recuperação da Informação, mas na incorporação paradigmas das áreas da Comunicação, das Ciências Comportamentais, Cognitivas, etc.

Consideram-se objetos de estudo da Ciência da Informação as questões ligadas ao conteúdo e interpretação da informação, o fenômeno da explosão informacional, a diversificação de suportes, o desenvolvimento de tecnologias de informação, dentre outros aspectos relativos à informação.

Tendo como base a definição de Masterman (1974) Sayão (2001, p. 86) aponta a Ciência da Informação como uma ciência multiparadigmática visto que pela sua própria natureza ampla e interdisciplinar, para mapear toda a sua realidade, teve obrigatoriamente de tomar, como seus, paradigmas e modelos de outras áreas, tais como informática, inteligência artificial, lingüística, economia, marketing

Assim como a Ciência da Computação, considera-se a década de 60 do século XX o início da consolidação da Ciência da Informação como matriz disciplinar com a formação de uma comunidade científica preocupada em solucionar problemas e estudar o tratamento documental e o uso de tecnologias nesse processo. Trata-se do início de sua institucionalização social com a constituição de uma comunidade científica e uma estrutura formal da matriz disciplinar que pressupõe a existência de mecanismos eficientes para a socialização e reprodução dos paradigmas. Oliveira, Mota e Alvarado (2004) destacam alguns elementos que caracterizam a institucionalização de uma matriz disciplinar: existência de instituições capazes de abrigar os grupos de pesquisadores, a existência de pesquisadores capacitados para o desenvolvimento e formação da atividade científica, canais de comunicação que permitam o fluxo de informação dentro da comunidade. Como salienta Kuhn (2003) no interior de uma matriz disciplinar a comunicação é relativamente ampla e os julgamentos profissionais relativamente unânimes. Utilizando-se dos pontos destacados anteriormente, Oliveira, Mota e Alvarado (2004) caracterizam a comunidade científica brasileira em Ciência da Informação através da identificação das instituições de pesquisa, dos canais de comunicação (periódicos e eventos) e das associações profissionais confirmando a cientificidade da Ciência da Informação.

2.3 Ciência da Informação, Ciência da Computação e o compartilhamento de objetos de estudo

Apresentadas as duas matrizes disciplinares, cumpre estabelecer qual o objeto de estudo comum a elas. Considerado o componente tecnológico da Ciência da Informação, a Ciência da Computação ou “tecnologia da computação” permite o compartilhamento de diversos temas de pesquisa como a Recuperação da Informação, a ambiência digital da informação, a estrutura hipertextual dos documentos digitais, a estruturação de metadados representativos da informação, etc.

A Recuperação da Informação foi um dos primeiros objetos de estudo destas duas matrizes: a Ciência da Informação e a Ciência da Computação. Saracevic (1996) afirma que a Recuperação de Informação pode ser considerada a vertente tecnológica da Ciência da Informação e é resultado da relação desta com a Ciência da Computação.

O conceito de recuperação da informação envolve um processo de identificação, dentro de um conjunto de documentos de um sistema, quais são os que atendem às necessidades de informação dos usuários. Estas necessidades são representadas através de um ‘expressão de busca’ que podem ser especificadas em linguagem natural ou através de linguagens artificiais e devem resultar na recuperação de um número de documentos que possibilitem a verificação de cada um deles para a seleção dos considerados úteis.

O advento da Internet como grande repositório de documentos e das informações neles contidas favoreceu o surgimento de pesquisas sobre recuperação de informação na Web. A Internet caracteriza-se por uma infra-estrutura de redes e servidores capazes de estabelecerem canais de comunicação para troca de informações em nível global através de recursos informacionais textuais e multimídias. Mecanismos de busca foram desenvolvidos especificamente para este novo ambiente informacional e atualmente com o objetivo de melhorar a recuperação de informação neste grande repositório, as pesquisas buscam encontrar formas de agregar valor semântico às páginas da Web.

Procura-se aumentar a eficiência dos mecanismos de busca e de outros tipos de ferramentas de processamento automático de documentos através da utilização de linguagens que permitam definir dados e regras para o raciocínio sobre esses dados. (FERNEDA, 2003, p. 110)

O desenvolvimento da Internet tem início no período pós-guerra com a Guerra Fria. O objetivo principal era descentralizar informações militares, em virtude do auge da Guerra Fria. Sua estrutura não possui um centro gerenciador, nem mesmo rotas únicas e específicas

para a transmissão dos dados. Qualquer dano ou impossibilidade de acesso de um centro da rede não prejudica o restante dos centros que compõem a rede.

No final da década de 1980, em 1989, Tim Berners-Lee começa a pesquisar e desenvolver uma tecnologia...

para o compartilhamento de informação usando documentos textuais que se referenciavam através de ligações. O objetivo inicial era construir uma ferramenta de comunicação baseada na Internet para compartilhar informação com diferentes universidades no mundo todo. Berners-Lee criou uma linguagem de marcação [...] e batizou-a de HTML (*HyperText Markup Language*). (FERNEDA, 2003, p. 92)

Berners-Lee elabora esta proposta que permite a criação de um ambiente gráfico para uso na Internet com a difusão de documentos textuais, sonoros e imagéticos de forma integrada. Ferneda (2003, p. 92) destaca que o pesquisador também desenvolveu “protocolos de comunicação para formar a espinha dorsal do seu novo sistema de informações em hipertexto, o qual denominou de *World Wide Web* ou simplesmente *Web*.” Ramalho, Vidotti e Fujita (2007, p.2) afirmam que a *Web* “tem como uma de suas principais funcionalidades a tarefa de disponibilizar conteúdos informacionais de modo que estes possam ser visualizados e interpretados por usuários humanos.”

O desenvolvimento tecnológico que proporcionou o surgimento da Internet e posteriormente da *Web* trouxe atrelado a si o desenvolvimento exponencial de documentos em ambientes *Web* e os problemas que dificultam a localização de informações nessa grande massa documental.

Para Mendez Rodriguez (2002) a Recuperação de Informação na *Web* é diferente da realizada nos sistemas tradicionais de informação, pois seus recursos informacionais não são fixos e estáveis podendo desaparecer; não passaram por um processo de seleção e representação; não são organizados de forma a facilitar o acesso; não são catalogados nem indexados.

As ferramentas de buscas existentes na *Web*, em geral, procuram estabelecer a relevância dos documentos em relação à expressão de busca através de *rankings* de relevância que ordenam uma lista de resultados que satisfaçam a expressão e na qual as páginas que possivelmente sejam as mais adequadas aparecem nas primeiras posições. (MENDEZ RODRIGUES, 2002; FERNEDA, 2003)

As ferramentas que trabalham com robôs de busca normalmente se interessam em selecionar a informação, alimentar os bancos de dados com a maior quantidade

possível de informação. Esse trabalho, efetuado pelos robôs, inicia na localização de uma URL, passa pela descrição automática das informações contidas no site e termina no registro das informações descritivas e temáticas na imensa base de dados, sem nenhuma interferência uma. (SANTAREM SEGUNDO, 2004, p. 83)

Para Cendón (2001, p.42)

A maioria dos motores de busca indexa [...] cada palavra do texto completo, apenas o URL, as palavras que ocorrem com frequência ou palavras e frases mais importantes contidas no título ou nos cabeçalhos e nas primeiras linhas, por exemplo. Alguns motores indexam outros termos que não fazem parte do texto visível, mas que contêm informações importantes e úteis.

Em 2001, Berners-Lee, Hendler e Lassila publicam no *Scientific American* uma proposta para solucionar problemas na recuperação de informações na Web: o estabelecimento de padrões ou metadados que permitam ao computador compreender a semântica dos textos digitais, disponíveis na Internet. Os autores denominaram esta nova metodologia para conteúdos digitais de 'Web Semântica'. Os autores consideram esta proposta como uma extensão e aprimoramento da Web atual, pois a informação passa a possuir um significado claro e definido, possibilitando uma maior e melhor interação entre os computadores e os seres humanos.

Recuperar a informação de forma qualitativa neste grande repositório informacional que é a Internet tem sido objeto de estudo de diversas matrizes disciplinares que através da organização de Congressos, Conferências, Seminários, Periódicos e Grupos de pesquisa buscam construir paradigmas capazes de explicar e aprimorar o sistema. Como afirmam Ramalho, Vidotti e Fujita (2007, p.2) “a expressão Web passou a ser disseminada como um título genérico que representa uma série de pesquisas que têm como objetivo principal possibilitar um melhor aproveitamento das potencialidades do ambiente Web”, para isso utiliza linguagens computacionais e metadados que permitem seu acesso automatizado de forma mais precisa.

É consenso dentro da Ciência da Informação que a informação organizada de forma adequada permite sua recuperação de forma precisa no menor tempo possível, meta principal das pesquisas envolvendo recuperação da informação. Estes e outros paradigmas regem a Ciência da Informação independente do suporte do documento onde estejam registradas as informações. McGarry(1999, p. 11) afirma que “a informação deve ser ordenada estruturada ou contida de alguma forma, senão permanecerá amorfa e inutilizável”, para isso ela necessita de um tratamento que permita sua representação de forma compreensível aos seres humanos.

A agregação de valores semânticos às páginas da Web e o estabelecimento de metadados para a descrição da informação, é um dos focos da agenda de pesquisa em Ciência da Computação e que muito se assemelham aos paradigmas da Ciência da Informação. As pesquisas envolvendo maior padronização das páginas da Web através de metadados podem ser consideradas um paralelo na agenda das duas matrizes disciplinares. Furgeri (2006, p. 28) considera que para a Ciência da Informação “a representação de uma informação passa pela definição de um conjunto de elementos básicos e de regras para a conjunção desses elementos” num processo semelhante ao que ocorre na Ciência da Computação no qual computadores representam a informação através de bits e utilizam regras específicas para agrupar esse bits na formação de caracteres. O autor analisa ainda algumas propostas de organização da informação na Web que estabelecem ligações entre a Ciência da Computação e a Ciência da Informação:

- Diretórios – tem como função básica organizar os conteúdos armazenados e facilitar sua recuperação Céndon (2001) considera os diretórios a primeira proposta para organização e localização da Web;
- Linguagens de marcação: codificação que permite descrever a estrutura lógica ou semântica de um documento, fornecendo assim elementos que permitem a apresentação do conteúdo de um arquivo;
- XML: linguagem computacional de marcação definida como um padrão para o ambiente Web, capaz de estabelecer valores semânticos os documentos, facilitando os processos de recuperação e disseminação da informação;
- XML Schema: estrutura de regras que permite a compreensão de documentos XML em diferentes contextos, define padrões que se deve seguir para que o documento seja considerado válido.

Portanto o paradigma da organização da informação digital através de normas e padrões como solução para o aumento do grau de precisão na recuperação da informação passa a ser compartilhado por duas ciências que constroem, analisam e comunicam suas pesquisas de forma diferenciada e para públicos específicos e distintos.

No capítulo seguinte procura-se apresentar o desenvolvimento científico da *World Wide Web*.

3 DESENVOLVIMENTO CIENTÍFICO DA WEB

A *World Wide Web*, comumente denominada Web, é resultado das pesquisas para o desenvolvimento científico e tecnológico da Internet. Como apresentado no Capítulo anterior, este ambiente informacional está baseado em protocolos de comunicação e numa linguagem de marcação que estabelece uma estrutura hipertextual entre documentos textuais, imagéticos, sonoros. Santarém Segundo (2004, p. 32) considera a Web “o serviço de informação da Internet que consiste em milhares de páginas, formadas por gráficos e fotografias, combinadas com texto” baseado no conceito de hipertexto

A rede Internet pode ser entendida e visualizada como um labirinto documental no qual as informações armazenadas e apresentadas na *World Wide Web* (WWW ou Web) são estruturadas em sites/home-pages em forma de redes hipertextuais. (VIDOTTI, 2001, p. 1)

Ferneda (2003) considera a Web a face hipertextual da Internet e uma das principais fontes de informação em diversas áreas. Para Santarém Segundo (2004) a Internet é atualmente o canal de comunicação que possibilita a maior interatividade entre as pessoas na busca por troca de informações e o maior repositório para armazenamento de informações existente.

A Web é composta unidades de informação denominadas páginas e/ou *sites* e/ou *home-pages*, que utilizam a linguagem de marcação HTML (*HyperText Markup Language*) que permite a criação de ambientes gráficos e de documentos digitais facilmente legíveis por humanos.

Estes *sites* possuem a seguinte estrutura:

- Um esquema de endereçamento denominado *Universal Resource Locator* (URL)
- Um protocolo *Hypertext Transfer Protocol* (http).
- E uma estrutura da página Web especificada através do padrão *HyperText Markup Language* (HTML)

A URL permite e é indispensável para a localização de um recurso disponível da Web, pois é este padrão que permite ao *browser* localizar um recurso por ele solicitado. E assim concretizar a transferências dos dados entre o computador-servidor e o computador-solicitante através de outro padrão aberto o protocolo HTTP. O terceiro padrão aberto a ser destacado é o HTML que é a linguagem padrão de marcação para o desenvolvimento das páginas Web.

Este padrão, segundo Ramalho (2003), possibilita a representação dos dados de forma simplificada e prioriza a forma de apresentação dos conteúdos em detrimento de uma melhor descrição das estruturas semânticas das páginas Web.

Esta linguagem de marcação apresenta um conjunto de códigos denominados *tags* utilizados para definir os componentes, a apresentação e funcionalidades das páginas Web como, por exemplo: título, autoria, figuras, etc. Como afirma Ferneda (2003) uma página Web pode incluir *tags* que redirecionem para outras URLs, estas ligações são chamadas de *links* e formam uma estrutura complexa que caracteriza a adoção do termo Web, teia em inglês. As *tags* de uma página em HTML informam aos *browsers* como o documento deve ser apresentado, descrevendo sua aparência e forma de apresentação.

É a junção de simplicidade e possibilidades de uso de inúmeros recursos na geração de páginas Web proporcionada pelo HTML, adicionado à adoção de outros padrões abertos que proporcionou a rápida disseminação da Web nos moldes que conhecemos atualmente. Trata-se de uma linguagem extremamente simples; possui estilo próprio para formatação dos documentos; permite a criação de ligações hipertextuais de maneira fácil; possibilita a interação Homem-máquina além de uma programação simplificada. A estrutura de *tags* de uma página forma um arquivo texto que pode ser perfeitamente criado num editor de texto.

Sua maior vantagem também é fonte de problemas, ao utilizar as *tags* para definir forma e estrutura ao mesmo tempo, a linguagem HTML impossibilita a atribuição de significados aos conteúdos das páginas Web, comprometendo a eficácia na recuperação da informação nos ambientes Web através das ferramentas de busca.

...usando *tags* tanto para marcar trechos do conteúdo do documento quanto para informar ao programa navegador como tais trechos devem ser exibidos [...]. Outro fator de limitação da linguagem HTML é que ela impossibilita a atribuição de significados aos conteúdos das páginas, fazendo com que os motores de busca, *search engines*, recuperem um grande volume de dados que não são relevantes para os usuários finais. (RAMALHO, 2006, p. 35)

3.1 Recuperação da informação na Web

Para identificar na Web a informação desejada existem diversos mecanismos para sua localização na Web, estes mecanismos são denominados ferramentas de busca, como por exemplo: diretórios, motores de busca ou *search engines*, metamotores.(CENDÓN, 2001; SANTARÉM SEGUNDO, 2004).

Os diretórios foram a primeira solução para organizar e localizar as informações na Web, organizam os *sites* de sua base de dados em categorias, subcategorias, sucessivamente de forma hierárquica de assunto. Como salienta Cendón (2001, p.39) “foram introduzidos quando o conteúdo da Web ainda era pequeno o suficiente para permitir que fosse coletado de forma não automática”.

Os motores de busca não possuem uma organização hierarquizada dos *sites* que compõem sua base de dados. Procuram armazenar o maior número possível de recursos informacionais através de *softwares* de busca denominados de robôs. Estes robôs varrem a Web seguindo os *links*, descrevendo e indexando de forma automática as informações coletadas. Encontra-se na literatura especializada diversas denominações para os robôs de busca: aranhas (*spiders*), agentes, viajantes (*wanderers*), rastreadores (*crawlers*) ou vermes (*worms*). (CENDÓN, 2001; SANTARÉM SEGUNDO, 2004)

A maioria dos motores de busca indexa, ou seja, inclui em seu índice, cada palavra do texto completo, apenas o URL as palavras que ocorrem com frequência ou palavras e frases mais importantes contidas no título ou nos cabeçalhos e na primeiras linhas, por exemplo Alguns motores indexam outros termos que não fazem parte do texto visível, mas que contém informações importantes e úteis. (CENDÓN, 2001, p. 42)

Os metamotores são ferramentas de busca que não possuem nenhuma base de dados própria, são serviços que utilizam *softwares* que pesquisam dados de outras ferramentas de busca. Esta ferramenta geralmente realiza um pré-processamento da consulta realizada por usuários preparando e estruturando a busca para submissão em cada ferramenta de busca da Web: diretórios e motores de busca. (CENDÓN, 2001; SANTARÉM SEGUNDO, 2004)

Como afirma Santarém Segundo (2004) as ferramentas de busca têm desenvolvido técnicas que proporcionem uma maior revelância na recuperação da informação na Web, no entanto algumas limitações persistem dificultando a satisfação dos usuários. O autor salienta questões como a grande quantidade de documentos e informação disponíveis na Web, além da velocidade com que estes documentos são produzidos e ao mesmo tempo desaparecem da Web e a baixa padronização dos documentos na Internet.

Santarém Segundo (2004, p. 106) destaca ainda outras limitações das ferramentas de busca nos ambientes Web desenvolvimentos com a linguagem HTML

A facilidade de interpretação que o ser humano tem em distinguir uma palavra em um determinado contexto não é encontrada nos computadores e nos robôs de busca,

não permitindo, assim, que os mesmos consigam entender o conteúdo significativo de uma página Web .

No processo de desenvolvimento tecnológico da Web, novas tecnologias propõem um novo formato para este ambiente, na busca por uma melhor forma de organização das informações. Ramalho (2006, p36) afirma que “verifica-se atualmente uma grande demanda de estudos relacionados ao desenvolvimento de mecanismos de representação de recursos informacionais”. As pesquisas atualmente desenvolvidas procuram encontrar padrões que possibilitem agregar um maior nível semântico às páginas Web. O objetivo destas pesquisas é “aumentar a eficiência dos mecanismos de busca e de outros tipos de ferramentas de processamento automático de documentos através da utilização de linguagens que permitam definir dados e regras para o raciocínio sobre esses dados”. (FERNEDA, 2003, p. 110)

A Web Semântica é a proposta de desenvolvimento da tecnologia Web que tem como objetivo proporcionar a execução de tarefas mais sofisticadas pelos computadores e é resultado destas pesquisas que buscam dar significado às informações disponibilizadas em ambientes Web.

3.2 Web Semântica

As pesquisas sobre Web Semântica têm como objetivo tornar possível um melhor aproveitamento das potencialidades dos ambientes Web, através do uso de linguagens de computação e metadados que permitam que o acesso automatizado às informações ocorra de maneira mais precisa. (RAMALHO, 2006) Para Berners-Lee, Lassila, Hendler (2001), precursores da Web Semântica, esta cria um ambiente propício para que os robôs de busca possam realizar tarefas mais sofisticadas e disponibilizá-las aos usuários. Os autores a consideram uma extensão da Web atual, sendo resultado do desenvolvimento científico e tecnológico que permite que a informação possua um significado claro e bem definido dentro dos ambientes Web, possibilitando uma melhor interação entre Homem-máquina. Para Santarém Segundo (2004, p. 107) a Web Semântica “permitirá que computadores e pessoas possam trabalhar cooperativamente no processo de descrição, armazenamento e recuperação de informações digitais”.

Ramalho (2006) salienta que fica claro que o objetivo final da Web Semântica é atender as pessoas e não os computadores, mas que para isso é necessária a construção de instrumentos que produzam sentido lógico e semântico para as máquinas. Os computadores não são capazes de estabelecer associações de significado, a proposta Web Semântica é

proporcionar um arranjo de idéias e suas associações de forma irrestrita sob o ponto de vista computacional e, desta forma proporcionar ao computador associações entre coisas não relacionadas explicitamente. (SANTARÉM SEGUNDO, 2004; CAMPOS, CAMPOS, CAMPOS, 2006; RAMALHO, 2006)

Santarém Segundo (2004) destaca como ponto fundamental da Web Semântica a criação de uma nova estrutura de armazenamento de dados. O ponto principal está na separação da apresentação do conteúdo e do conteúdo da estrutura como componentes independentes da informação em ambientes Web.

Nafria (2001 apud FEITOSA, 2006) afirma que desta forma a Web se tornará um ambiente com capacidade de processamento automático da informação e não mais um repositório de documentos.

Segundo Codina e Rovira (2006) a Web Semântica é um projeto a curto, médio e longo prazo que inclui transformações que estão afetando a criação, edição e publicações de páginas Web e que seguirá tendo uma importância crescente no futuro. Acrescentar regras semânticas aos ambientes Web alterará sua natureza radicalmente, de um ambiente que somente exibe informações para um ambiente no qual estas mesmas informações podem ser interpretadas, trocadas e processadas por máquinas (computadores).

Para Ramalho (2006, p.40) é a capacidade de associação de relacionamentos não explícitos entre documentos que sustenta o uso de ontologias na estrutura da Web Semântica “de modo que se espera que, com o desenvolvimento de ontologias formais, seja possível descrever as informações semânticas dos recursos Web, possibilitando o compartilhamento e manipulação de informações”.

3.2.1 Infra-estrutura da Web Semântica

O ponto chave para o desenvolvimento da Web atual está no desenvolvimento e aplicação de tecnologias capazes de estruturar a informação para que de forma automatizada possa ser processada e compartilhada. Senso Ruiz (2006) destaca a importância dos metadados que oferecem possibilidades de trabalho em diferentes níveis dos objetos digitais através de facilidades como:

- Incremento do acesso: um conjunto de metadados que descrevam corretamente um ou vários objetos digitais aumenta a possibilidade de acessá-los;

- Diminuição do tráfego na rede: para se indexar a representação de um objeto e não o objeto em si não exige uma configuração de banda muito larga para busca na Web;
- Expansão do uso da informação: os metadados facilitam a difusão das versões digitais dos objetos;
- Permite a gestão legal dos objetos: pode-se estabelecer restrições de acesso, informar sobre direitos autorais, permite o controle de uso, etc;
- Preservação do objeto original.

Codina e Rovira (2006) listam as tecnologias como normas, protocolos, linguagens disponíveis para que os objetivos da Web Semântica sejam alcançados:

- Em primeiro lugar está a codificação de páginas Web, através da linguagem de marcação XML (*eXtensible Markup Language*), com o uso de *tags* que tenham uma carga semântica;
- Em segundo lugar o uso de metadados descritores das páginas Web com formato compatível com a estrutura da Web permitindo a interoperabilidade entre diferentes sistemas. Para esta camada utiliza-se o padrão RDF (*Resource Description Language*)
- Em terceiro lugar um sistema de ontologias que permitam a especificação de conceitos em diversos domínios do conhecimento através de uma linguagem baseada na lógica simbólica e que permita sua interpretação por computador. Para esta camada da Web Semântica utiliza o padrão OWL (*Web Ontology Language*)

Todo desenvolvimento científico é baseado no intercâmbio informações resultantes das pesquisas realizadas num processo acumulativo, no caso da Web Semântica este desenvolvimento permitiu o estabelecimento de uma nova estrutura de metadados, linguagens, protocolos e padrões conforme a figura 3.

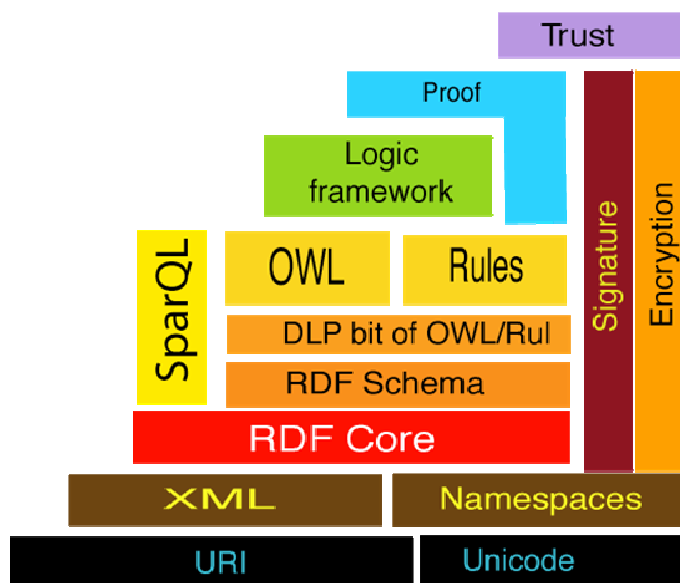


Figura 3. Arquitetura da Web Semântica (BERNERS-LEE, 2005)

É através do uso desta estrutura em camadas na construção da nova Web que os computadores processam automaticamente as informações contidas nos documentos digitais. Estas camadas são construídas através de linguagens de marcação e metadados em processos que se assemelham as representações descritivas e temáticas da Ciência da Informação.

Conforme destaque de Codina e Rovira (2006) segue-se um resumo das principais bases da Web Semântica de acordo com a atual proposta de arquitetura:

- XML: linguagem computacional que permite a estruturação dos dados através de definições de elementos e atributos, possibilita o estabelecimento de regras semânticas.
- *Namespace*: conjunto de nomes utilizados em documentos XML para validação de elementos e atributos
- RDF Core: núcleo que compreende as especificações do modelo e a sintaxe da RDF (*Resource Description Language*), permite a descrição dos recursos através de suas propriedades.
- RDF Schema: padrão usado para a descrição do vocabulário RDF, permite a definição hierárquica dos recursos.
- DLP: tecnologia que permite a intersecção de dois paradigmas de representação automatizada do conhecimento: a Lógica Descritiva e a Programação Lógica. Ramalho (2006, p 49) alerta que a DLP “ainda não é considerada atualmente como uma linguagem de representação do conhecimento”.

- OWL: linguagem computacional recomendada para o desenvolvimento de ontologias.

3.2.2 Ontologias

As ontologias são os elementos chave para a construção da Web Semântica, acredita-se que suas estruturas favoreceram o compartilhamento de informações dentro de contextos sobre um determinado domínio do conhecimento. O termo Ontologia no contexto da Ciência da Computação, especificamente no âmbito da Web Semântica, é utilizado como a modelagem para a representação de domínios do conhecimento. Tendo como objetivo o uso de linguagens computacionais capazes de processar a informação e realizar inferências automáticas. (CAMPOS, CAMPOS, CAMPOS, 2006; FEITOSA, 2006; RAMALHO, 2006)

Para Lima-Marques (2006, p. 50), na Ciência da Computação as ontologias podem ser descritas como um vocabulário de representação que reflete “o conjunto de objetos de um domínio representado em um formalismo declarativo, bem como os relacionamentos entre os objetos”.

Feitosa (2006, p. 121) em seu resumo considera que

do ponto de vista da representação do conhecimento, uma ontologia não deve ser concebida apenas como um vocabulário informal [...] na gênese das linguagens de marcação e de especificação de ontologias para a Web Semântica encontra-se a XML, que fornece uma sintaxe básica para a estrutura de documentos, mas não possui elementos que lhe permitam impor restrições semânticas ao significado de tais elementos. Fundados nessa linguagem estão os padrões XML-Schema, RDF, RDF-Schema, OWL, entre outros.

O desenvolvimento de ontologias é considerado um aspecto importante para o aprimoramento dos ambientes Web, pois permite o compartilhamento e reutilização de informações dentro deste ambiente informacional. “Uma ontologia constitui um documento ou arquivo que define formalmente as relações entre termos” (LIMA-MARQUES, 2006, p. 52).

O conceito de ontologia presente na Ciência da Computação se assemelha ao conceito de uma ferramenta de representação desenvolvida pela Ciência da Informação, o tesauro.

Tesauro é um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente, cobrindo um domínio específico do conhecimento, funcionando como um dispositivo de controle terminológico [...] são estruturas sistemáticas de conceitos. (CAMPOS, CAMPOS, CAMPOS, 2006, p. 60)

Em relação à ontologia Campos, Campos, Campos, (2006, p.61) afirmam que “muita confusão se tem feito em torno do conceito de ontologia, que não pode ser considerado somente como um vocabulário controlado”. Feitosa (2006, p.73) também ressalta que uma ontologia possui características que a difere do tesouro “do ponto de vista da representação do conhecimento, uma ontologia não deve ser concebida apenas como um vocabulário informal, ou mesmo uma linguagem de termos estruturados – como um tesouro” O autor considera necessária a interpretação algorítmica dos seus significados, portanto permitindo que o processamento dos significados seja realizado por computadores. “Uma ontologia requer a explicitação lógico-formal de significados e palavras, que devem ser expressos por meio de construtos matemáticos” (FEITOSA, 2006, p. 73)

Todas as tecnologias desenvolvidas no âmbito da Web Semântica trazem cada vez mais componentes que a situam como tema de fronteira entre a Ciência da Computação e a Ciência da Informação interligando instrumentos, ferramentas e conceitos presentes nas pesquisas de ambas as matrizes disciplinares. Ramalho, Vidotti, Fujita (2007, p.7) identificam “uma tendência de aproximação entre as áreas de Ciência da Informação e Ciência da Computação, principalmente no que tange ao desenvolvimento de novos instrumentos de representação e recuperação de recursos informacionais” Desde impulso inicial na Recuperação da Informação na Web até a construção de estruturas computacionais capazes de representar o conhecimento em determinada área e estabelecer inferências entre informações não relacionadas.

Retomando o percurso histórico de ambas as matrizes, estabelecem-se suas institucionalizações sociais na década de 1960, tendo a Ciência da Computação sua estrutura baseada na área de exatas e a Ciência da Informação sendo estruturada a partir de disciplinas das ciências sociais, humanas, além da própria Ciência da Computação. Portanto vale salientar que matrizes disciplinares distintas apresentam ‘estilos de pensamentos’ também distintos, como afirmam Ramalho, Vidotti, Fujita (2006, p. 7) “apesar de relacionadas, é possível identificar um distanciamento teórico entre tais Ciências”.

Sendo a Web Semântica um tema presente na agenda de pesquisa da Ciência da Informação e na Ciência da Computação, nacional e internacionalmente, existe a cooperação destas duas Ciências na construção deste conhecimento científico? Particularmente a produção científica brasileira em Web Semântica integra os pesquisadores da Ciência da Informação e da Computação? O capítulo seguinte procura identificar iniciativas de

cooperação nacional entre estas Ciências, explicitadas através de autorias e citações presentes na produção científica brasileira sobre Web Semântica e tecnologias relacionadas.

4 UM RECORTE DA PESQUISA BRASILEIRA EM WEB SEMÂNTICA

Os avanços tecnológicos na área da informação têm despertado os interesses de diversas áreas do conhecimento na análise de suas aplicações e impactos no mundo pós-moderno. Grupos multidisciplinares são formados, departamentos e sociedades científicas são criados, seminários e outros eventos são organizados, artigos científicos são publicados visando discutir, compreender, divulgar e solucionar questões envolvendo o mundo digital: sua organização, recuperação e impacto na sociedade da informação.

Os protocolos desenvolvidos em função da Internet podem ser considerados como uma preocupação da Ciência da Computação em promover o intercâmbio de informações de forma ampla, precisa e padronizada. Sem a existência destes padrões, a comunicação dos dados seria restrita a grupos que compartilhassem protocolos específicos e isolados entre si. No entanto, o desenvolvimento da Web provocou a inquietação dos cientistas da computação preocupados com a recuperação de informações. Pesquisas buscando estabelecer estruturas de metadados que agreguem valor semântico às páginas Web, tecnologia denominada Web Semântica, têm se consolidado na agenda de pesquisas de diversas matrizes disciplinares, com destaque para a Ciência da Computação e Ciência da Informação. Como salientam Ramalho, Vidotti, Fujita (2007, p. 11) “as tecnologias subjacentes ao projeto Web Semântica e os instrumentos de representação de informações desenvolvidos no âmbito da área da Ciência da Informação possuem como objetivo comum propiciar meios mais adequados de representar e organizar conteúdos informacionais” sendo, portanto um espaço propício para aplicação de alguns paradigmas da Ciência da Informação na representação de documentos Web.

Este capítulo procura caracterizar a autoria da produção científica brasileira sobre o tema através dos textos de doutores cadastrados numa base de dados de âmbito nacional, num esforço em identificar a participação destas e de outras matrizes disciplinares na construção do conhecimento científico no âmbito nacional. Um levantamento bibliográfico em paralelo identificou as teses e dissertações sobre Web Semântica e tecnologias relacionadas disponíveis nas diversas iniciativas de bancos de textos completos, a fim de identificar as referências bibliográficas citadas nos trabalhos e a influência de quais matrizes disciplinares nas pesquisas da Ciência da Informação e da Ciência da Computação.

4.1 O conhecimento científico e sua comunicação

A construção e/ou desenvolvimento do conhecimento científico envolve um processo de socialização de idéias pelos canais formais ou informais de comunicação, pela atividade científica nos laboratórios ou grupo de pesquisa, pela formação acadêmica, pelo aliciamento de seguidores, todos os fatores que caracterizam uma comunidade científica. Podemos, portanto afirmar que o desenvolvimento do conhecimento científico está fortemente ligado ao processo de comunicação dentro da comunidade de pesquisadores.

O sujeito acessa o conhecimento cumulativo, construído por outros e, com base na própria capacidade de apreensão, análise e reflexão, gera novo conhecimento. No entanto, acredita-se que o conhecimento somente será de fato construído, com sua socialização aos outros. (VALENTIM, 2005, p. 11)

A produção do conhecimento científico flui principalmente no âmbito da comunicação científica, seja ela formal ou informal. A publicação científica é uma etapa importante da atividade científica, permitindo a análise e legitimação pelos pares dos resultados finais ou parciais de uma pesquisa (MEADOWS, 1999) Griffith (1989 apud VANZ; CAREGNATO, 2003) considera a comunicação científica a única etapa comum a todas comunidades científicas independente da matriz disciplinar a que pertençam, visto que as outras etapas da atividade científica diferem conforme os procedimentos e técnicas de cada área do conhecimento. Comunicação científica é entendida como um ato, um processo de intermediação que permite a troca de idéias entre pesquisadores (TARGINO, 2000). Para Lyman (1997 apud MUELLER, 2006) a comunicação científica é a infra-estrutura da comunidade científica.

É a comunicação científica que favorece ao produto (produção científica) e aos produtores (pesquisadores) a necessária visibilidade e possível credibilidade no meio social em que o produto e produtores se inserem. (TARGINO, 2000, p. 46)

TARGINO (2000) destaca algumas funções da comunicação dentro do campo da ciência: fornecer respostas a perguntas específicas; promover a atualização profissional dos pesquisadores; introduzir os pesquisadores nos campos científicos; estimular descobertas científicas; divulgar áreas científicas emergentes; legitimar novos conhecimentos; redirecionar ou ampliar as áreas de interesses dos pesquisadores; fornecer *feedback* para o aprimoramento da produção científica.

Considerando-se a ciência uma atividade social passível de análise, este trabalho realiza um estudo cienciométrico, segmento da Sociologia da Ciência ou dos Estudos Sociais

da Ciência, sobre pesquisa brasileira em Web Semântica através de suas publicações científicas.

4.2 A Web Semântica na Plataforma Lattes

Procura-se caracterizar através da identificação de Instituições de pesquisa, matrizes disciplinares, estabelecer indicadores de cooperação na comunidade científica responsável pela construção e desenvolvimento da pesquisa em Web Semântica considerando-se como representantes desta comunidade os cientistas doutores. A opção por um estudo por amostragem fundamenta-se no conceito kuhniano da necessidade de formação dos pesquisadores dentro de suas matrizes disciplinares para amadurecimento e consolidação dos paradigmas dominantes daquela ciência. Visto que a pós-graduação no nível de doutorado permite uma extensa socialização dentro da matriz disciplinar, considera-se os pesquisadores doutores amadurecidos e efetivamente praticantes dos paradigmas consolidados em sua formação. Acredita-se que o estudo deste segmento da comunidade científica pesquisadora sobre Web Semântica refletirá adequadamente a cooperação em sua atual estrutura.

A abordagem cienciométrica teve como fonte para a coleta dos dados a Plataforma Lattes desenvolvida pelo Conselho Nacional de Pesquisa (CNPq) através do Currículo Lattes dos pesquisadores doutores. A Plataforma Lattes “representa a experiência do CNPq na integração de bases de dados de currículos e de instituições da área de ciência e tecnologia em um único Sistema de Informações” (CONSELHO). Uma dessas bases de dados tem como função registrar as atividades profissionais de pesquisadores independente do grau de formação: graduação, especialização, mestrado, doutorado, pós-doutorado, inclusive estudantes ainda não graduados ou licenciados no ensino superior, trata-se do Currículo Lattes. “O Currículo Lattes registra a vida pregressa e atual dos pesquisadores sendo elemento indispensável à análise de mérito e competência dos pleitos apresentados às agências de fomento” (CONSELHO).

Optou-se pelo uso da Plataforma Lattes para identificação dos pesquisadores brasileiros pois as informações que compoem a Plataforma como currículos de pesquisadores, grupos de pesquisa e projetos em ciência e tecnologia realizados no país são públicas e possibilitam a identificação e extração de dados sobre C&T do Brasil. (BALANCIERI et al., 2005)

Foram recuperados em abril de 2007 os Currículos Lattes dos pesquisadores doutores que satisfizeram a seguinte estratégia de busca: através do formulário de Busca Avançada por Assunto foi preenchido o campo ‘Construa uma consulta com... esta frase exata’ com o termo Web Semântica (Figura 5). Segundo as instruções da Base de Dados através da estratégia adotada foram recuperados os Currículos Lattes dos pesquisadores doutores que tivessem no título ou no campo assunto de suas produções bibliográficas o termo ‘Web Semântica’.

The image shows a screenshot of the 'Plataforma Lattes' search interface. At the top left is the CNPq logo, and at the top right is the 'Plataforma Lattes' logo. Below the header is a search form titled 'Buscar pesquisadores (Busca Avançada por Assunto)'. The form has two main sections: 'Construa uma consulta com...' and 'Pesquisa com...'. In the 'Construa uma consulta com...' section, the 'esta frase exata' field is filled with 'web semântica'. There are also fields for 'todas essas palavras', 'qualquer uma dessas palavras', and 'e nenhuma dessas palavras'. The 'Pesquisa com...' section has a field for 'esta expressão booleana'. Below the form, there are options for 'Nas bases de dados de:' with 'Doutores' selected. At the bottom, there are filters for 'Aplicar filtro aos resultados por:' with 'Pesquisadores do CNPq' and 'Bolsistas do CNPq' options. A portrait of Carlos Chagas is visible on the right side of the form.

Figura 4. – Pesquisa avançada no Currículo Lattes utilizando a frase exata: “web semântica”

Foram recuperados 141 Currículos Lattes que satisfizeram a estratégia de busca. Porém após a análise destes Currículos foram excluídos 38 pesquisadores doutores que não eram autores ou co-autores de publicações bibliográficas sobre Web Semântica totalizando um universo de pesquisa de 103 doutores responsáveis pela pesquisa científica brasileira em Web Semântica.

Analisou-se os Currículos Lattes destes doutores buscando coletar dados que permitissem o mapeamento das Instituições de pesquisa nas quais eles atuam, identificação da matriz disciplinar a que pertencem, e estabelecer relações de cooperação entre estes pesquisadores.

4.3 Quem contribui para a pesquisa em Web Semântica no Brasil?

Procurou-se coletar e analisar os dados presentes nos Currículos Lattes dos pesquisadores doutores cuja produção científica incluisse a pesquisa em Web Semântica com o intuito de identificar o grau de participação da Ciência da Informação e da Ciência da Computação na construção do conhecimento científico nesta área.

A pesquisa em Web Semântica pode ser considerada como uma das áreas de fronteira entre essas duas matrizes disciplinares, pois faz uso de diversas estruturas em camadas de metadados que agregam valor semântico às páginas Web. Diversos conceitos se assemelham, em teoria, aos padrões de descrição bibliográfica e temática presentes na Ciência da Informação e têm sido estudados pelos pesquisadores da área. Sendo assim, a Recuperação de Informação na Web, principalmente o desenvolvimento da Web Semântica é um tema comum às duas matrizes disciplinares.

A seguir apresenta-se a atual estrutura da comunidade científica pesquisadora sobre Web Semântica no Brasil. Dos 103 doutores: 86 (83,49%) são da Ciência da Computação, 16 (15,54%) da Ciência da Informação, 1 (0,97%) da Linguística com interface com a Ciência da Computação (Tabela 2). Merece destaque a identificação de 4 (3,88%) pesquisadores da Ciência da Computação que possuem interface com a Ciência da Informação através do uso de seus canais de comunicação.

Tabela 1. Matriz Disciplinar dos pesquisadores

Matriz Disciplinar	Porcentagem (%)
Ciência da Computação	83,49
Ciência da Informação	15,54
Linguística	0,97

A presença de um maior número de pesquisadores dentro a Ciência da Computação era esperada visto que se trata de um paradigma gerado dentro desta matriz disciplinar. É esta matriz que assume o papel central no desenvolvimento das pesquisas. A Ciência da Informação tem papel coadjuvante ao proporcionar novas perspectivas paradigmáticas.

A presença de pesquisadores da Ciência da Computação no sistema de comunicação científica da Ciência da Informação tem caráter positivo, porém espera-se que o inverso também aconteça para que a intersecção de paradigmas possa se concretizar no âmbito nacional. Prática importante para o crescimento científico e tecnológico, as pesquisas transdisciplinares promovem a formação de redes de interação entre diferentes matrizes disciplinares que contribuem para o fortalecimento da área por elas compartilhada. Ao

analisarmos posteriormente os canais de comunicação adotados pelos pesquisadores procuramos identificar a participação da Ciência da Informação na comunicação científica da Ciência da Computação .

Sob o ponto de vista da tipologia documental das produções bibliográficas encontradas nos Currículos Lattes, os destaques são os trabalhos apresentados em eventos (72,18%). São os eventos os canais de comunicação mais rápidos para a socialização de uma pesquisa e geralmente são usados na divulgação de pesquisas em andamento ou recém concluídas (MEADOWS, 1999; MUELLER, 2005; SILVA et al. 2006). No entanto, 18,23% da produção bibliográfica científica brasileira está registrada em periódicos científicos, considerado o canal de comunicação científica mais respeitado pelos cientistas, principalmente nas ciências exatas, da saúde e sociais aplicadas. (MUELLER, 2005)

Meadows (1999) e Mueller (2005) destacam que diferentes matrizes disciplinares ou áreas do conhecimento possuem preferências por canais de comunicação distintos como consequência da natureza e especificidades de cada área que leva à adoção de práticas diferentes da atividade científica, e conseqüentemente a adoção de formas diferentes de comunicação e socialização dos conhecimentos gerados em suas pesquisas.

Tabela 2. Canais de comunicação da pesquisa em Web Semântica na Ciência da Computação

Tipo de Publicação	Porcentagem (%)
Artigo de Periódico	18,23
Capítulo de Livro	6,00
Livro	2,39
Relatório Técnico	1,20
Trabalho de Evento	72,18

Os livros (2,39%) e capítulos de livros (6%) em geral são pouco utilizados como canal de comunicação dentro da Ciência da Computação, confirmando o baixo índice de adoção destes canais que também pode ser reflexo do estágio inicial do conhecimento e da pesquisa científica em Web Semântica (MUELLER, 2005).

Dentro da matriz Ciência da Informação a opção principal como canal de comunicação também foram os trabalhos de eventos (48,58%) e os artigos de periódicos (40%), quase em igual proporção.

Tabela 3. Canais de comunicação da pesquisa em Web Semântica na Ciência da Informação

Tipo de Publicação	Porcentagem (%)
Artigo de Periódico	40,00
Capítulo de Livro	5,71
Livro	5,71
Trabalho de Evento	48,58

Destaca-se a existência da participação da Ciência da Informação nos canais de comunicação da Ciência da Computação através dos trabalhos de eventos, prática a ser incentivada para a construção de um conhecimento científico transdisciplinar em Web Semântica. O oposto também ocorreu com a publicação nos canais da Ciência da Informação de trabalhos realizados por pesquisadores da Computação.

A seguir procura-se identificar as relações de cooperação explicitadas pela prática da co-autoria de trabalhos científicos, objetivando principalmente localizar as iniciativas de cooperação entre matrizes disciplinares. O fenômeno da co-autoria ou autoria múltipla está presente em todas as matrizes disciplinares, porém com diferentes graus de adesão, visto que “a política vigente das agências de fomento também concorre para a crescente autoria múltipla, priorizando os projetos integrados de pesquisa em vez de trabalhos individuais” (TARGINO, 2005, p. 46). Outro fator que incentiva a adesão às autorias múltiplas é a maior visibilidade das pesquisas em colaboração através das citações recebidas (MEADOWS, 1999).

Os 103 pesquisadores produziram 420 publicações sobre Web Semântica que apresentam as características descritas anteriormente. As maiores incidências de parcerias estão apresentadas na tabela 4.

Tabela 4. Colaboração em co-autoria

Co-autorias	Trabalhos
BULCÃO NETO, RF; PIMENTEL, MGC	12
ALVES, IMR; CHISHMAN, RLO	10
GUIZZARDI, G; WAGNER, G	10
FALBO, RA; GUIZZARDI, G	08
FARIA, CG; GIRARDI, R	08
FALBO, RA; PEREIRA FILHO, JG	06
GIRARDI, R; LINDOSO, A06	06

As maiores incidências de autorias múltiplas estão na Ciência da Computação, pois são todas as que compõem a Tabela 4. No entanto foram identificadas duas publicações com co-autoria entre pesquisadores da Ciência da Informação e a Ciência da Computação.

Portanto, pode-se considerar que existem iniciativas nacionais objetivando a cooperação entre essas matrizes disciplinares.

Outra metodologia cienciométrica que permite identificar e analisar as influências paradigmáticas de uma pesquisa científica são os estudos de citação. A seguir procurou-se identificar quais autores mais contribuíram para a construção e desenvolvimento do conhecimento científico sobre Web Semântica.

4.4 Quem influencia a pesquisa em Web Semântica no Brasil?

Macias-Chapula (1998) destaca um trabalho de Weinstock no qual o autor lista diversas funções que uma citação bibliográfica pode desempenhar: homenagear pioneiras da área; dar crédito a trabalhos relacionados; disponibilizar literatura científica básica; identificar métodos, técnicas, equipamentos, etc; retificar o próprio trabalho, bem como o trabalho de outros; sustentar declarações; destacar trabalhos pouco conhecidos; contestar trabalhos de outros; estabelecer direito de propriedade e de prioridade sobre o tema; identificar fontes de informação. O importante a se destacar é que a literatura sobre estudos métricos da informação comprova que as citações presentes num trabalho científico refletem a influência do autor citado no trabalho citante (VANZ ; CAREGNATO, 2003).

Os estudos de citação

Nos permite mapear um campo emergente ou consolidado, identificar seus principais atores e as relações que se estabelecem entre eles e identificar uma série de características do comportamento de uso da informação recuperada. Assim, os estudos de citação constituem um importante indicador da atividade científica, pois contribuem para entender a estrutura e o desenvolvimento da ciência e também identificar as regularidades básicas de seu funcionamento. (VANZ; CAREGNATO, 2003, p. 255)

Com o intuito de identificar as influências científicas que a pesquisa brasileira em Web Semântica recebe, optou-se por analisar as citações presentes nas dissertações e teses defendidas e cadastradas nos bancos de textos completos disponibilizados pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Estabeleceu-se a mesma estratégia de busca nos dois bancos de textos, foram definidos como palavras-chaves os termos: Web Semântica, Ontologia e/ou Ontologias, RDF. A partir dos resultados obtidos realizou-se uma seleção temática visto que os termos Ontologia e

Ontologias recuperaram trabalhos não relacionados à Web Semântica. Após a seleção foram identificadas 52 dissertações ou teses sobre Web Semântica e suas tecnologias pertencentes as seguintes matrizes disciplinares:

Tabela 5. Matrizes disciplinares das Dissertações e Teses brasileiras sobre Web Semântica

Matriz Disciplinar	Porcentagem (%)
Ciência da Computação	88,47
Ciência da Informação	9,61
Letras	1,92

Mais uma vez a predominância esperada da Ciência da Computação como matriz disciplinar com a agenda mais focada nas pesquisas sobre Web Semântica. A presença da área de Letras como mais uma matriz disciplinar envolvida na pesquisa sobre o tema reforça a tese da Web Semântica como um projeto transdisciplinar que se localiza em diversas fronteiras do domínio do conhecimento.

As citações presentes nestas publicações foram analisadas isoladamente por matriz disciplinar, desta forma é possível identificar as influências que cada Ciência recebe e comparar os resultados com a intenção de identificar a existência de similaridades. Foram selecionados como autores mais citados aqueles que receberam pelo menos seis citações no total das publicações. A seguir a Tabela 6 apresenta a relação dos autores mais citados dentro da Ciência da Informação:

Tabela 6. Autores mais citados na Ciência da Informação

Autor	Porcentagem (%)
Berners-Lee, T	18,57
Guarino, N	12,86
Gruber, T	11,43
Almeida, MB	10
Alvarenga, L	10
Miller, E	10
Bax, MP	10
Gómez-Perez, A	8,57
Lassila, O	8,57

A tabela 7 apresenta a relação dos autores mais citados na Ciência da Computação:

Tabela 7. Autores mais citados na Ciência da Computação

Autor	Porcentagem (%)
Girardi, R	18,24
Staab, S	12,04
Maedche, A	9,51
Cossentino, M	7,13
Fensel, D	7,13
Guarino, N	7,13
Lindoso, A	6,35
Noy, N	6,35
Faria, C	5,54
Gruber, T	5,54
Oliveira, I	5,54
Deloach, S	4,75
Lassila, O	4,75

Ao compararmos as citações presentes na Ciência da Informação com as da Ciência da Computação de imediato verificamos uma maior concentração de citação num menor número de autores na Ciência da Informação. Comparando as tabelas 6 e 7 que identifica as matrizes disciplinares dos autores citados na Ciência da Informação e da Computação percebe-se que a Ciência da Informação cita, além dos autores de sua própria matriz, autores considerados clássicos dentro do tema Web Semântica: Berners-Lee, Guarino, Lassila, Gruber. Identificando o perfil temático destas citações nota-se a predominância das questões envolvendo a construção de Ontologias. Considerando-se que as citações representam influências dos autores citados na matriz disciplinar citante pode-se inferir que a Ciência da Informação identifica como ponto de confluência entre as Ciências as pesquisas sobre a representação e organização temática dos ambientes Web.

Ao mesmo tempo identificou-se a citação de autores na Ciência da Computação também focados na pesquisa envolvendo a construção de Ontologias e Inteligência Artificial, porém todos os autores citados pertencem a mesma matriz disciplinar. Os autores citados na Ciência da Computação têm seus trabalhos direcionados mais à pesquisa aplicada em comparação aos autores citados na Ciência da Informação. Acredita-se que esta é a característica que propiciou a citação de autores diferentes dentro do mesmo tema: Ontologias.

Tabela 8. Matriz disciplinar dos autores citados nas Dissertações e Teses sobre Web Semântica

Dissertações e Teses	Autores da CI (%)	Autores da CC (%)
Ciência da Informação	30	70
Ciência da Computação		100

Os resultados obtidos no estudo de citações se assemelham aos resultados obtidos na análise de co-autoria: são poucas as iniciativas concretas de cooperação transdisciplinar nas pesquisas sobre Web Semântica no Brasil. Aparentemente somente a Ciência da Informação identifica integrações temáticas entre as duas Ciências. Uma tentativa de explicar este fenômeno baseia-se na concepção kuhniana que a comunicação científica é abundante dentro de uma comunidade científica que compartilha os mesmos canais de comunicação, valem-se das mesmas fontes de informação e indiretamente promovem a exclusão de outras matrizes disciplinares que não compartilham destas mesmas fontes e canais.

Considera-se a Web Semântica uma área de pesquisa na fronteira da Ciência da Computação com a Ciência da Informação e outros domínios do conhecimento como a Linguística, por exemplo. A cooperação efetiva destas e outras matrizes disciplinares no desenvolvimento tecnológico da Web Semântica depende da integração das matrizes disciplinares na construção e desenvolvimento do conhecimento científico de forma transdisciplinar, através do compartilhamento de canais de comunicação e fontes de informação que permitam sua disseminação dentro de cada comunidade científica. Acredita-se ser este o caminho para a ampla divulgação das potencialidades que cada matriz tem para aprimorar essa nova configuração dos ambientes Web.

5 CONSIDERAÇÕES FINAIS

A ampliação dos espaços informacionais com o desenvolvimento da Web provocou um impacto nas pesquisas de diversas matrizes disciplinares dentre as quais podemos destacar a Ciência da Informação e a Ciência da Computação. Questões envolvendo os impactos tecnológicos na sociedade ou em comunidades específicas, a recuperação da informação na Web, a arquitetura de informação, usabilidade de sistemas de informação, têm sido objeto de estudo e análise destas e de outras matrizes disciplinares.

Pode-se considerar que as pesquisas envolvendo a Internet e a Web tem um posicionamento de fronteira dentro dos limites das matrizes disciplinares. Estes temas fronteiriços podem ser considerados como transdisciplinares permitindo a adoção e contribuição de diversos paradigmas de diferentes áreas do conhecimento.

Procurou-se neste trabalho identificar e analisar as parcerias e influências que a Ciência da Informação e a Ciência da Computação desempenham na construção e no desenvolvimento das pesquisas brasileiras sobre o projeto Web Semântica. Vale destacar que esta análise ficou comprometida em virtude da falta de instrumentos padronizados e estruturados para o desenvolvimento de estudos cienciométricos sobre C&T desenvolvida no Brasil. A iniciativa nacional da Plataforma Lattes não disponibiliza abertamente sua estrutura de metadados para que se avalie sua compatibilidade com os diversos recursos automatizados para estudos cienciométricos. Em relação às dissertações e teses produzidas pelos programas de pós-graduação os problemas enfrentados foram em relação à falta de normalização até mesmo dentro dos mesmos programas de pós-graduação. Maior ainda o problema da normalização quando os programas eram de Instituições diferentes e pertencentes a matrizes disciplinares distintas. Esta falta de padronização no registro das referências bibliográficas impediu a coleta automatizada de dados, também comprometendo uma análise mais detalhada e ampla das citações presentes nos documentos. Todos esses fatores influenciaram nas práticas adotadas para coleta e análise dos dados e apesar das dificuldades encontradas foi possível caracterizar a atual situação da pesquisa brasileira sobre o projeto Web Semântica.

A Web Semântica está posicionada numa fronteira de saberes de diversos domínios do conhecimento e nos trabalhos estudados tem recebido contribuições tanto da Ciência da Computação quanto da Ciência da Informação, além da Linguística, na construção e desenvolvimento de sua tecnologia. Porém cada matriz disciplinar praticamente isolada em

sua comunidade. Optou-se pela análise cienciométrica deste campo de pesquisa que permitiu através do estudo das publicações científicas a identificação de quais as matrizes disciplinares publicam sobre o tema, quais canais de comunicação são preferenciais em cada matriz e quais são as influências representadas em citações bibliográficas procurando estabelecer um painel do cenário atual da pesquisa sobre Web Semântica no Brasil.

A participação dos pesquisadores da Ciência da Computação em mais de 80 % da produção científica sobre o tema era um fato esperado no estudo, no entanto revelou questões importantes como um tímido intercâmbio entre as duas matrizes que pode indicar o início de processo de compartilhamento de informações através dos canais de comunicação mais utilizados em cada matriz.

Esperava-se caracterizar um ambiente de cooperação transdisciplinar no desenvolvimento científico da Web, no entanto foram identificadas ações isoladas de integração entre a Ciência da Informação e da Computação através de suas publicações científicas. A partir da observação das citações presentes nas publicações da Ciência da Informação percebe-se que esta matriz identifica esta transdisciplinaridade, ao citar autores de ambas as matrizes que divulgam suas pesquisas sobre o tema, porém tal percepção é praticamente inexistente na Ciência da Computação.

Como possíveis explicações para o atual cenário de cooperação entre as disciplinas no âmbito das pesquisas sobre Web Semântica, pode-se destacar a incompatibilidade do conceito informação para as cada uma das Ciências. Enquanto na Ciência da Informação a informação tem que necessariamente comportar um sentido e estar registrada num suporte, para a Ciência da Computação o conceito de informação está atrelado à capacidade desta ser representada através de uma teoria matemática que modele a transmissão de mensagens e as trocas de sinais presentes nesta transmissão sem qualquer preocupação com a semântica dos dados.

Outro fator que pode dificultar a cooperação entre as disciplinas é o uso de canais de comunicação e fontes de informação específicos de cada Ciência. A seleção dos canais de comunicação para a disseminação dos resultados de pesquisa interfere na capacidade de integração dos pesquisadores de diferentes matrizes disciplinares. Toda informação que não se tem acesso, inexistente para o usuário.

Mais um fator que pode contribuir para a realidade observada é o número reduzido de dissertações e teses na Ciência da Informação que tratam do assunto, cabe aos pesquisadores e

profissionais da informação melhor se familiarizarem com as novas tecnologias que permeiam a Web, possibilitando novas contribuições ao desenvolvimento deste ambiente informacional.

A integração entre comunidades científicas formadas em matrizes disciplinares diferentes que, no entanto compartilham o mesmo objeto de estudo é um processo difícil, nem sempre possível, mas que encontra um ambiente favorável nos espaços de fronteira que demarcam as matrizes disciplinares. O cenário nacional atual da pesquisa em Web Semântica pode ser início deste processo, para isso espera-se maior integração dessas comunidades científicas através da inter ou transdisciplinaridade na criação de grupos de pesquisa, no incentivo das agências de fomento e na escolha de diferentes canais de comunicação para a disseminação dos resultados das pesquisas capazes de atingir diversas matrizes disciplinares.

Sob uma perspectiva otimista pode-se comprovar o que se acredita ser o início do processo de cooperação ao observarmos a comunidade pesquisadora sobre Web Semântica no Brasil, composta por pesquisadores de diversas matrizes como a Linguística, a Ciência da Informação e a Ciência da Computação que procuram compartilhar suas pesquisas em diversos canais de comunicação, principalmente os eventos científicos e periódicos de suas e de outras matrizes disciplinares. Sob a perspectiva pessimista, este mesmo cenário pode indicar que a pesquisa brasileira sobre Web Semântica estruturada de forma cooperada entre a Ciência da Informação e a Ciência da Computação ainda depende de iniciativas isoladas de pesquisadores, tendo ainda um longo caminho a percorrer.

Apoiando-se na perspectiva otimista, espera-se que este cenário permaneça e evolua gradativa e rapidamente para um maior intercâmbio de informações entre as matrizes disciplinares, proporcionando a cooperação multiparadigmática que a Web Semântica necessita. Acredita-se que a melhor estratégia para a consolidação da pesquisa transdisciplinar e cooperada é a ampliação dos canais de comunicação da Ciência da Informação e a inserção de seus pesquisadores em grupos de pesquisa da Ciência da Computação.

Referências Bibliográficas

ALVES, R. C. V.; SANTOS, P. L. V. A. C. Web semântica: uma análise focada no uso de metadados. In: SIMPÓSIO EM FILOSOFIA E CIÊNCIA (6., Marília, 2005) **Universidade e contemporaneidade** : produção do conhecimento e formação profissional. 1 CD-ROM.

ARAÚJO, C. A. A. A ciência da informação como ciência social. **Ci. Inf.**, Brasília, v. 32, n. 3, set./dez. 2003a. Disponível em:
http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652003000300003&lng=pt&nrm=iso&tlng=pt Acesso em: 18 de agosto de 2006.

BALANCIERI, R. et al. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo da Plataforma Lattes. **Ci. Inf.**, Brasília, v. 34, n. 1, p. 64-77, jan./abr. 2005.

BERNERS-LEE, T.; HANDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, may 2001. Disponível em:
<http://www.scientificamerican.com/2001/0501issueberners-lee.html> Acesso em: 27 de novembro de 2004.

BERNERS-LEE, T. **Semantic web concepts**. Disponível em:
<http://www.w3.org/2005/Talks/0517-boit-tbl/> Acesso em: 25 de outubro de 2006.

BROWN, R. H. Modern Science: institutionalization of knowledge and rationalization of power. **The Sociological Quarterly**, v. 34, n. 1, p. 153-168, 1993.

CAMPOS, M. L. M; CAMPOS, M. L. A; CAMPOS, L. M. Web semântica e a gestão de conteúdos informacionais. In: MARCONDES, C. H et al. (orgs.) **Bibliotecas digitais: saberes e práticas**. 2. ed. Salvador/Brasília: UFBA/IBICT, 2006. p. 55-74.

CENDÓN, B. V. Ferramentas de busca na web. **Ci. Inf.**, Brasília, v. 30, n. 1, p. 39-49, jan./abr. 2001.

CODINA, L.; ROVIRA, C. La web semântica. In: TRAMULLAS, J. (coord.) **Tendências em documentación digital**. Gijon: Trea, 2006. p. 09-54.

CONSELHO NACIONAL DE PESQUISA. **Plataforma Lattes**. Disponível em:
http://lattes.cnpq.br/conheca/con_apres.htm Acesso em: 09 de abril de 2007.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. **Portal periódicos CAPES**. Disponível em:
<http://www.periodicos.capes.gov.br/portugues/index.jsp> Acesso em 06 de outubro de 2007.

DENCKER, A. F. M. Paradigmas científicos e ensino superior. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 24., 2001, Campo Grande. **Anais...** São Paulo: Intercom, 2001. CD-ROM. Disponível em:

<http://reposcom.portcom.intercom.org.br/handle/1904/5175> Acesso em: 08 de janeiro de 2007.

DENNING, P.J. et al. Computing as a discipline. **Communication of the ACM**, v. 32, n. 1, p. 9-23, 1989

EGGHE, L. Bridging the gaps: conceptual discussions on informetrics. **Scientometrics**, v. 30, n. 1, p. 35-47, 1994.

EUGÊNIO, M.; FRANÇA, R.O.; PEREZ, R. C. Ciência da informação sob a ótica paradigmática de Thomas Kuhn: elementos de reflexão. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, p. 27-39, jan./jun. 1996.

FEITOSA, A. **Organização da informação na Web**. Brasília: Thesaurus, 2006

FERNEDA, E. **Recuperação da informação**: análise sobre a contribuição da ciência da computação para a ciência da informação. 2003.137 f. Tese (Doutorado em Ciência da Comunicação). Escola de Comunicações e Artes - USP, São Paulo, 2003.

FLECK, L. **Genesis and development of a scientific fact**. Chicago: University of Chicago Press, 1979.

FONSECA FILHO, C. **História da computação**: teoria e tecnologia. São Paulo: LTr, 1999.

FURGERI, S. **Representação de informação e conhecimento**: estudo das diferentes abordagens entre a ciência da informação e a ciência da computação. 2006. 161 f. Dissertação (Mestrado em Ciência da Informação). Pontifícia Universidade Católica de Campinas, Campinas, 2006.

GRIFFITH, B. C. Understanding science: studies of communication and information. **Communication Research**, v. 16, n. 5, p. 600-614, 1989.

HAYTHORNTHWAITE, C. Learning and knowledge networks in interdisciplinary collaborations. **JASIST**, v. 57, n. 8, p. 1079-1092, 2006.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA.

Biblioteca Digital de Teses e Dissertações Disponível em:

<http://btdt.ibict.br/utilitarios/sobre.jsp> Acesso em 06 de outubro de 2007.

KOBASHI, N.Y.; SANTOS, R. N.M. Arqueologia do trabalho imaterial: uma aplicação bibliométrica à análise de dissertações e teses. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (7.: Marília, 2006) **A dimensão epistemológica da Ciência da Informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação**. Disponível em: <http://www.portalppgci.marilia.unesp.br/enancib/viewpaper.php?id=263> Acesso em: 07 de setembro de 2007.

KUHN, T. **A tensão essencial**. Lisboa: Edições 70, 1989.

KUHN, T. **A estrutura das revoluções científicas**. São Paulo: Perspectiva, 2003.

LARA, M. L. G. Termos e conceitos da área de comunicação e produção científica. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M (orgs.) **Comunicação & produção científica: contexto, indicadores e avaliação**. São Paulo: Angellara, 2006. p. 387-416.

LE COADIC, Y.-F. **A ciência da informação**. 2. ed. Brasília: Briquet de Lemos, 2004.

LIMA, R. A; VELHO, L. M. S.; FARIA, L. I. Delimitação de uma área multidisciplinar para análise bibliométrica de produção científica: o caso da Bioprospecção. **Transinformação**, v. 19, n. 2, p. 153-168, maio/ago. 2007.

LIMA-MARQUES, M. **Ontologias: da filosofia à representação do conhecimento**. Brasília: Thesaurus, 2006

LÓPEZ YEPES, J. El análisis cualitativo de citas como instrumento para el estudio de la creación y transmisión de las ideas científicas. **Documentación de las Ciencias de la Información**, v. 26, p. 41-70, 2003.

LOWY, I. Ludwick Fleck e a presente história das ciências. **Manguinhos**, v.I, n. 1, p. 7-18, 1994.

LYMAN, P. Digital documents and the future of the academic community. In: CONFERENCE ON SCHOLARLY COMMUNICATION AND TECHNOLOGY, 1997.

MACHADO, C. J. S.; TEIXEIRA, M. O. Descoberta, invenção e inovação segundo os estudos sociais anglo-saxões e europeus das ciências. **DataGramaZero**, v. 8, n. 2, abr. 2007. Disponível em: http://www.dgz.org.br/abr07/F_I_aut.htm Acesso em: 06 de abril de 2007.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ci. Inf.**, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998.

MASTERMAN, M. A natureza de um paradigma. In: LAKATOS, I.; MUSGRAVE, A. (orgs.) **A crítica e o desenvolvimento do conhecimento**. São Paulo: Cultrix, 1974. p.59-90.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MENDEZ RODRIGUES, E. **Metadados y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijon: Trea, 2002.

MERTON, R. K. Os imperativos institucionais da ciência. In: DEUS, J. D. **A crítica da ciência**. Rio de Janeiro: Zahar, 1979. p. 37-52

MOOERS, C. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, v.2, n.1, p. 20-32, 1951.

MUELLER, S. P. M. O crescimento da ciência, o comportamento científico e a comunicação científica: algumas reflexões. **R. Esc. Biblioteconomia UFMG**, Belo Horizonte, v. 24, n. 1, p. 63-84, jan./jun. 1995.

MUELLER, S. P. M. A publicação da ciência: áreas científicas e seus canais preferenciais. **DataGramZero**, v.6, n. 1, fev. 2005. Disponível em: http://www.datagramazero.org.br/fev05/F_I_art.htm Acesso em: 13 de março de 2005.

MUELLER, S. P. M. A comunicação científica e o movimento de cesso livre ao conhecimento. **Ci. Inf.**, v. 35, n. 2, p. 27-38, 2006.

MUGNAINI, R.; CARVALHO, T.; CAMPANATTI-OSTIZ, H. Indicadores de produção científica: uma discussão conceitual. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M (orgs.) **Comunicação & produção científica**: contexto, indicadores e avaliação. São Paulo: Angellara, 2006. p.313-340.

NAFRIA, I. **El futuro de Internet tiene nombre**: la web semântica. 23/05/2001. Disponível em: <http://www.baquia.com/com/20010523/art00008.html>

NEHMY, R. M. Q. et al. A ciência da informação como disciplina científica. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n.1, p. 9-25, jan./jun. 1996.

NUNES, D. J. Projetos de planos pedagógicos orientados a problemas. **Sociedade Brasileira de Computação**, 21 jun.2004. Disponível em : http://143.54.83.4/ArquivosComunicacao/Artigo_Plano_Pedagogico_daltro.pdf Acesso em: 25 de outubro de 2006.

OLIVEIRA, M.; MOTA, F. R. L.; ALVARADO, R. U. Comunidade científica e cientificidade da ciência da informação. In: CONGRESSO DA BAD, 8., 2004, Lisboa. Disponível em: <http://sapp.telepac.pt/apbad/congresso8/com27.pdf> Acesso em: 21 de janeiro de 2005.

PINHEIRO, C. B. F.; VIDOTTI, S. A. B. G. A pesquisa em Web Semântica : análise métrica do seu desenvolvimento. In: SIMPÓSIO EM FILOSOFIA E CIÊNCIA (6., Marília, 2005) **Universidade e contemporaneidade** : produção do conhecimento e formação profissional. 1 CD-ROM.

PINHEIRO, L. V. R.; LOURENÇO, J. M. M. Traçados e limites da ciência da informação. **Ci. Inf.**, Brasília, v. 24, n. 1, jan./abr. 1995. Disponível em: <http://www.ibict.br/cienciadainformacao/viewarticle.php?id=576&layout=abstract> Acesso em: 18 de agosto de 2006.

POMBO, O. Contribuição para um vocabulário sobre interdisciplinaridade. In: POMBO, O.; LEVY, T.; GUIMARÃES, H. **A interdisciplinaridade**: reflexão e experiência. Lisboa: Texto, 1993. Disponível em: <http://www.educ.fc.ul.pt/docentes/opombo/mathesis/vocabulario-interd.pdf> Acesso em 09 de março de 2008.

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

RAMALHO, R. A. S.; VIDOTTI, S. A. B. .G.; FUJITA, M. S. L. Web semântica: uma investigação sob o olhar da Ciência da Informação. **DataGramaZero**, v.8, n. 6, dez. 2007. Disponível em: http://www.dgz.org.br/dez07/F_I_art.htm Acesso em: 21 de janeiro de 2008.

SANTAREM SEGUNDO, J. E. **Recursos tecno-metodol[ogicos para 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.

SANTOS, R. N. M. Indicadores estratégicos em ciência e tecnologia: refletindo a sua prática como dispositivo de inclusão/exclusão. **Transinformação**, v. 15, Ed. Esp., p. 129-140, set./dez. 2003.

SARACEVIC, T. A natureza interdisciplinar da ciência da informação. . **Ci. Inf.**, Brasília, v. 24, n. 1, jan./abr. 1995. Disponível em: <http://www.ibict.br/cienciadainformacao/viewarticle.php?id=575&layout=abstract> Acesso em 18 de agosto de 2006.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n.1, p. 41-62, jan./jun. 1996.

SAYÃO, L. F. Modelos teóricos em ciência da informação – abstração e método científico. **Ci. Inf.**, Brasília, v. 30, n. 1, p. 82-91, jan./abr. 2001.

SENSO RUIZ, J. A. Sistemas de metadados para la web semântica. In: TRAMULLAS, J. (coord.) **Tendências em documentación digital**. Gijon: Trea, 2006. p. 55-80.

SILVA, E.L. et al. Panorama da pesquisa em ciência da informação no Brasil. **Inf. & Soc.: Est**, v. 16, n. 1, p. 205-228, 2006.

TARGINO, M. G. Comunicação científica: uma revisão de seus elementos básicos. **Inf. & Soc.: Est**, v. 10, n. 2, p. 37-85, 2000.

TARGINO, M. G. Artigos científicos: a saga da autoria e co-autoria. In: FERREIRA, S. M. S. P.; TARGINO, M.G. **Preparação de revistas científicas: teoria e prática**. São Paulo: Reichmann, 2005. p.35-54

VALENTIM, M. L. P. Construção do conhecimento científico. In: VALENTIM, M. L. P.(org.) **Métodos qualitativos de pesquisa em ciência da informação**. São Paulo: Polis, 2005. p. 07-28.

VANZ, S. A. S.; CAREGNATO, S. E. Estudos de citação: uma ferramenta para entender a comunicação científica. **Em Questão**, v. 9, n. 2, p. 295-307, 2003.

VELHO, L. The meaning of citation in the context of a scientifically peripheral country. **Scientometrics**, v. 9, n. 1-2, p. 71-89, 1986.

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

WERSIG, G. Information science: the study of postmodern knowledge. **Information Processing & Management**, v. 29, n.2, p. 229-239, 1993.

WHITLEY, R. Cognitive and social institutionalization of scientific specialities and research areas. In: WHITLEY, R (ed.) **Social processes of scientific development**. London: Routledge and Kegan, 1974, p. 69-95.