



Fábio Augusto Guimarães Teixeira

A Recuperação da Informação e a colaboração de usuários na Web

Brasília

2010

Fábio Augusto Guimarães Teixeira

A Recuperação da Informação e a colaboração de usuários na Web

Dissertação apresentada à Faculdade de
Ciência da Informação da Universidade de
Brasília, em 19 de julho de 2010, como
requisito parcial para a obtenção do título
de Mestre

Orientador: Prof. Dr. Cláudio Gottschalg Duque

Brasília

2010

Fábio Augusto Guimarães Teixeira

A Recuperação da Informação e a colaboração de usuários na Web

Dissertação apresentada à Faculdade de
Ciência da Informação da Universidade de
Brasília, em 19 de julho de 2010, como
requisito parcial para a obtenção do título
de Mestre

Data de aprovação:

Banca examinadora:

Prof. Dr. Cláudio Gottschalg Duque
UnB/PPGCInf

Prof. Dr. Jaime Robredo
UnB/PPGCInf

Profa. Dra. Fernanda Lima
UnB/CIC

Brasília
2010

*Dedicado ao meu sobrinho João Augusto, recém-nascido
em um mundo onde a informação, cada vez mais presente,
só tem real valor quando usada para o bem do próximo*

Agradecimentos

A Deus, que não apenas me deu o dom da vida, mas todos os dias me dispõe de força, saúde e inteligência para procurar fazer dela um instrumento de boa vontade.

À minha mãe, meu exemplo, motivação para minhas lutas e minhas vitórias, apoio para meus dias mais difíceis.

À minha irmã e meu cunhado, meus melhores amigos, pela preocupação e carinho sempre à disposição.

À minha família, com quem aprendi meus valores mais importantes.

Ao Prof. Dr. Cláudio Gottschalg Duque, pela paciência e orientação.

Aos professores e colegas do CID, com quem aprendi a expandir meus horizontes e direcionar minhas inquietações.

Aos meus amigos em Brasília (e arredores) e Belém, pela simples e tão necessária amizade e pelo companheirismo.

A todos os que, direta ou indiretamente, contribuíram para a realização deste trabalho.

“Não há fatos, só interpretações”

Friedrich Nietzsche

Resumo

Investiga a existência de características da Recuperação da Informação em iniciativas de colaboração de usuários na Web como forma de sugerir novos desenvolvimentos para a área. Estuda definições de Recuperação da Informação e de sistemas, contextualizando-os com o maior valor conferido à informação na sociedade atual, caracterizada pelo advento de Tecnologias de Informação e Comunicação, notadamente a Internet e a Web. Define tópicos que representam objetivos e preocupações da Recuperação da Informação para pesquisa em artigos recuperados na base *Web of Science*, com indexação contendo as palavras-chave “Web”, “user” e “collaboration” e publicados nos últimos cinco anos em áreas previamente selecionadas para o estudo de iniciativas de colaboração de usuários na Web. Examina, dentre os artigos recuperados, aqueles pertinentes ao escopo da pesquisa e com conteúdo acessível, verificando que todos apresentaram características que os enquadram como de interesse para a Recuperação da Informação. Discute a aplicação e o potencial uso de outros tópicos relacionados à Recuperação da Informação nos artigos analisados e em novos estudos. Indica possíveis fontes de interesse e inovação, não identificadas antecipadamente como características tradicionais, para a pesquisa na área a partir da descrição das iniciativas de colaboração de usuários na Web nos artigos analisados.

Palavras-chave: Recuperação da Informação. Web. Web 2.0. Colaboração de usuários na Web. *Web of Science*. Tecnologias da Informação e Comunicação. Sociedade da Informação. Sistemas de Recuperação da Informação. Redes sociais.

Abstract

The existence of Information Retrieval characteristics in users' collaboration on the Web research is investigated as a mean to propose new developments to the field. Information Retrieval and systems definitions are reviewed and contextualized with the greater value given to information in current society, characterized by the emergence of Information and Communication Technologies, especially the Internet and the Web. Topics representing Information Retrieval objectives and concerns are defined in order to analyze articles retrieved from the Web of Science database, indexed with "Web", "user", and "collaboration" as keywords, and published in the last five years in previously selected fields approaching users' collaboration on the Web research. Articles relevant to the study scope and with accessible content were examined, finding that all presented characteristics that classified them as of interest to Information Retrieval. The application and potential use of other topics related to Information Retrieval in the articles analyzed and in new research are discussed. Possible sources of interest and innovation for investigations in the field, not previously identified as traditional characteristics, are indicated based on the description of users' collaboration on the Web initiatives in the articles analyzed.

Keywords: Information Retrieval. Web. Web 2.0. Users' collaboration on the Web. Web of Science. Information and Communication Technologies. Information Society. Information Retrieval systems. Social networks.

Lista de figuras

Figura 1 - A estrutura em laço de documentos na Web (MANNING; RAGHAVAN; SCHÜTZE, 2009, p. 427)	74
Figura 2 - Página da ferramenta ISI <i>Web of Knowledge</i> para busca na base de dados <i>Web of Science</i>	95
Figura 3 - Parâmetros utilizados para busca na base de dados <i>Web of Science</i>	97
Figura 4 - Publicações por ano dos artigos recuperados na base <i>Web of Science</i> (THOMSON REUTERS, 2009).....	103
Figura 5 - Citações por ano dos artigos recuperados na base <i>Web of Science</i> (THOMSON REUTERS, 2009).....	103
Figura 6 - Situação de análise dos artigos recuperados na base <i>Web of Science</i> ..	110
Figura 7 - Quantidade de artigos analisados por tipo.....	111
Figura 8 - Características de RI do grupo A encontradas nos artigos analisados ...	112
Figura 9 - Características de RI do grupo B encontradas nos artigos analisados ...	114
Figura 10 - Características de RI do grupo C encontradas nos artigos analisados.	116
Figura 11 - Características de RI do grupo D encontradas nos artigos analisados.	118
Figura 12 - Características de RI do grupo E encontradas nos artigos analisados.	120
Figura 13 - Média de citações por grupo nos artigos analisados	121
Figura 14 - Quantidade de características de RI encontradas por artigo analisado	123

Lista de quadros

Quadro 1 - Recuperação de Dados ou Recuperação da Informação? (VAN RIJSBERGEN, 1979, p. 2, tradução nossa).....	23
Quadro 2 - Linha de tempo dos navegadores Web (Adaptado de ADEVERIA, 2010)	54
Quadro 3 - Linha de tempo de evolução da Web e do W3C (Adaptado de WORLD WIDE WEB CONSORTIUM, 2004c)	57
Quadro 4 - Exemplo de registro de informação em formato XML	75
Quadro 5 - Características de RI localizadas nos artigo analisados	108
Quadro 6 - Características de RI do grupo A por artigo analisado	112
Quadro 7 - Características de RI do grupo B por artigo analisado	114
Quadro 8 - Características de RI do grupo C por artigo analisado.....	116
Quadro 9 - Características de RI do grupo D por artigo analisado.....	118
Quadro 10 - Características de RI do grupo E por artigo analisado	119
Quadro 11 - Pontos de interesse para a RI encontrados nos artigos analisados....	125

Lista de tabelas

Tabela 1 - Áreas de assuntos de artigos recuperados na base <i>Web of Science</i>	98
Tabela 2 - Artigos recuperados na base <i>Web of Science</i> , organizados pelo número de citações	100
Tabela 3 - Quantidade de artigos recuperados na base <i>Web of Science</i> de acordo com o país de publicação.....	142
Tabela 4 - Quantidade de artigos recuperados na base <i>Web of Science</i> de acordo com o tipo de publicação.....	143
Tabela 5 - Quantidade de artigos recuperados na base <i>Web of Science</i> de acordo com o idioma.....	143
Tabela 6 - Quantidade de artigos recuperados na base <i>Web of Science</i> de acordo com o ano de publicação	143
Tabela 7 - Quantidade de artigos recuperados na base <i>Web of Science</i> de acordo com as principais fontes de publicação.....	143
Tabela 8 - Quantidade de artigos recuperados na base <i>Web of Science</i> de acordo com as áreas de assunto	144
Tabela 9 - Quantidade de citações por ano dos artigos recuperados na base <i>Web of Science</i>	145

Sumário

1	Introdução	12
1.1	<i>Objetivos</i>	13
1.2	<i>Justificativas</i>	14
1.3	<i>A pesquisa</i>	15
2	A Recuperação da Informação	19
2.1	<i>Sistemas de Recuperação da Informação, avaliação e relevância</i>	25
2.2	<i>A sociedade da informação e a Recuperação da Informação</i>	33
3	A Web	47
3.1	<i>A Web social</i>	58
3.2	<i>Recuperação da Informação na Web</i>	67
4	Colaboração de usuários na Web	78
4.1	<i>Tecnologias de colaboração entre usuários na Web</i>	84
5	A pesquisa por iniciativas de colaboração de usuários na Web	93
5.1	<i>A base de dados Web of Science</i>	93
5.2	<i>Parâmetros utilizados para pesquisa</i>	95
6	Discussão	105
7	Conclusão	129
7.1	<i>Sugestões de trabalhos futuros</i>	133
	Referências	134
	APÊNDICE A - Características do conjunto de artigos recuperados na base <i>Web of Science</i>	142
	APÊNDICE B - Análise individual dos artigos recuperados na base <i>Web of Science</i>	146

1 Introdução

Mais do que experimentação científica ou uma rotina organizacional, recuperar informação se tornou uma necessidade cotidiana. Na sociedade humana atual é notável a facilidade para difusão e acesso à informação, motivando o acréscimo do valor dado à utilidade da informação. Este fato serve de base para a presença no dia a dia das pessoas de um ciclo infundável que pode ser descrito em algumas etapas: a descoberta da necessidade de informação para responder a uma dúvida ou à exigência para uma atividade específica; a busca pela informação de forma que ela seja suficiente para responder a dúvida anterior; a geração de novo conhecimento e/ou novas dúvidas a partir da assimilação da informação obtida anteriormente; e então tornar disponível este novo conhecimento para outras pessoas.

Esse ciclo não surgiu recentemente, e se encontra descrito de formas diversas em estudos de Recuperação da Informação (RI) em diferentes épocas. Mas se torna cada vez mais patente na vida de um número cada vez maior de pessoas, não apenas se resumindo a grupos científicos ou acadêmicos. É uma revolução cuja responsabilidade principal pode ser atribuída a duas ferramentas que modificam a forma como o homem vive em sociedade: a Internet e a Web, o que delimita um vínculo indissociável entre estas últimas e a RI.

Como conseqüência dessa relação, a visão de que o interesse em estudos de RI deveria estar concentrado em grupos de bibliotecários e especialistas em informação está sendo abandonada (BAEZA-YATES; RIBEIRO-NETO, 1999). A propagação de diversas ferramentas de busca na Web e a constante sensação de frustração ao buscar conteúdo naquele ambiente, seja pela não relevância dos resultados retornados ou mesmo pela grande quantidade deles, acabam por ressaltar a importância da RI.

Recentemente na Web, novas ferramentas têm surgido com o intuito de facilitar a colaboração entre seus usuários para a divulgação de informação e conhecimento. Essas aplicações ou ferramentas colaborativas são chamadas comercialmente de aplicações Web 2.0 (O'REILLY, 2005). Da mesma forma como a Internet e a Web apresentam correlação com a RI, é possível supor que aplicações colaborativas baseadas na Web apresentam não apenas características de interesse

daquela área de estudo, mas também desafios já abordados pela RI em seus experimentos. Entretanto, esta relação entre aplicações colaborativas na Web e a RI não se encontra satisfatoriamente descrita em publicações científicas de forma a corroborar ou mesmo refutar tais suposições, motivando o desenvolvimento desta pesquisa. Define-se, desta forma, não apenas a temática de atuação deste estudo, a RI em ferramentas colaborativas na Web, mas também o problema não abordado por outros estudos: quais as características de interesse da RI presentes na colaboração de usuários na Web?

1.1 Objetivos

Objetivo geral

- Identificar a existência de conceitos da RI em estudos de colaboração de usuários na Web, definindo-os como de interesse para a área e com características que representam potencial desenvolvimento para a pesquisa em RI.

Objetivos específicos

- Pesquisar sobre a RI, utilizando seus conceitos para estudar os reflexos do advento de novas tecnologias e o valor da informação na sociedade atual;
- Estudar pesquisas em RI influenciadas pelo advento da Web e por características da sociedade da informação;
- Descrever algumas características da colaboração de usuários na Web e tecnologias que a utilizam; e
- Definir características conceituais e de interesse da RI para a análise de iniciativas de colaboração de usuários na Web.

1.2 *Justificativas*

A RI se ocupa com “os aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregados para o desempenho da operação” (MOOERS, 1951 apud SARACEVIC, 1996, p. 44). Com o advento da Web, o interesse em pesquisa na área se torna cada vez maior, pois a facilidade de compartilhamento de informação de uma maneira tão livre quanto descontrolada a tornou um repositório universal de conhecimento. Localizar informação efetivamente relevante para o contexto do usuário, aqui classificado como aquele que detém necessidades de informação, se tornou uma tarefa árdua (BAEZA-YATES; RIBEIRO-NETO, 1999).

Em tempos mais recentes surgiu uma tendência de desenvolvimento de aplicações que empregam a colaboração de usuários na construção de conhecimento na Web. Segundo Berners-Lee (2000), o compartilhamento de informação e a colaboração de usuários são os objetivos principais pelos quais a Web foi concebida. Agrupados em comunidades virtuais, os usuários de serviços colaborativos na Web se tornam ao mesmo tempo fonte e público-alvo de informação, podendo vir a influenciar tanto a descrição como a efetiva recuperação da informação, assim como a avaliação dos resultados de suas buscas. Com isso é possível supor que da mesma forma como corrobora para a criação de mais conteúdo, a colaboração entre usuários encerra características que podem facilitar a execução de tarefas de RI.

É necessário utilizar conceitos tradicionais da RI para primeiro compreendê-la e depois relacionar ao seu escopo de estudo o fenômeno que caracteriza a sociedade atual, de valorização da informação como vantagem competitiva para o desenvolvimento humano, notadamente influenciado pelo advento da Web. A colaboração de usuários na Web para criação e compartilhamento de informação e conhecimento surge como um novo fator que influencia não apenas a Web, mas também a própria relação desta com a RI. A abordagem de RI por iniciativas de colaboração de usuários na Web permitiria propor inovações para o emprego e o desenvolvimento da RI.

1.3 A pesquisa

Considerando a pesquisa como um processo de investigação do problema proposto, é a partir deste e dos objetivos do estudo que deve ser definido o método científico a ser empregado a fim de obter resultados que sejam passíveis de análise e generalização (CERVO; BERVIAN; DA SILVA, 2007). Segundo estes autores, o método científico define quem são os envolvidos na pesquisa, quando e onde ela deve ser realizada e principalmente a forma como deve ser feita. A pesquisa se torna composta, portanto, pela dúvida ou problema, pelo método e pelos resultados obtidos e as possíveis soluções que estes proporcionam a partir das discussões e inferências dos resultados.

No caso específico do problema a que esta pesquisa se dispõe a estudar (quais as características de interesse da RI presentes na colaboração de usuários na Web?) e dos objetivos propostos, observa-se a necessidade de seu enquadramento como sendo em nível de estudo descritivo e com base lógica indutiva, conforme classificação de Viegas (2007).

A obtenção de premissas e a generalização de propriedades comuns como base para posterior confrontação entre estes elementos e os indícios efetivamente obtidos na pesquisa, de forma a permitir a discussão e a conclusão advindas da investigação científica, são características do processo indutivo de fundamentação lógica (CERVO; BERVIAN; DA SILVA, 2007). Os autores assim descrevem a indução científica: “processo que generaliza a relação de causalidade entre dois fatos ou fenômenos” (p. 45). Marconi e Lakatos (2007) afirmam que o processo indutivo aponta o caminho para conclusões prováveis e não definitivamente verdadeiras, sendo composto pelas seguintes etapas: observação dos fenômenos, descoberta de relação entre eles e generalização da relação.

Sobre a pesquisa e o estudo descritivos, é possível citar CerVO, Bervian e Da Silva (2007, p. 61-62):

A pesquisa descritiva observa, registra, analisa e correlaciona fatos ou fenômenos (variáveis) sem manipulá-los. Procura descobrir, com a maior precisão possível, a frequência com que um fenômeno ocorre, sua relação e conexão com outros, sua natureza e suas características. Busca conhecer as diversas situações e relações que ocorrem na vida social, política, econômica e demais aspectos do comportamento humano, tanto do indivíduo tomado isoladamente como de grupos e comunidades mais complexas.[...]

a) *Estudos descritivos*: trata-se do estudo e da descrição das características, propriedades ou relações existentes na comunidade, grupo ou realidade pesquisada. [...]

Desta forma, a pesquisa descritiva ocorreu com a análise de conceitos de RI, da Web e das ferramentas de colaboração de usuários na Web a partir de estudo do referencial teórico a respeito dos três assuntos. Inicialmente, no capítulo 2, houve o estudo de referencial teórico a respeito de definições tradicionais da RI. Os conceitos pertinentes à área foram contextualizados com a identificação das necessidades informacionais das pessoas, em virtude do maior valor empregado à informação na sociedade atual, caracterizada pelas novas tecnologias de informação e comunicação. A Web, considerada até o momento como o maior exemplo do advento dessas novas tecnologias, foi alvo de análise no terceiro capítulo. De acordo com os objetivos da criação e as características (benefícios, potencialidades e riscos) tanto da Web quanto de serviços de busca lá desenvolvidos, foram relacionados impactos da existência do ambiente sobre a sociedade e sobre a RI. As ferramentas de colaboração na Web foram abordadas no capítulo 4 através da observação dos vínculos dos serviços que desempenham para seus usuários e dos seus respectivos efeitos sobre a RI, descrevendo tecnologias mais frequentemente empregadas.

Uma vez descritos fatos e/ou fenômenos pertinentes a estes tópicos, foram definidas generalizações que relacionam a RI tradicional e o seu papel com a Web. Posteriormente, foi verificado em iniciativas de colaboração de usuários na Web, trabalhos em publicações acadêmicas para as quais foi realizada consulta em base de dados específica, se houve alguma abordagem às caracterizações inicialmente obtidas.

A base de dados utilizada foi a *Web of Science*, integrante da ferramenta *ISI Web of Knowledge*, criada pelo *Institute for Scientific Information* e adquirida pelo grupo *Thomson Reuters*, sendo um dos maiores e mais compreensivos serviços de indexação de publicações acadêmicas disponíveis atualmente. Ela compreende de forma completa publicações de cerca de 9.760 periódicos científicos (além de publicações selecionadas individualmente de cerca de 10.100 periódicos não completamente indexados) em 150 disciplinas científicas, 50 ciências sociais e em disciplinas de artes e humanidades (THOMSON REUTERS, 2009). O fato de que a base de dados abrange prioritariamente publicações em língua inglesa não é

considerada prejudicial pelo fato de que a maior parte do material disponível e de interesse no contexto da pesquisa se encontra publicada naquela língua.

A pesquisa foi executada em março de 2010 utilizando as palavras-chave “Web”, “user” e “collaboration” em publicações realizadas nos cinco anos anteriores e classificadas como sendo das seguintes áreas:

- Ciência da Computação, Sistemas de Informação (*Computer Science, Information Systems*);
- Ciência da Computação, Teoria e Métodos (*Computer Science, Theory & Methods*);
- Ciência da Computação, Engenharia de Software (*Computer Science, Software Engineering*);
- Ciência da Informação e Biblioteconomia (*Information Science & Library Science*);
- Engenharia, Elétrica e Eletrônica (*Engineering, Electrical & Electronic*);
- Telecomunicações (*Telecommunications*);
- Ciência da Computação, Aplicações Interdisciplinares (*Computer Science, Interdisciplinary Applications*);
- Pesquisa Operacional e Ciência de Gerenciamento (*Operations Research & Management Science*);
- Ciências Sociais, Interdisciplinar (*Social Sciences, Interdisciplinary*);
- Gerenciamento (*Management*); e
- Comunicação (*Communication*).

O intervalo de tempo de publicação foi estabelecido com o intuito de restringir o quantitativo de contribuições analisadas visando à manutenção da viabilidade temporal da pesquisa e também considerando o amadurecimento das tecnologias de colaboração no período abordado. As áreas escolhidas são as consideradas pertinentes ao desenvolvimento de aplicações Web, enfoque do estudo.

A partir da análise das publicações acadêmicas e uma vez evidenciadas as relações das ferramentas de colaboração na Web com a RI, foram descritos no capítulo 6, destinado às discussões da pesquisa, os vínculos entre as generalizações obtidas na descrição do referencial teórico e as características das iniciativas de colaboração com o intuito de discutir sobre benefícios, potencialidades

e dificuldades do emprego destas ferramentas para o desenvolvimento da área de RI na sociedade de informação. O legado final desta pesquisa é um panorama proposto com a relação entre a colaboração de usuários na Web e a RI, apresentado no sétimo e último capítulo, servindo de base para pesquisas futuras.

2 A Recuperação da Informação

No artigo “As we may think” de Vannevar Bush publicado na edição de julho de 1945 da revista *The Atlantic Monthly* (atualmente chamada apenas de *The Atlantic*) reside o que muitos acreditam ser o primeiro pensamento em direção ao estabelecimento da Ciência da Informação (CI). Segundo o editor do artigo, a opinião do autor poderia ser sintetizada com o apelo de que “os cientistas deveriam se voltar para a tarefa massiva de tornar mais acessível o nosso desconcertante estoque de conhecimento” (BUSH, 1945, tradução nossa).

O uso da palavra “desconcertante” representava a angústia de lidar com a grande quantidade de conhecimento em constante geração até então pela humanidade (a chamada “explosão da informação”) de forma que ela pudesse ser não apenas controlável, mas, principalmente, útil. No contexto em que houve a publicação do artigo, o processo de criação do conhecimento havia sido acelerado pelo desenrolar da Segunda Guerra Mundial, já que as necessidades de desenvolvimento tecnológico da guerra favoreceram o compartilhamento de informação entre cientistas não só de diversas nacionalidades, mas de diferentes áreas.

“Um registro, se for para a utilidade da ciência, deve ser continuamente estendido, deve ser armazenado e acima de tudo ele deve ser consultado” (BUSH, 1945, tradução nossa). Ao mesmo tempo em que incitava a criação de uma ciência preocupada com o estudo da informação, Bush determinava o interesse dela pelo armazenamento e recuperação da informação. Uma vez aproximados os tempos de paz, surgia a oportunidade de registro e emprego destes avanços em conhecimento para fins pacíficos e voltados para o desenvolvimento humano. Além disso, tornavam-se patentes os indícios benéficos do compartilhamento e da colaboração como vetores do progresso da ciência.

Nos dias atuais, é possível sugerir que a preocupação com a grande quantidade de informação disponível e o seu uso para fins benéficos não só permanece, mas se intensifica cada vez mais. Primeiramente porque não existe previsão para que se encerre o processo de geração de conhecimento humano. E em segundo lugar pelo constante advento de novas tecnologias de informação e comunicação, que permitem não apenas compartilhar conhecimento, mas também a

aproximação entre pessoas em lugares distintos, com o intuito de colaborar para obtenção de novo conhecimento. Existem ainda diversos outros fatores que explicam este fenômeno, mas que aqui não serão abordados por não estarem diretamente ligados ao âmbito desta pesquisa.

Definindo a área de atuação da CI, Borko (1968, p. 4, tradução nossa) descreve a essência de sua pesquisa como a investigação das “propriedades e o comportamento da informação, o uso e a transmissão da informação e o processamento de informação para armazenamento e recuperação ótimas”. Para Belkin (1978), o problema abordado pela CI pode ser considerado como sendo facilitar a comunicação efetiva de informação desejada entre gerador e receptor.

Situando a criação de uma nova área de pesquisa sob o escopo da CI, em 1951 Calvin Mooers cunhou o termo *Information Retrieval* (Recuperação da Informação – RI) (SARACEVIC, 1995) e definiu as questões com as quais ela se ocuparia:

Recuperação da Informação é o nome para o processo ou método segundo o qual um provável usuário da informação é capaz de converter a sua necessidade por informação em uma lista real de citações para documentos armazenados contendo informação útil para ele. [...] Recuperação da Informação engloba os aspectos intelectuais de descrição da informação e sua especificação para busca e também quaisquer sistemas, técnicas ou máquinas que são utilizadas para executar a operação (MOOERS, 1951 apud SARACEVIC, 1995, p. 139, tradução nossa).

A RI, portanto, deveria se ocupar tanto de fatores humanos ao lidar com a subjetividade interpretativa de cada indivíduo ao descrever informações para armazená-las e também as suas necessidades informacionais para recuperá-las, quanto de fatores técnicos para o desempenho do processo de armazenamento e recuperação. Cabe ressaltar que nesta clássica fronteira imaginária entre homens e máquinas, elementos técnicos já foram desempenhados por atores humanos, mas o que se percebe atualmente é a busca constante pela substituição destes por tecnologias presentes nas etapas de descrição, busca e operação.

Ampliando o leque das definições sobre RI, Robredo (2005) a descreve como parte integrante do processo global de gestão da informação e do conhecimento. Para tanto, não deve ser dissociada da necessidade de estruturação de dados e informação na descrição, na organização e na armazenagem. Por sua vez, Van Rijsbergen (1979) descreve o problema de armazenar e recuperar informação como o dilema entre a grande quantidade de informação disponível e a cada vez maior

dificuldade em ofertar acesso correto e rápido a ela. Manning, Raghavan e Schütze (2009, p. 1, tradução nossa) abordam RI da seguinte maneira:

Recuperação da Informação (RI) é encontrar material (normalmente documentos) de uma natureza não estruturada (normalmente texto) que satisfaz uma necessidade de informação em grandes coleções (normalmente armazenadas em computadores).

Outra definição relevada com o aumento da informação disponível refere-se à Arquitetura da Informação (AI). A AI é preocupada com a organização de padrões em dados de forma a facilitar o acesso, a compreensão e a construção de conhecimento pelas pessoas. O termo foi cunhado pelo arquiteto Richard Saul Wurman em 1976 ao lidar com o fenômeno de crescimento exponencial da informação (por ele denominado de “explosão de informação”), efeito dos avanços da tecnologia e da maior facilidade de transmissão e armazenamento da informação, gerando ao mesmo tempo novas oportunidades de compartilhamento e criação de conhecimento e um grande volume de informação não organizada que se confronta com a capacidade limitada de percepção humana (WURMAN, 1991; WYLLYS, 2000). A relação da AI com a RI existe no destaque dado à apresentação da informação para facilitar a sua localização e a sua utilidade. Define-se o papel do arquiteto da informação da seguinte forma:

1) o indivíduo que organiza os padrões inerentes em dados, tornando o complexo claro. 2) a pessoa que cria a estrutura ou mapa de informação que permite que outros encontrem seus caminhos pessoais para o conhecimento. 3) a emergente ocupação profissional do século 21 que aborda as necessidades da época centrada em clareza, compreensão humana e a ciência da organização da informação (WURMAN, 1996 apud WYLLYS, 2000, tradução nossa).

Etimologicamente proveniente da expressão “dar forma” (CUNHA, 1986), o significado da palavra informação é alvo de exaustiva discussão. Como exemplo é possível citar o compreensivo artigo “O conceito de informação” de Rafael Capurro e Birger Hjørland. A informação é abordada como dependente do contexto e da relação em que aparece com outras palavras e do significado que adquire de acordo com a função em que é utilizada. Mais especificamente, na maneira em que comunica conhecimento a um indivíduo de acordo com suas necessidades e habilidades de interpretação (CAPURRO; HJORLAND 2007).

A informação, por depender da interpretação humana, pode proporcionar sentidos distintos para diferentes pessoas e para uma mesma pessoa em diferentes ocasiões. De acordo com Davenport e Prusak (1999), é o receptor e não o emissor

quem define se a mensagem constitui informação. Eis a subjetividade que ocupa os processos de descrição e busca estudados pela RI. A própria definição etimológica da palavra recuperação, que designa o “ato ou efeito de recobrar o perdido, de adquirir novamente” (CUNHA, 1986), sugere a apropriação de sentidos dada exclusivamente pelos seres humanos.

O dilema entre a quantidade de informação disponível e o maior acesso exposto por Van Rijsbergen permite sugerir que as tecnologias de informação e comunicação influenciam tanto o aumento do volume informacional quanto as ferramentas para efetiva recuperação da informação. Primeiramente, por permitirem a criação, difusão e armazenamento de informação de maneira cada vez mais simples e direta, o compartilhamento e a colaboração entre indivíduos, suplantando barreiras físicas e geográficas; e posteriormente por nelas ser depositada a esperança do emprego de técnicas desenvolvidas tanto para organizar a informação automaticamente (de forma a possibilitar a fácil recuperação), quanto para a execução de procedimentos para a recuperação da informação realmente desejada.

Entretanto, as expectativas depositadas nas técnicas automatizadas são até hoje motivo de discussões e estudos. Ainda em 1979, Van Rijsbergen abordou obstáculos para a caracterização automática do conteúdo de documentos com o intuito de organizá-los para armazenamento. Entre eles, a dificuldade de fazer os computadores simularem o processo humano de leitura, interpretação e extração de informação de documentos, de forma a considerar corretamente sentidos sintáticos e semânticos e utilizá-los para avaliar a relevância (o que por si só também representa outro obstáculo, abordado com mais detalhes na análise da avaliação em RI) de um documento como resposta a consultas realizadas ao sistema de recuperação. O próprio processo humano, onde a caracterização da informação contida em um documento normalmente é feita considerando o conhecimento da pessoa que executa a ação e as suas expectativas quanto ao que espera que seja buscado pelos usuários, é extremamente suscetível às subjetividades que a interpretação pessoal provoca.

Além de dificuldades quanto aos objetivos de RI, existem ainda barreiras na definição dos meios utilizados para atingi-los. A descrição da informação utilizando a linguagem natural, usada na comunicação humana, gera tanto problemas de compreensão quanto de armazenamento (e logo, de posterior recuperação) pela possibilidade iminente de dubiedade nos termos aplicados a determinado objeto

(sintaxe) e também aos sentidos praticados para cada termo (semântica). Já o emprego de uma linguagem artificial desenvolvida para um sistema de recuperação, mesmo com o intuito de aproximar-se à linguagem natural para facilitar o compartilhamento e a troca de informação (FREIRE, 2006), obriga o usuário a representar a sua necessidade informacional em uma linguagem que não lhe é de uso cotidiano, o que pode vir a dificultar a execução de tarefas que já são subjetivas.

A abordagem de sistemas informatizados de recuperação evoca a participação de outra ciência que também possui amplos interesses em RI. Estudos em Ciência da Computação apresentam a indissociável relação entre as técnicas de recuperação citadas anteriormente e as tecnologias computacionais. De fato, muito da pesquisa em RI é proveniente de experiências que envolvem essas tecnologias. Porém, a definição clássica para aquela ciência do que é informação (dado para o qual é concedido algum sentido) pode explicar algumas das limitações de sua abordagem. Como fazer um elemento artificial expressar sentido (normalmente ligado à interpretação individual humana) para um objeto? Pesquisas em Inteligência Artificial continuamente buscam essa resposta. Entretanto, por muito tempo, experimentos computacionais em RI trabalharam na verdade com a recuperação de dados, e não necessariamente de informação, panorama que vem sendo oportunamente abandonado. O quadro 1 apresenta algumas das características da recuperação de dados confrontadas com as da RI. As características objetivas e determinísticas da recuperação de dados sem significação e sem contexto, embora aparentemente facilitem a execução do processo, se distanciam da real necessidade informativa das pessoas.

	Recuperação de Dados (RD)	Recuperação da Informação (RI)
Coincidência	Coincidência exata	Coincidência parcial, melhor coincidência
Inferência	Dedução	Indução
Modelo	Determinístico	Probabilístico
Classificação	Monotética	Politética
Linguagem de consulta	Artificial	Natural
Especificação de consulta	Completa	Incompleta
Itens desejados	Coincidentes	Relevantes
Resposta de erro	Sensitiva	Não sensitiva

Quadro 1 - Recuperação de Dados ou Recuperação da Informação? (VAN RIJSBERGEN, 1979, p. 2, tradução nossa)

A RI possui uma esfera de atuação mais rica e ao mesmo tempo mais subjetiva que a recuperação de dados. Lida com a idéia de correspondência entre

busca e resultado mais aplicada ao contexto e a indução de resultados para recuperação de itens relevantes. Por estar aberta a essa capacidade interpretativa, revela-se não como uma tarefa única, mas como um conjunto de ações onde a interação contínua com o ente que utiliza seus serviços (daqui por diante designado como usuário) é sempre presente.

Dessa forma, definições como as de Lancaster (1968 apud VAN RIJSBERGEN, 1979, p. 1), segundo o qual a RI não informa um usuário sobre um assunto, mas sim quais são os documentos que tratam do assunto buscado, além de contraditórias, tornam-se cada vez mais ultrapassadas. São contraditórias pois um sistema de RI, ao comparar descrições de necessidades de informação formuladas por um usuário com descrições de informação armazenadas, não pode deixar de realizar algum tipo de avaliação ou inferência como método de busca ou mesmo de ordenação da apresentação de resultados. Logo, até mesmo esses resultados podem efetivamente representar informações para o usuário do sistema. Além disso, se o sentido utilizado para formular a necessidade informacional apresenta-se extensivamente representado nos resultados trazidos pelo sistema, isso pode dizer que a base de informações que ele contém seja de grande valia para os objetivos do usuário, assim como também pode dizer que o sistema não aplica triagem suficientemente satisfatória. Por outro lado, se a necessidade informacional aparece pouco representada nos resultados, pode ser que a base utilizada não apresente a informação desejada ou a apresente em poucas representações, ou até mesmo que a própria necessidade informacional, como formulada, não se apresenta suficientemente compreensível ou representativa para o sistema. São ultrapassadas pelo fato de que não se preocupam com o valor cada vez mais esperado das informações obtidas nestes sistemas.

Calado (2009) abordou o percurso histórico de RI considerando o emprego de sistemas informatizados de RI. Iniciado pelo processamento de dados (considerados como fatos ou conceitos), o ciclo passa ainda pelo processamento de texto, pelo processamento de contexto, processamento da informação (organizada, estruturada, com significado, contexto e interpretação) e finalmente o raciocínio. Aplicada aos sistemas de RI, a recuperação da informação propriamente dita depende que eles inicialmente compreendam dados isolados e depois os registros textuais de informação, aplicando o contexto da atividade de recuperação e do próprio usuário.

A fase de raciocínio depreende conexões entre informações processadas para recuperação de forma a inferir novo conhecimento.

Pelo percurso histórico do desenvolvimento da RI, como descrito em forma de analogia ao crescimento humano por Lesk (1995), a área passou da infância a uma crise de meia idade, ultrapassando a idade escolar, a idade adulta e a maturidade, buscando o futuro cumprimento de seus objetivos e por fim a sua aposentadoria. Atualmente, da visão de Vannevar Bush, se detém que os equipamentos já se encontram defasados pelas novas tecnologias, e mesmo assim os objetivos ainda não foram atingidos. Desenvolvimentos passados e ainda exercitados em interfaces de sistemas, mesclando conhecimentos de computação e de ciências tradicionais como a lingüística; além da maior disponibilidade de meios de distribuição de informação e de acesso a cada vez mais pessoas comprovam que a preocupação em RI ultrapassa questões técnicas como performance ou desempenho.

Durante a discussão teórica sobre o papel e a abrangência da RI, é possível observar a citação constante dos sistemas de RI, apresentados desde aplicações de busca de acervo em bibliotecas até complexos serviços de busca na Internet. É através deles que a RI se aproxima do cotidiano e normalmente aplica suas pesquisas. Os elementos que configuram estes sistemas e a importância de avaliar seus objetivos são descritos a seguir.

2.1 *Sistemas de Recuperação da Informação, avaliação e relevância*

O objetivo final da RI é a aproximação da necessidade informacional apresentada por um usuário às descrições de informação obtidas como resultado de uma consulta realizada por ele. Segundo definição de Van Rijsbergen, “o objetivo de uma estratégia de recuperação automática é recuperar todos os documentos *relevantes* e ao mesmo tempo recuperar tão pouco dos *não relevantes* quanto possível” (VAN RIJSBERGEN, 1979, p. 6, tradução nossa). É possível conceber que a única forma de mensurar como um sistema de RI trata um ou outro objetivo é através da avaliação de sua efetividade, utilizando para isso cenários e variáveis específicos para o contexto de atuação do sistema.

Antes mesmo de definir do que trata a avaliação em RI, é necessário esclarecer o papel dos sistemas na área. Segunda Cunha (1986), etimologicamente a palavra sistema define um “conjunto de elementos, materiais ou ideais, entre os quais se possa encontrar ou definir alguma relação” e também um “método, processo”.

Substituindo os termos da definição etimológica e considerando o viés de administração corporativa, Laudon e Laudon (2007, p. 9) caracterizam sistemas de informação da seguinte maneira:

Um sistema de informação pode ser definido tecnicamente como um conjunto de componentes inter-relacionados que coletam (ou recuperam), processam, armazenam e distribuem informações destinadas a apoiar a tomada de decisões, a coordenação e o controle de uma organização.

Os autores complementam sua definição salientando que, diferentemente do que é normalmente imaginado, sistemas de informação não são restritos a tecnologia. Na verdade, esta seria apenas uma das dimensões que compõem os sistemas, juntamente com as organizações em que eles são aplicados e para quem são construídos, e as pessoas que os utilizam.

Sob o âmbito da CI, é possível definir sistemas de informação como conjuntos formados por rotinas de processamento de informação, registros informacionais e usuários, com o objetivo de permitir a manipulação da informação contida nos registros pelos usuários para atender às suas necessidades. Em sistemas de RI, a manipulação citada refere-se à descrição, à organização, ao armazenamento e ao acesso à informação. Especificamente, o acesso à informação é composto por etapas de recebimento de descrições de necessidades informacionais dos usuários, de comparação destas com os registros armazenados no sistema e a apresentação dos resultados propostos pelo sistema.

Somente é possível definir a aplicação de avaliação sobre o sistema de RI depois de delimitado o seu contexto de atuação, obtendo o que se está avaliando e de que forma a avaliação pode ser feita. A avaliação, segundo Saracevic (1995, p. 138, tradução nossa), significa “medir a performance ou o valor de um sistema, processo (técnica, procedimento...), produto ou política”. Quando define a avaliação em RI, o autor retoma a “explosão da informação” cunhada por Vannevar Bush e a relaciona com o papel da RI como solução para os problemas causados pelo excesso de informação, gerando o questionamento a seguir:

Como a RI dá suporte às pessoas em situações nas quais elas são confrontadas com problemas de buscar, encontrar, utilizar e interagir com informação da massa de informação existente e a miríade de escolhas disponível? (SARACEVIC, 1995, p. 138, tradução nossa).

Ellis (1990) abordou a relação entre RI e a avaliação dos sistemas de RI. Ao descrever a sessão de testes sobre sistemas de RI denominados Cranfield I pela *Association of Special Libraries and Information Bureaux* (ASLIB), na cidade de Cranfield no Reino Unido em 1957, duas variáveis básicas eram calculadas: “revocação” e “relevância”. A primeira era a proporção entre documentos considerados relevantes na coleção e aqueles efetivamente recuperados. Já a segunda era a proporção dos documentos recuperados que seriam de fato relevantes. Associou-se então a relevância à avaliação de sistemas de RI.

Com a segunda sessão de testes denominada de Cranfield II, os conceitos de revocação e relevância surgidos anteriormente foram evoluídos e redefinidos, dando origem às variáveis de avaliação de sistemas de RI mais utilizadas até hoje: “revocação”, agora entendida como a relação entre os documentos relevantes recuperados e o total de documentos relevantes existentes na base do sistema; e “precisão”, a relação entre os documentos relevantes recuperados e o total de documentos recuperados. Na mesma sessão de testes, outra hipótese importante levantada foi a de que estas variáveis possuiriam uma relação contraditória entre si. Isto é, caso um sistema de RI apresentasse ou fosse ajustado para boa revocação, provavelmente apresentaria piora na precisão, e vice-versa. Entretanto, a discussão a respeito de revocação e precisão engloba um problema que pode ser considerado anterior às variáveis, por estar presente na composição de ambas. Como designar corretamente (ou simplesmente designar) a relevância de um documento (ou informação) para um usuário?

Os testes denominados ASTIA (*Armed Services Technical Information Agency*) – Uniterm realizados nos Estados Unidos, foram descritos por Cleverdon (CLEVERDON; MILLS; KEEN, 1966; ELLIS, 1990). Mesmo havendo apenas dois grupos de pessoas envolvidos, um da própria ASTIA e o outro de funcionários da Documentation, Inc., empresa de Mortimer Taube, inventor do sistema de RI Uniterm utilizado nos testes, o consenso a respeito de quais itens seriam efetivamente relevantes esteve longe de ser obtido. Com isso, o resultado dos testes pôde ser questionando em relação a este ponto. Considerando que um dos grupos envolvidos

neste caso provavelmente estava interessado em lucrar com o produto sendo testado, o que se percebe é que a definição de relevância foi alvo de discussões pois os interesses (e os fatores que impactam a relevância) sempre se mostraram distintos, em maior ou menor grau, para usuários e contextos de recuperação diferentes em um mesmo sistema.

Cleverdon definiu que para um mesmo usuário existem dois tipos de relevância, um denominado “relevância declarada” (no original, *stated relevance*) e o outro “relevância do usuário” (*user relevance*), diferentes entre si. A primeira seria encontrada no pedido de informação expresso pelo usuário ao sistema e a segunda na real necessidade do usuário. A relevância declarada pode ser considerada apenas como uma forma artificial de definição de relevância para tentar facilitar a execução da tarefa de RI. Isso porque a relevância do usuário é a que essencialmente representa o senso de efetividade do sistema de RI para o usuário, apresentando a subjetividade da própria interpretação dele sobre sua necessidade informacional e até mesmo de outros fatores, como os resultados de sua interação com o sistema.

Um mesmo usuário pode apresentar noções diferenciadas de relevância para um mesmo conjunto de informações e para um mesmo sistema uma vez que os contextos (e, conseqüentemente as necessidades informacionais) tendem a nunca se repetir. Exemplificando, imagine que alguém deseje saber mais sobre computadores. Em um primeiro momento, ele busca e recupera alguns registros que, se não respondem inteiramente a sua necessidade, provocam alterações no seu estado de conhecimento. Com isso, uma segunda busca já apresenta embasamento e requisitos que a primeira não possuía. Um documento considerado relevante inicialmente pode agora se tornar inútil. O contexto de busca já foi modificado, assim como as necessidades informacionais do usuário. Decorre de todos estes fatores a afirmação de que a natureza e a confiabilidade do julgamento de relevância se tornam o problema intelectual de maior dificuldade na avaliação da efetividade de sistemas de RI para seus usuários (CLEVERDON; MILLS; KEEN, 1966; ELLIS, 1990).

Apesar do problema cada vez mais enfatizado da determinação de relevância, de uma maneira geral as sessões de testes realizadas nas décadas de 50 e 60 (como o ASTIA-Uniterm e os Cranfield) definiram um padrão referente à necessidade de avaliação dos sistemas de RI. Estes testes apresentavam um

formato segundo o qual as variáveis seriam controladas como em um laboratório, o que valeu a denominação “paradigma físico” para avaliações dessa natureza. Porém, o mesmo controle aplicado sobre os parâmetros e que permitia a execução das rotinas de avaliação tornava o ambiente de teste distante da realidade de aplicação dos sistemas de RI por usuários, desconsiderando fatores como a importância relativa dos termos para cada usuário e a ordenação dos resultados apresentados, e a difícil utilização por não peritos (CALADO, 2009). A partir do reconhecimento desta falha, aspectos cognitivos e de comportamento dos usuários passaram a ser levados em consideração na execução de avaliações de sistemas de RI.

Estudos em RI originaram experimentos e produtos como serviços, sistemas e aplicações específicos para a área, normalmente explicitados em tecnologias constantemente aprimoradas. Entretanto, conforme lembra Saracevic (1996), por mais que sejam utilizadas abundantemente, a eficácia do uso dessas tecnologias somente pode ser mensurada através de avaliações humanas. Por isso defende que o conceito de eficácia seja recolocado como aspecto crítico central para avaliação de avanços na CI (e por consequência na RI), em detrimento ao fascínio provocado pelas aplicações tecnológicas. O próprio autor menciona uma constatação, segundo a qual apesar do aumento exponencial da quantidade de publicações existe uma concentração das que são efetivamente citadas em outras publicações, de forma a corroborar o papel que a eficácia desempenha nas avaliações e no uso da informação:

Por muito tempo, o principal critério para se focar a eficácia foi a relevância e/ou utilidade da informação. Mas, mais recentemente, tem-se escutado apelos por outros critérios – como qualidade, seletividade, veracidade, síntese e/ou impacto da informação. O fato estatístico chave da explosão informacional está bem estabelecido – por exemplo, em ciência e tecnologia, o número de publicações dobra aproximadamente a cada quatorze anos. A qualidade dessas publicações varia amplamente. Algumas investigações estabeleceram leis empíricas, demonstrando que apenas uma pequena parte é altamente utilizada, considerada de alta qualidade, ou citada, e o uso, qualidade e citação das restantes declinam rapidamente. Parece que um processo natural, darwiniano, de seleção está em curso.[...] (SARACEVIC, 1996, p. 56-57).

A afirmação de Saracevic pode ser entendida como um alerta para todos aqueles que acreditam que apenas avanços em tecnologia poderão resolver os problemas da humanidade. No caso específico de RI, os sistemas informatizados só apresentaram evoluções consideráveis quando, a despeito de sua origem

tecnológica, englobaram o conceito de interação com os usuários. Não apenas por uma necessidade de melhor atender as suas necessidades (objetivo mínimo que garante a existência de um sistema), mas, em uma vertente cada vez mais importante hoje em dia, também com o intuito de apreender deles melhores formas de descrever ou recuperar a informação. Com isso, um sistema que é levado a compreender que quanto mais citações um documento possui em publicações de autores de uma mesma área, maior é a probabilidade de ele ser relevante para aquela área, possivelmente tem maiores chances de sucesso do que outro que desconsidera esse exemplo de comportamento humano. A partir daí é possível tentar inferir outros pontos como quais periódicos possuem artigos mais citados, em que áreas, quais são seus autores, e assim por diante. Surge então uma miríade de alternativas de desenvolvimento, concentrada nos aspectos cognitivos e comportamentais dos usuários. Porém, ao mesmo tempo, um sistema que seja levado a processar informações desta forma pode desconsiderar conhecimento relevante em artigos menos citados, não visíveis segundo essa metodologia. Avanços feitos pelo paradigma físico da pesquisa em RI não poderiam trazer alternativas a problemas como este?

O histórico da mudança de paradigmas para a eficácia de sistemas de RI citados brevemente por Saracevic foi debatido de forma mais ampla por Bates em uma palestra ministrada no encontro anual da *American Society for Information Science and Technology* em 2002. Denominada de “Conceituando usuários e usos” (no original, *Conceptualizing Users and Uses*) a sessão tinha como objetivo debater sobre o que a autora denominou de “provincianismo” da área de Ciência da Informação (especificamente no campo de comportamento de busca de informação) ao lidar com as mudanças de enfoques na área. Nessas situações, o que normalmente ocorria era que a definição de um novo paradigma significava o total abandono dos avanços obtidos com o paradigma anterior. Além disso, até então, todos os enfoques vislumbrados para a área eram provenientes de outras áreas de estudos, trazendo desta forma restrições às suas aplicações no campo da CI (BATES, 2002).

Pela ordem, primeiramente os métodos de pesquisa obedeciam a uma metodologia objetivista, também chamada de métodos quantitativos. Importada das ciências naturais (matemática, física, etc.), essa metodologia consistia em manipulação e mensuração de variáveis controladas nos ambientes de pesquisa,

sem riscos de desqualificação da pesquisa, tornando inválidas as influências do comportamento dos usuários na interação com os sistemas (BATES, 2002). Não coincidentemente, essas são características dos testes em sistemas de RI efetuados nas décadas de 50 e 60 e descritos anteriormente, como os testes ASTIA-Uniterm e os Cranfield. Estes, servindo de exemplo para pesquisas que os seguiram, influenciaram para que utilizassem igualmente o paradigma objetivista, mesmo com as discussões e controvérsias que apresentava, então simplesmente desconhecidas ou mesmo desconsideradas.

Em um segundo momento, a metodologia quantitativa foi substituída por outra qualitativa, baseada na subjetividade típica da pesquisa em ciências humanas. Enquanto a anterior clamava por ignorar as influências dos usuários sobre os sistemas, o novo paradigma era baseado exclusivamente na perspectiva única que cada usuário possui ao interagir com ele. Com isso, a forma de se obter definições válidas seria através do estabelecimento de consenso entre os usuários para as questões abordadas, mesmo que este consenso não representasse nenhuma “razoavelmente válida observação” do universo de estudo (BATES, 2002). Reside neste porém uma característica intrínseca das ciências humanas e que pode representar restrições para o emprego em ciências sociais aplicadas. Como obter definições efetivamente aplicáveis para problemas estudados se elas podem ser concebidas sem respeitar o contexto, o domínio ou até mesmo sem delimitar o universo do problema em si?

Da mesma forma como o enfoque objetivo pode se desprender da realidade de aplicação dos sistemas de RI ao desconsiderar a interação com seus usuários, o enfoque subjetivo pode simplesmente tornar-se inútil para fins de avaliação dos sistemas de RI ao não restringir algum tipo de domínio para uso destes sistemas pelos usuários. Entretanto, concluir que ambos os paradigmas devem ser desprezados de forma total é errôneo. Eles possuem características e avanços que, combinados, podem servir ao propósito avaliativo almejado.

Bates (2002) aborda que a substituição de um paradigma por outro não responde integralmente aos problemas de estudos de comportamento de busca de informação em CI, ainda mais se os avanços obtidos em cada um deles não puderem ser analisados com o intuito de reaproveitamento e até mesmo de combinação para melhoria dos métodos. Indo mais além, a autora ainda sugere que

o possível surgimento de uma metodologia específica de ciências sociais deve ser resultante da combinação de outras metodologias:

Eu gostaria de propor outra forma de olhar para estes vários métodos [objetivos e qualitativos]. Em minha laaaaaarga [sic] (!) experiência, o que eu descobri foi que não é o método em si que faz boa ou má pesquisa, mas sim a qualidade da reflexão atrás do desenho, a adequação do método à questão perguntada e o cuidado com que o trabalho é executado que determina se os resultados serão válidos, significativos ou úteis (BATES, 2002, p. 2, tradução nossa).

O que é primordial para definir como deve ser feita a avaliação sobre um sistema de RI é conhecer os objetivos deste sistema (e por que não, de seus usuários). Somente dessa forma é possível saber quais características o sistema deve apresentar, o que o usuário espera dele, de que forma ele enxerga a sua interação, entre outros fatores. A partir destes detalhes um avaliador pode definir variáveis de estudo, base de dados de teste e o próprio domínio de aplicação, ou seja, a metodologia da avaliação, combinando peculiaridades de um paradigma ou de outro que sejam proveitosas para o caso específico. Ater-se não a um modelo pré-definido, mas ao problema pesquisado.

Tal forma de abordar a avaliação da RI como dependente não apenas de paradigmas, mas principalmente do contexto (quem são os usuários, quais os objetivos do sistema, quais as características do sistema) de aplicação do sistema avaliado encontra outro defensor em Saracevic (1996). O autor percebeu que o contexto da avaliação deve influenciar a abordagem escolhida porque esta, por sua vez, influencia os resultados obtidos, que devem responder questões do contexto com informações úteis, formando um ciclo de avaliação. Este ciclo coincide ainda com o papel da interatividade entre usuário, questionando e reformulando suas consultas de acordo com as respostas que recebe, e sistema.

Os métodos de avaliação baseados em contexto, quando comparados aos métodos mais tradicionais, acabam apresentando distanciamento do ambiente controlado proposto pelo paradigma físico ou quantitativo, justamente pela dificuldade em propor fronteiras ao sistema avaliado. Mas, da mesma forma, devem vir a propor observações válidas e com critérios definidos pelo menos de porções destes universos de pesquisa, o que também os afasta dos métodos puramente qualitativos das ciências humanas.

Seguindo os passos de Saracevic (1996), são cinco os itens que precisam ser estudados de acordo com o contexto de forma a definir a avaliação de um sistema de RI:

- O sistema em si (ou sua representação e as técnicas que emprega);
- Critérios (representando os objetivos do sistema);
- Métricas (baseadas nos critérios);
- Instrumentos de medição; e
- Metodologia (que agrega os demais itens).

O planejamento das avaliações em RI, considerando as características dos objetivos da área da ciência nos dias atuais de Internet e de sistemas de alta interatividade com o usuário, deve ser feito de maneira a entender o contexto das suas aplicações, determinando métricas que visem a responder questionamentos ligados aos objetivos dos usuários e dos sistemas. Somente dessa forma será possível corresponder ao papel da RI no auxílio às pessoas para armazenar, buscar, encontrar e utilizar informação útil em sistemas de informação.

2.2 A sociedade da informação e a Recuperação da Informação

Ao abordar a origem e a evolução da CI, Saracevic (1996) opina que a RI influencia a “emergência, a forma e a evolução da indústria informacional”. Por indústria informacional é possível entender o leque de alternativas quanto a áreas para estudos e aplicação de técnicas que envolvem o uso efetivo da informação para o conhecimento humano.

O trabalho determinado pela necessidade de recuperar informações suscitou questões e promoveu pesquisas exploratórias de fenômenos, processos e variáveis, bem como das causas, efeitos, comportamentos e manifestações relacionados. Historicamente, este fato conduziu a estudos teóricos e experimentais sobre a natureza da informação, a estrutura do conhecimento e seus registros (incluindo bibliometria), o uso e os usuários, levando a estudos do comportamento humano frente à informação; a interação homem-computador, com ênfase no lado humano da equação; relevância, utilidade, obsolescência e outros atributos do uso da informação juntamente com medidas e métodos de avaliação dos sistemas de recuperação da informação; economia, impacto e valor da informação, dentre outros (SARACEVIC, 1996, p. 44-45).

Estudos em RI cresceram em importância na mesma medida em que a informação se tornou mais abundante e influente no cotidiano, em que se passou a notar o valor da informação não somente para estudos científicos, mas para obter vantagens competitivas ou simplesmente para suplantar dificuldades no cumprimento de tarefas humanas. Por isso, a abordagem dos conceitos históricos da RI, de seu papel, de seus sistemas e de seus métodos de avaliação é feita não apenas com o intuito de conhecer suas peculiaridades, mas sim de confrontá-la com os novos desafios propostos pela era de desenvolvimento tecnológico e de difusão de conhecimento atualmente vivenciada.

Freire (2006, p. 10) nota que “a característica marcante da atual sociedade não seria apenas a apropriação da informação e do conhecimento pela sociedade, mas a transformação de ambos em forças produtivas”. Com isso, opina que a RI deve ter o intuito de tornar disponível a informação para utilidade dos seus usuários potenciais e de toda a sociedade. Trata-se, portanto, de uma nova era em que a importância da informação influencia diretamente o bem estar e a melhoria da qualidade da vida humana, devendo se ocupar cada vez mais das condições de acesso como forma de subjugar desigualdades nesta sociedade doravante denominada de sociedade da informação (PROGRAMA SOCIEDADE DA INFORMAÇÃO, 2000).

Estabelecendo o valor do conhecimento para as organizações empresariais, Davenport e Prusak (1999) citam os pensamentos de Peter Drucker, que define o conhecimento como base de concorrência na sociedade pós-capitalismo, e do economista de Stanford, Paul Romer, que o define como o único recurso ilimitado na sociedade humana. Propõem ainda uma definição:

Conhecimento é uma mistura fluida de experiência condensada, valores, informação contextual e *insight* experimentado, a qual proporciona uma estrutura para a avaliação e incorporação de novas experiências e informações. Ele tem origem e é aplicado na mente dos conhecedores. Nas organizações, ele costuma estar embutido não só em documentos ou repositórios, mas também em rotinas, processos, práticas e normas organizacionais (DAVENPORT; PRUSAK, 1999, p. 6).

Como característica desta nova sociedade existe a constante interligação entre pessoas sem a necessidade de respeitar fronteiras senão as de interesses comuns com o objetivo de compartilhar conhecimento para gerar produto diretamente ligado a ele e novos conhecimentos, e não para a criação de um produto industrial (CASTELLS, 1999). Entre outros fatores influenciados, o autor

destaca o papel da individualidade de cada ser perante a rede em que se encontra inserido, em áreas como a economia, a relação entre empresas, o mercado de trabalho, a difusão, o acesso à cultura e o próprio fluxo da informação. Tapscott (1998) define a cultura dessa sociedade como sendo de inovação, colaboração e de inter-relacionamento entre as pessoas (no original, *networking*).

“Mercado da informação” é outro termo utilizado para simbolizar esta época (DERTOUZOS, 1997), sendo compreendido como o meio que reúne pessoas, redes informais, computadores, meios de comunicação e serviços para processamento e transmissão de informações. Neste ambiente, similar a um mercado de compra e venda, o valor da informação é dado comparativamente ao de um bem ou serviço útil, de forma a permitir trocas entre participantes cujo intuito seja a necessidade ou a oferta de informação e onde o valor do trabalho intelectual se assemelha ao do trabalho físico. Tal tipo de interação permitiria não apenas o aumento da produtividade humana, majorando a sua capacitação e minorando o seu esforço para realização de tarefas, mas também afetaria a “qualidade de vida, conveniência, acesso ao conhecimento, paz de espírito e relacionamentos humanos aprimorados” (DERTOUZOS, 1997, p. 124). O autor ainda designa da seguinte forma o valor econômico da informação:

A informação tem valor econômico quando leva à satisfação dos desejos humanos. Uma pequena parcela é formada por produtos finais, que derivam seu valor da oferta e da procura. A porção majoritária, porém, cabe aos bens intermediários, cujo valor deriva substancialmente do valor dos bens e serviços que deles se utilizam (DERTOUZOS, 1997, p. 297).

Na observação do autor há uma relação que não existe apenas na assim chamada sociedade da informação, mas também esclarece o ritmo de desenvolvimento (ou subdesenvolvimento) das nações. Por mais rica em recursos naturais (animais, vegetais ou minerais) que seja uma nação, somente se torna desenvolvida em padrões econômicos e de qualidade de vida aquela que saiba utilizar estes recursos como insumos para outros produtos mais elaborados. De forma análoga, este é um cenário que virá a se repetir com a informação. Assim como preza pelo acesso indiscriminado à informação e por “dar voz” a culturas então marginalizadas para conservação e publicidade de seu legado, na sociedade da informação, riscos existem que as desigualdades se aprofundem caso iniciativas de desmarginalização não sejam organizadas.

O desenvolvimento das Tecnologias de Informação e Comunicação (também referenciadas pela sigla TIC) está intrinsecamente ligado ao advento da sociedade da informação. Estas tecnologias englobam, entre outras, as transmissões de rádio, telefonia e TV bem como os dispositivos utilizados para elas. O computador, inicialmente pensado como uma máquina que serviria apenas para realizar cálculos matemáticos, passou a agregar funções de produtividade (como a criação de texto, planilhas, apresentações eletrônicas e impressão) e entretenimento (jogos, multimídia – sons, músicas e vídeos) utilizadas diretamente por usuários não tão especializados, e se tornou um exemplo de TIC cada vez mais comum não só em ambientes empresariais mas também domésticos.

A relação entre a tecnologia¹ e a sociedade da informação passa pela incorporação da segunda pela primeira e pela utilização da primeira pela segunda (CASTELLS, 1999). A incorporação é explicada ao abordar que a evolução tecnológica apenas tende a seguir ou reproduzir os anseios humanos. No chamado modo informacional de desenvolvimento batizado pelo autor, “a fonte de produtividade acha-se na tecnologia de geração de conhecimentos, de processamento da informação e de comunicação de símbolos” (CASTELLS, 1999, p. 53), a fim de atingir os objetivos de maior acúmulo de conhecimento e maiores níveis de complexidade do processamento da informação. Como fatores influenciados pelo aumento da produtividade, é possível entender desde a satisfação de objetivos pessoais até os processos de gerenciamento e de mudança em organizações (cuja adaptação a um novo modelo alimentado por TICs deve ser precedente à adoção ou escolha das tecnologias). Essas tecnologias, por sua vez, devem permitir a difusão dos seus efeitos, o agrupamento de idéias e de pessoas em redes de compartilhamento, a flexibilidade para adaptação e evolução, a convergência de várias tecnologias (inclusive de diferentes áreas) para um bem comum e a escalabilidade, termo que designa a capacidade para suportar o aumento do número de acessos.

A presença dos computadores em meio à sociedade iniciou pelas antigas máquinas de grande porte, confinadas em CPDs (Centrais de Processamento de Dados) com acesso restrito a cientistas/tecnólogos para fins especializados. E

¹ “O uso de conhecimentos científicos para especificar as vias de se fazerem as coisas de uma maneira reproduzível” (BROOKS, 1971 apud CASTELLS, 1999 p. 67; BELL, 1976 apud CASTELLS, 1999, p. 67).

evoluiu a ponto de fazer parte da interação diária de qualquer indivíduo na realização de suas tarefas cotidianas, presentes em eletrodomésticos e outros dispositivos de bolso, compreendendo o aprendizado digital a que são submetidas as novas gerações. Aproxima-se o lado técnico do advento dos computadores às necessidades humanas de seu emprego.

Assim como existem limitações humanas para lidar com complexidades sociais de relacionamento (inclusive referente à quantidade de pessoas com que se relaciona), existem limitações para lidar com a complexidade tecnológica (DERTOUZOS, 1997). A incorporação do uso dos computadores pelos seres humanos passou por melhorias sensíveis dos primeiros em relação à forma como interagem com os últimos. O desenvolvimento de interfaces aparece como uma das áreas onde isso é notável. Como interface de usuário, entende-se o meio em que há comunicação entre usuário e computador. Inicialmente através apenas de texto, digitação de comandos escritos em telas monocromáticas, a interação entre eles passou a ser realizada através de janelas gráficas, de uso amigável (em inglês, *user friendly*), utilizando dispositivos apontadores como um *mouse*, reconhecimento da fala, ou telas sensíveis ao toque (*touch screen*) e monitores coloridos, com apresentação de gráficos e até mesmo ambientes de imersão e simulação de situações reais como uma forma de realidade virtual, apresentando novas formas de experiência para os usuários.

A partir do constante desenvolvimento desses dispositivos, a criação de tecnologias de comunicação para interligá-los em rede se tornou o próximo passo. Inicialmente restritas a ambientes acadêmicos, corporativos ou governamentais específicos, essas redes de comunicação passaram a ser interligadas entre si, com abrangência cada vez maior. A expansão da infra-estrutura de telecomunicações, a sua digitalização, o aumento das velocidades de transmissão e a própria melhoria do desempenho dos computadores conectados em rede foram influenciadas e ao mesmo tempo impulsionaram o aumento do tráfego de comunicação da informação (BAR; BORRUS, 1993 apud CASTELLS, 1999). Uma dessas redes, a ARPANET (projeto da ARPA – *Advanced Research Projects Agency* – Agência de Projetos de Pesquisa Avançada) conectava instituições militares e acadêmicas americanas e serviu de base para a criação daquela que pode ser considerada a chave para a evolução do compartilhamento de informação: a Internet.

A Internet, tecnicamente especificada como a rede mundial de computadores, apresenta outras definições, técnicas ou não, mais aprofundadas. Ela é um conjunto de tecnologias e protocolos (padrões para comunicação de dados) que permite que computadores (e mais recentemente outros dispositivos, como telefones celulares) sejam interligados em redes, que por sua vez estão interligadas a outras redes, de forma crescente até atingir escala mundial, com o intuito básico de intercambiar informações. Não importa qual o tipo de computador ou dispositivo utilizado, se ele for apto a compreender o conjunto de tecnologias Internet, ele pode enxergar e compartilhar informação com qualquer outro na rede.

A Internet é uma rede de redes. Ela compreende uma grande coleção de tecnologias de computação, telecomunicações, entretenimento, publicação e outras. Ela abrange texto digitalizado, sons, imagens e vídeo e está rapidamente envolvendo outros formatos de informação [...]. Na medida em que dispositivos de automóveis a discos de hóquei se tornam objetos comunicativos, eles se tornam parte da Net. O assim chamado ciberespaço se expande cada vez que alguém se conecta (TAPSCOTT, 1998, p. 24, tradução nossa).

Por mais que sejam desempenhadas tarefas tão díspares quanto trocar mensagens eletrônicas (os conhecidos *e-mails*), compartilhar arquivos (unidade lógica segundo a qual informação de algum tipo é organizada para armazenamento e processamento por um computador) de música, baixar (um neologismo que não significa nada mais do que copiar algo pela Internet para o seu próprio dispositivo de acesso) vídeos ou simplesmente acessar páginas com conteúdo visual, no fundo todas elas representam o mesmo objetivo, que é permitir que um usuário (seja ele um humano ou mesmo virtualmente um computador ou outro dispositivo) tenha acesso a informações de outro(s) usuário(s).

Ao oferecer essa grande quantidade de serviços para troca de informações, a Internet permite que os usuários tenham a possibilidade de tornar disponível todo e qualquer conteúdo armazenado nos dispositivos que a acessam para os demais usuários. Por exemplo, se um poeta assim desejar, pode utilizar tecnologias da Internet para publicar seus manuscritos e poemas para quaisquer outros usuários da rede sem necessidade de pedir autorização ou mesmo hierarquizar o controle de acesso aos seus documentos. Existem restrições tais como usuários que bloqueiam o acesso a parte ou a todo o conteúdo sob o seu controle, ou mesmo políticas restritivas de acesso e divulgação, como para atender a leis autorais, por parte de empresas ou usuários comuns. Resumindo, o compartilhamento de conteúdo (e,

portanto, de informação) através da Internet pode ser feito de forma universal e sem um limite de escala para crescimento.

A comunicação da informação permitida pela Internet é símbolo da nova era. A mídia digital cerca a sociedade. Um mundo de informação armazenado em *bits* (menor unidade lógica para armazenamento de informação em formato digital) e não apenas em átomos (NEGROPONTE, 1995). A transmissão da informação neste formato digital, conseqüentemente, possui maior velocidade e maior abrangência, se tornando mais viável economicamente, mais conveniente para uso pela sociedade e capaz de suplantando barreiras geográficas. Torna-se, ao mesmo tempo, aliado para o uso democrático de informação e um obstáculo para dominação e controle da população por um regime opressor. Segundo o autor, a era digital é caracterizada pela “descentralização, globalização, harmonização e capacitação” (NEGROPONTE, 1995, p. 196).

O uso dos avanços tecnológicos, já imprescindível no cotidiano, modificou a forma como as pessoas realizam tarefas tão distintas quanto entreter-se, lecionar e aprender, se comunicar, comprar e nas relações e ferramentas de trabalho (TAPSCOTT, 1998). Segundo o autor, apesar desta invasão tecnológica obrigar o aprendizado por vezes forçado de gerações mais antigas de pessoas, para as novas gerações ela é totalmente transparente e normal, pois simplesmente elas já nasceram ou convivem desde cedo com as tecnologias estabelecidas, não imaginando o mundo sem elas.

A nova cultura juvenil na sociedade da informação é definida pelo autor como sendo de “padrões de comportamento, costumes, atitudes e códigos tácitos, crenças e valores, artes, conhecimento e formas sociais transmitidos socialmente e compartilhados” (TAPSCOTT, 1998, p. 55, tradução nossa). Temas normalmente abordados por essa cultura são a inclusão de pessoas em comunidades virtuais², a investigação baseada na maior disponibilidade de informação e a constante autenticação das informações obtidas de forma a se poder confiar nelas. Já no trabalho os temas são a abertura intelectual a novas idéias, a colaboração, o inter-relacionamento entre as pessoas (no original, *internetworking*) e as relações de confiabilidade e confiança entre empregados motivados e empregadores. Entretanto,

² “Agregações sociais que emergem da Net [Internet] quando pessoas suficientes se encarregam de discussões públicas suficientes, com suficiente sentido humano, para formar teias de relacionamentos pessoais no ciberespaço” (REINGOLD, 1993 apud TAPSCOTT, 1998, p.56, tradução nossa).

da mesma forma como este compartilhamento de padrões sugere o enriquecimento das relações e do conhecimento pessoal, ele também pode ser utilizado para difundir padrões que, caso equivocadamente empregados, resultem em degradação social.

Abordando o aspecto de globalização, o avanço da Internet influencia a difusão em escala mundial de conhecimento alojado em qualquer lugar, criando um novo meio de comunicação, assim como a radiodifusão e as transmissões televisivas. Porém, ao invés de usuários de televisão ou rádio que apenas recebem passivamente as informações enviadas por um ponto central (sendo por isso chamados de espectadores), os usuários de Internet possuem papel interativo (dando origem ao termo interatividade) como beneficiários da informação em todas as faixas de idade (LÉVY, 1999; TAPSCOTT, 1998) e facilitando o agrupamento de pessoas de acordo com interesses em comum.

Para Lévy (1999), a Internet representa um espaço onde as pessoas podem experimentar coletivamente formas de comunicação diferentes das mídias clássicas, explorando suas potencialidades de comunicação. Este espaço, denominado pelo autor de ciberespaço, é composto não apenas pelas pessoas, mas por todo o tipo de informação nele compartilhado e pelas próprias experiências de interação, dando origem ao que ele chama de cibercultura:

O ciberespaço (que também chamarei de "rede") é o novo meio de comunicação que surge da interconexão mundial dos computadores. O termo especifica não apenas a infra-estrutura material da comunicação digital, mas também o universo oceânico de informações que ela abriga, assim como os seres humanos que navegam e alimentam esse universo. Quanto ao neologismo "cibercultura", especifica aqui o conjunto de técnicas (materiais e intelectuais), de práticas, de atitudes, de modos de pensamento e de valores que se desenvolvem juntamente com o crescimento do ciberespaço (LÉVY, 1999, p. 17).

A Internet contém um novo veículo de comunicação, sendo classificada como um tipo de CMC – Comunicação Mediada por Computador. Este meio de comunicação se caracteriza pela representação e definição direta da cultura de seus usuários, pela existência de padrões de comunicação, pelo índice de penetração, descentralização, flexibilidade e pela possibilidade de gerar a chamada multipersonalização, que seria a representação de uma mesma informação para uma grande quantidade de pessoas, porém de forma personalizada para cada uma delas (CASTELLS, 1999). O emprego da Internet como CMC tem como exemplos a criação de comunidades pessoais de acordo com vínculos de afinidade em grupos

de discussão e as chamadas redes sociais, abordadas com maiores detalhes posteriormente.

Lévy (1999) indica ainda outras peculiaridades da Internet como CMC: a possibilidade de registro da inteligência coletiva baseada na colaboração e coordenação descentralizada; o apoio do meio de comunicação na digitalização da informação, o que possibilita maior velocidade do seu processamento e também de sua transmissão; o uso intenso das interfaces de comunicação de forma a facilitar a interação dos usuários; e o aspecto da virtualidade, segundo a qual as manifestações de um mesmo usuário podem ocorrer em diferentes lugares simultaneamente ou em momentos distintos. Porém, da mesma forma como pode promover a interação e a socialização, a Internet pode vir a se tornar um veículo de alienação da realidade e até mesmo de solidão de acordo com a forma como é utilizada por seus usuários.

Entretanto, um aspecto que precisa ser reforçado é a concentração de acessos em países e regiões ou públicos-alvo específicos, o que ainda representa uma dificuldade para a catalisação do desenvolvimento social em escala mundial. Por exemplo, países desenvolvidos como os Estados Unidos ou pertencentes à União Européia apresentam uma quantidade muito maior de acessos à Internet que o Brasil. No próprio país também existe essa diferenciação entre diferentes regiões e estados. Trata-se de um problema não apenas de acesso, mas também de representatividade, em se tratando de inclusão social e digital. Neste caso, não só o novo meio de comunicação não atinge seus benefícios, mas por ser o principal meio de divulgação de novo conhecimento para trabalho e geração de riquezas, acaba por resultar no aumento das desigualdades.

Acessar informação compartilhada através da Internet representa hoje grande parte do cotidiano. Notícias circulam rapidamente através do globo terrestre, não sendo necessariamente provenientes apenas de jornalistas ou de empresas de comunicação; pessoas conversam entre si através de programas de texto, de som e até mesmo de vídeo sem respeitar fronteiras geográficas; a divulgação da comunicação científica ultrapassa os limites de grupos compostos apenas pelos indivíduos diretamente envolvidos; informações armazenadas antes em sistemas restritos se tornam acessíveis *on-line*, isto é, em resposta direta a consultas realizadas pelos usuários.

A própria quantidade de informação disponível para as pessoas cresceu (e ainda cresce) de forma vertiginosa com o advento da Internet. A rede se alinha como tecnologia que fomenta o acesso ao conhecimento com o intuito de agregar valor produtivo à informação para a sociedade da informação. Porém, da mesma forma como permite o crescimento exponencial da Internet, a falta de controle sobre a qualidade e a quantidade da informação tornada disponível representa um grave obstáculo para o usufruto pleno de suas potencialidades por parte de seus usuários.

Outros problemas a serem abordados e que ainda apresentam empecilhos para o desenvolvimento das TICs referem-se a fatores como a replicação de tarefas que deveriam ter suas rotinas substituídas mas que passam a ser feitas tanto da forma anterior quanto com o emprego das tecnologias, a dificuldade de melhorias significativas em algumas tecnologias, o excesso de informação, a complexidade e o excesso de recursos desnecessários para determinadas tarefas, a dificuldade de mensurar a relação custo/benefício entre o tempo gasto no uso das tecnologias e a real necessidade da tarefa e a própria dependência das máquinas (DERTOUZOS, 1997).

Uma vez interligados os dispositivos em redes de comunicação, o próximo passo naturalmente seria a automação do compartilhamento de conhecimento entre eles. Aplicativos normalmente denominados de agentes (BERNERS-LEE, 2000; DERTOUZOS, 1997; NEGROPONTE, 1995) são os responsáveis por manter, requisitar e compartilhar informação nessas redes, normalmente utilizando cabeçalhos descritores existentes em cada registro de informação, de forma similar a índices³ em RI. Porém, a falta de infra-estrutura de informação e de padrões de comunicação culmina na necessidade de criar conexões entre conceitos de diferentes áreas e permitir a interligação de conhecimento pelos agentes. O mapeamento pela própria fonte da informação muitas vezes impacta o processo criativo de sua descrição, pela incapacidade humana de organizar previamente o que se gera. Além de prejudicar a representação, pode influenciar na utilidade da informação, desvirtuando a área de aplicação inicialmente prevista pelo autor. Este problema invariavelmente também ocorre quando o mapeamento é realizado por outras pessoas, considerando a influência de suas próprias interpretações no processo.

³ Palavras-chave ou qualquer outro tipo de dado que referencie e facilite o acesso a determinada informação

Nessa situação, considerando a diferenciação entre dados sem significância e informação, a rede passa a apresentar um excesso de dados para seus usuários, e não necessariamente de informação. A ansiedade da informação, conforme descrita por Wurman (1991), se manifesta quando, por mais que se tenha acesso à informação, não se é capaz de aproveitá-la como tal, ou simplesmente não há compreensão. Segundo o autor, das três atividades ligadas à comunicação (transmissão, armazenamento e compreensão), a última é a que deve depreender maiores esforços para melhoria.

A compreensão da informação na nova sociedade deve passar pela valorização do papel humano no desenvolvimento das tecnologias, até mesmo pelo fato de ser ele o interessado final. Não se trata de exigir do usuário o conhecimento sobre tudo, mas sim de capacitá-lo para encontrar a informação desejada. Isso dá através não apenas da exatidão da informação, mas com todo o processo pelo qual o usuário formula suas perguntas e o valor que elas carregam; espera organizar suas respostas; extrai e personaliza os sentidos que delas obtém, de acordo com o seu ponto de vista; e realiza conexões com o seu conhecimento anterior e o contexto em que se encontra inserido. A partir destas etapas, o usuário não apenas interage com a informação, mas também se torna capaz de utilizá-la de acordo com a sua necessidade.

No conceito de Ecologia da Informação definido por Davenport (1998), consta a necessidade de estudar o ambiente informacional de acordo com o seu contexto, incluindo as empresas, as pessoas, as políticas e então a Tecnologia da Informação, considerada apenas como parte da gestão do conhecimento. A evolução deste contexto passa pela definição de políticas que influenciem todos estes pontos e a própria estruturação dos dados.

O envolvimento do usuário visando ao melhor uso da informação passa não apenas pela interação com as tecnologias, mas também com a percepção de sua personalidade e de suas necessidades. Com isso, menos informação pode representar maior utilidade. A própria forma de apresentação da informação pode influenciar a facilidade de visualização e, portanto, de emprego pelo usuário. Além disso, na maioria das vezes, ao usuário é apresentada uma quantidade de informação muito maior do que a necessária, considerando não só a sua capacidade limitada de tratamento, mas o próprio objetivo de sua consulta. Por mais avanços que existam no fluxo de criação e transmissão da informação, eles só se tornam

úteis quando atendem a necessidades e objetivos de seus usuários. A informação, vista como produto ou como fonte de serviços, deve proporcionar melhorias no desempenho de suas atividades.

Nesse contexto, retoma-se o papel da RI como atividade social responsável pela interação entre produtores e usuários de informação em meio à explosão informacional, descrita como um problema social, cognitivo, cultural e de comunicação (SARACEVIC, 1995). A RI transcende desta forma a sua conjuntura tradicional de atuação, antes limitada a experimentos de biblioteconomia ou de teóricos de computação, para um campo de atividade fervilhante e cada vez mais influente com o advento da sociedade da informação, que é o estudo do impacto do volume de informação disponível na Internet. Isso envolve questões como o acesso e extração de informações efetivamente úteis para agregação de valor e construção de conhecimento, empregando técnicas como a lingüística (DUQUE, 2005) e estudos de usuários, atuando como intermediador necessário em meio à grande quantidade de informação inútil.

Por esses fatores, maior é a dimensão dos novos investimentos em pesquisa em RI e até mesmo os diversos interesses em que a pesquisa em RI passou a ser direcionada. “Quase do dia para a noite, a RI conquistou um lugar com outras tecnologias no centro do palco” (BAEZA-YATES; RIBEIRO-NETO, 1999, p. 3, tradução nossa). Torna-se natural a busca pela criação de uma rede de recuperação e distribuição equivalente à ampliação do sistema tecnológico de comunicação da informação (CASTELLS, 1999).

O que aparentemente representa apenas boas notícias traz também um alerta. A concentração de esforços em RI na sua relação com a Internet aglutina pesquisadores das mais diversas áreas. Filósofos, sociólogos e principalmente cientistas ligados à computação, entre outros, passaram a compor os cenários de estudo do campo. Com formações tão distintas, estas pessoas podem simplesmente não conhecer o vínculo de suas pesquisas com a RI e todo o seu legado, ou mesmo desconsiderá-lo por falta de real conhecimento sobre a área. De uma forma ou de outra, o risco de haver desenvolvimento de pesquisas que não levam em consideração conceitos clássicos e fundamentados da RI com o intuito de resolver os problemas dos quais ela se ocupa pode levar a percepções restritas dos universos pesquisados e emprego de esforço para analisar temas para os quais já existem estudos com escopo similares.

Esse tipo de situação ocorre com certa freqüência em pesquisas de Ciência da Computação que compartilham problemas com a RI. Notadamente jovem como área científica (da mesma forma que a CI), a Computação apresentava-se basicamente desafiada com o desenvolvimento de tecnologias para processamento da informação em computadores. Somente com o passar do tempo e o advento de novas dessas tecnologias, percebeu-se que muitas delas tinham o intuito básico de interagir com usuários humanos, ou mesmo tentar simular seus comportamentos. Os sistemas informatizados de RI são exemplos dessas tecnologias. Com isso, pesquisas de Computação e de RI com intuítos similares foram e ainda são desempenhadas em paralelo. Porém, da mesma forma como a mudança entre paradigmas de avaliação da RI não precisa implicar em total descarte dos conceitos e metodologias empregados pelo paradigma anterior, estes estudos podem levar em consideração avanços tanto da Computação quanto da RI.

Apesar da afirmação de que o intuito básico da Internet é o compartilhamento de informações, é possível observar que até agora se falou mais do acesso a recursos de forma virtualmente ilimitada que a rede possibilita. Obviamente, acesso neste contexto representa uma forma indireta de compartilhar a informação para a construção do conhecimento. Indireta pois, da maneira como foi até agora retratada, a Internet apresenta-se apenas como um gigantesco repositório de documentos e registros que somente podem ser copiados para objetivos próprios de cada usuário, mas não relacionados entre si ou imediatamente modificados de forma colaborativa.

Em uma biblioteca um usuário pode suprir suas necessidades informacionais tendo acesso a informações contidas em livros, periódicos, registros multimídia, entre outros suportes. Entretanto, ele não pode interagir com estes meios de forma a relacioná-los e construir conhecimento imediatamente registrável e incorporável ao acervo da biblioteca. Neste ambiente, um produto de conhecimento proveniente de um ou vários de seus usuários só se torna disponível para os demais após o devido processo de registro em algum suporte, que é então passível de catalogação. A cópia de registros informacionais na visão da Internet como repositório, para posterior publicação de novo conhecimento obtido, se assemelha às características de um sistema bibliotecário. E de fato, no início foi assim.

A evolução da revolução trazida pela Internet responde pela sigla WWW (World Wide Web – Rede de alcance mundial – ou simplesmente Web). O novo ambiente, explorado com maiores detalhes no capítulo seguinte, enfim permite o

compartilhamento de informações relacionadas e a colaboração *on-line* entre seus usuários. Novas possibilidades trazem ainda novos desafios. Se já se tornara difícil organizar e encontrar informação que pudesse ser útil às necessidades dos usuários em uma rede sem limites nem controle de expansão, o que dizer desta mesma rede dotada de conexões informacionais e possibilidades de registros colaborativos de forma imediata?

3 A Web

A partir da infra-estrutura sobre a qual a transmissão de dados na Internet ocorre, descrita no capítulo anterior, é possível definir o que é a Web: “A WWW é uma aplicação e um desmembramento da Internet que permite a difusão e transferência de informações e arquivos multimídia mediante a navegação através de *links* hipertextuais⁴” (ROBREDO, 2005, p. 248).

A Web foi inicialmente concebida por Tim Berners-Lee enquanto trabalhava no CERN (*Conseil Européen pour la Recherche Nucléaire* – Conselho Europeu para a Pesquisa Nuclear, atualmente *European Organization for Nuclear Research* – Organização Européia para a Pesquisa Nuclear) (BERNERS-LEE, 2000). O autor cita que a idéia de desenvolver tal ferramenta surgiu da forma como enxergava os significados que podem ser atribuídos a determinado registro de informação. Segundo Berners-Lee, a informação só faz sentido quando relacionada com outra informação, e depende da forma como este relacionamento é estabelecido. Tal visão já havia sido abordada por Vannevar Bush em seu artigo “As we may think” de 1945, quando ele propôs a construção de uma máquina chamada Memex que criaria e seguiria referências cruzadas entre documentos microfilmados. Uma forma de criar tais relacionamentos entre qualquer tipo de informação seria através da utilização de hipertexto, conforme relatado no experimento Xanadu, proposto por Ted Nelson ainda em 1965. Assim, Berners-Lee afirma que as idéias que levaram à concepção da Web já existiam, mas eram consideradas futuristas em face do estado da tecnologia na época de suas propostas. Com a Internet, surgia a oportunidade de colocá-las em prática.

O projeto WorldWideWeb surgiu no CERN em 1990 como uma forma de permitir o acesso à informação em sistemas documentários desenvolvidos em diferentes linguagens de programação (métodos de escrita de instruções para execução por programas de computador) e executados em plataformas (infra-estrutura composta por computadores – *hardware*, sistemas operacionais que os

⁴ *Links* hipertextuais, *hyperlinks* ou simplesmente *links* são elos entre dois elementos, que podem ser textos, sons, vídeos e outros, na Web. Por permitirem o acesso direto, ou navegação, entre esses elementos sem a necessidade de obedecer a uma ordem seqüencial (como nos livros, onde para atingir determinada página é necessário passar pelas anteriores), são um tipo de hipertexto.

controlam, outras aplicações – *software*, e redes de comunicação) distintas. Quando se tornava necessário permitir o acesso de informação entre sistemas nessas condições, era necessário que um fosse migrado para a plataforma de outro, ou fosse construída alguma solução específica de integração.

A ferramenta Web não propunha nada mais do que um padrão para comunicação de forma a resolver o problema. Para tais sistemas deveriam ser construídas soluções de integração Web para permitir a comunicação entre si. Porém, o padrão Web foi construído de forma a ser simples de implementar e ao mesmo tempo poderoso o suficiente para permitir a integração de quaisquer registros e tipos de informação independente de plataforma, sem necessidade de um controle central que pudesse limitar a escalabilidade ou mesmo a escolha dos relacionamentos, de forma análoga a conexões cerebrais, evitando dificuldades no crescimento da rede de conexões. O nome em inglês Web (em português, teia) surgiu pela característica de que qualquer informação poderia ser acessada através de outra pois estariam todas interligadas em uma grande teia, pela qual o usuário caminharia, ou navegaria.

As conexões entre as informações seriam realizadas através de hipertexto, permitindo a não linearidade de acesso e a descentralização de controle. Isso quer dizer que determinada informação poderia ser acessada diretamente a partir de qualquer parte de outra, sem necessidade de acesso seqüencial. Da mesma forma, não era necessário registrar tal relacionamento em lugar nenhum, a não ser na informação que originou a conexão através de um elo, ou *link*. Unidirecionais, os *links* apenas apontam da origem para o destino, sem o caminho inverso. Com isso, a informação destino não sabe sobre a existência de registros de informação que apontam para ela. Assim como permite que estas ligações sejam quebradas facilmente (com a exclusão ou mudança da informação destino sem a respectiva desativação do *link*, por exemplo), isso possibilitou e ainda possibilita o crescimento exponencial das conexões e da própria Web.

O padrão Web era composto basicamente por três elementos:

- Um protocolo de comunicação que padroniza a troca de dados entre diferentes plataformas e permite a compreensão de hipertextos e a navegação através de *links*, denominado HTTP (*Hypertext Transfer Protocol* – Protocolo de transferência de hipertexto);

- Um esquema de endereçamento e identificação aplicável a todo registro de informação conectado na rede de comunicação de forma a torná-lo único, denominado então de UDI (*Universal Document Identifier* – Identificador universal de documento); e
- Uma linguagem de marcação que permitisse a descrição de informação na Web, bem como dos próprios *links*, denominada de HTML (*Hypertext Markup Language* – Linguagem de marcação de hipertexto), baseada na família SGML (*Standard Generalized Markup Language*). Linguagens de marcação, ou etiquetagem, permitem a classificação dos elementos em um registro de informação, distinguindo, por exemplo, texto simples, *links*, imagens, etc.

Todos estes elementos se mantêm como base da Web. Porém com melhorias, como novas versões do protocolo HTTP e da linguagem HTML, ou mudanças de conceitos, e em meio a novos padrões.

Ainda nos primórdios da Web, em discussões na IETF (*Internet Engineering Task Force*), comunidade de discussão sobre a arquitetura Internet, o termo *Universal Document Identifier* foi considerado demasiado audacioso para uma aplicação iniciante como a Web (BERNERS-LEE, 2000). Por isso, ao invés das palavras universal, documento e identificador, foram utilizadas uniforme, recurso e localizador: *Uniform Resource Locator* (URL), sigla que se tornou sinônimo dos endereços Web até hoje, mesmo que o esquema já seja denominado atualmente de *Uniform Resource Identifier* (URI).

A linguagem HTML é um subconjunto fixo da família de linguagens SGML com algumas novas etiquetas de classificação (ou *tags*). A família SGML permitia que qualquer um criasse o seu próprio conjunto de *tags*. A falta do suporte à flexibilidade foi criticada quando do surgimento da Web, mas hoje se reconhece que o conjunto limitado de *tags* presentes na HTML permitiu a facilidade de compreensão e adoção por desenvolvedores em todo o mundo, culminando no crescimento exponencial de documentos Web a cada ano. Mesmo assim, a linguagem HTML foi revisada e expandida algumas vezes. Além disso, uma nova linguagem, denominada XML (*eXtensible Markup Language* – Linguagem de marcação extensível) que permite a flexibilidade original da família SGML para

criação de conjuntos personalizados de *tags*, foi criada e estabelecida como outro padrão disponível na Web.

Como padrão de comunicação na própria Internet, a Web permitia não apenas resolver problemas de incompatibilidade entre plataformas, mas o fazia sem demandar mudanças nos formatos originais de armazenamento das informações, uma vez que apenas eram criadas conexões entre elas. De fato, sem a Web, a Internet se apresentava meramente como um meio de disponibilizar e transportar informações sem se importar com o relacionamento delas entre si. A informação não persistia no meio (no caso, a Internet), ela se encontrava espalhada em diversos servidores de rede ao redor da Internet. Já o espaço Web era universal: “A idéia de universalidade era chave: a revelação básica era que um espaço informacional poderia incluir todos os outros, provendo grande poder e consistência” (BERNERS-LEE, 2000, p. 33, tradução nossa).

Entretanto, não se pode esquecer que a Web é uma aplicação que utiliza a infra-estrutura da Internet para existir. Com isso, o papel dos servidores de rede continua em destaque. A diferença é que, ao invés de serem invocados por códigos de protocolos Internet de difícil compreensão, passam a deter URIs criados para facilitar a difusão de endereços e de informação através de citações. Como exemplo, é possível acessar o portal do Governo Brasileiro por dois endereços: “<http://www.brasil.gov.br>” ou “<http://161.148.24.13>”. Em ambos há a descrição do uso do protocolo HTTP. Mas apenas no primeiro existe um endereço que permite acesso a um servidor do governo brasileiro (na verdade, do Serviço Federal de Processamento de Dados – SERPRO) utilizando uma linguagem mais palatável, que é a URI. O segundo endereço consiste de um endereçamento de protocolo IP (*Internet Protocol*), um dos padrões de dados que compõem a infra-estrutura da Internet, mas desenvolvido separadamente da Web. É necessário registrar o URI para determinada entidade em um serviço hierárquico de controle na Internet, de forma a permitir o endereçamento correto ao servidor ou conjunto de servidores de rede da entidade.

Uma vez utilizando os padrões de comunicação Web e uma aplicação para se tornar servidor Web, os servidores de rede passam a permitir o acesso às informações (páginas Web, documentos, etc.) que contêm. Estas informações passam a ter um URI próprio (normalmente complementares ao URI do servidor, que depois de registrado passa também a ser chamado de nome de domínio do servidor

por permitir acesso ao seu conjunto de informações) e que permitem o acesso direto a elas ou através dos relacionamentos construídos em outras páginas ou documentos.

O acesso às páginas Web define a necessidade de criação de outro tipo de programa de computador. Denominados navegadores (em inglês, *browsers*), estes programas têm como intuito a leitura e interpretação de código HTML a fim de permitir ao usuário visualizar texto, imagem e outros elementos conforme relacionados nas páginas Web. Cada vez que o URI de determinada página ou *link* para ela é utilizado, o navegador acessa o servidor em que a página está localizada e a recupera, montando a sua visualização em um monitor, de forma dinâmica. Isso quer dizer que as páginas são abertas e visualizadas de acordo com a demanda do usuário de um navegador. Caso elas sofram modificações, estas se tornam imediatamente disponíveis para qualquer navegador que a acesse a partir da sua alteração.

Entre as funcionalidades inicialmente previstas por Berners-Lee para os navegadores Web, a única não descrita no parágrafo anterior é a de permitir a edição das páginas Web remotamente. Berners-Lee incentivou desde o advento da Web a inclusão da capacidade de edição de páginas nos navegadores de forma a salvar as alterações diretamente nos servidores que as mantivessem. Diferentemente de ferramentas que permitissem a edição local de páginas Web para posterior publicação, o desejo do autor era a possibilidade de publicação imediata. Embora alguns navegadores⁵ com essa característica tenham sido construídos principalmente em meios acadêmicos, eles não se tornaram difundidos fora destes meios e nem se mostraram comercialmente interessantes para desenvolvimento:

Sem um editor de hipertexto, as pessoas não teriam as ferramentas para realmente usar a Web como um meio colaborativo íntimo. Navegadores as deixariam encontrar e compartilhar informação, mas elas não poderiam trabalhar juntas intuitivamente. Parte do motivo, imaginei, era que colaboração requeria muito mais uma mudança social em como as pessoas trabalhavam. E parte disso [do motivo] era que editores eram muito mais difíceis de escrever [desenvolver] (BERNERS-LEE, 2000, p. 57, tradução nossa).

⁵ O navegador Navipress lançado pela empresa Navisoft em fevereiro de 1994 era um dos poucos navegadores comerciais com a capacidade de editar páginas Web. A empresa foi comprada pela America Online (AOL), que tornou o Navipress o navegador padrão para seus usuários até que em 1996, após acordo com a Microsoft, foi substituído pelo Internet Explorer, vindo a ser descontinuado.

Dotada de um navegador e de uma aplicação para servidor Web primitivos, da primeira versão da linguagem HTML e do protocolo HTTP, no final de 1990 a equipe de Berners-Lee registrou e colocou no ar o primeiro servidor Web da história. Identificado com a URI “info.cern.ch”, o servidor continha páginas com as anotações do projeto WorldWideWeb e especificações dos padrões HTTP, URI e HTML. O primeiro sistema integrado à Web foi a agenda de telefones do CERN. Apesar de parecer uma aplicação simples, o acesso à agenda de telefone era extremamente necessário por sistemas em diferentes plataformas, o que demandava grande esforço de sua equipe de manutenção para construir interfaces de comunicação entre a agenda e estes sistemas, realizado a partir de então diretamente pela Web.

O primeiro servidor Web fora da Europa foi montado nos laboratórios da *Stanford Linear Accelerator* (SLAC) em Palo Alto, Estados Unidos, com o intuito de promover acesso mundial à base bibliográfica interna de documentos produzidos pelo centro.

Um passo importante para o avanço da Web foi dotar o protocolo HTTP de acesso às informações trafegadas na Internet através do protocolo FTP (*File Transfer Protocol* – Protocolo de Transferência de Arquivos). Como o próprio nome diz, era o padrão responsável pela transferência de arquivos pela Internet, sendo de grande uso até os dias de hoje. Através do protocolo FTP, notícias eram trafegadas e grupos de discussão (*newsgroups*) trocavam mensagens. E foi utilizando um destes grupos, o “alt.hypertext”, de entusiastas por hipertexto, que o advento da Web foi anunciado e comentado amplamente como aplicação de hipertexto.

A partir daí as funcionalidades proporcionadas pela Web foram alvo de elogios, críticas, sugestões e desenvolvimento efetivo de melhorias de forma a constatar que os próprios usuários de Internet direcionaram os desdobramentos da Web. Tratou-se de um processo aberto de desenvolvimento social de aplicações para Web (BERNERS-LEE, 2000), onde ela era debatida em grupos de discussão, basicamente com envolvimento de pessoas e não de empresas com intuito comercial.

O número de servidores Web e a porcentagem de tráfego Internet direcionada para a Web eram relativamente pequenos, mas cresciam exponencialmente. Em março de 1993 o tráfego Web equivalia a 0,1% do tráfego total da Internet. Em setembro, atingia 1%. Em dezembro, 2,5% (BERNERS-LEE, 2000). No mesmo ano, a *University of Minnesota*, detentora dos direitos do Gopher, outra aplicação para

compartilhamento de informação na Internet, similar, mas então mais popular que a Web, decidiu cobrar uma licença anual para certas classes de usuários que utilizassem a aplicação. Isso fez com que caísse em desuso no meio comercial e logo deixasse de ser atrativa também no meio acadêmico. A Web, diferentemente, foi liberada como tecnologia de domínio público geral pelo CERN, culminando na difusão de seu uso em ambos os setores. Surgiram, por exemplo, as redes Web internas, de uso corporativo para visualização de páginas e documentos em empresas.

Um dos primeiros navegadores lançados para a Web foi o Mosaic. Criado por estudantes do NCSA (*National Center for Supercomputing Applications*) localizado na *University of Illinois*, Estados Unidos, ele já era dotado de interface gráfica, o que viria a se tornar o diferencial para uso em massa dos navegadores. O primeiro navegador Web comercial viria a ser lançado pela Netscape, empresa criada pelos estudantes que trabalharam no projeto Mosaic, como Netscape Navigator em abril de 1994. De forma diferente de outros programas de computador, o navegador não era vendido, mas distribuído livremente. A Netscape se tornou mais do que uma empresa de *software*, se tornou uma empresa cujo lucro se basearia na venda de serviços via Web, uma mudança de filosofia comercial. Não coincidentemente, os navegadores se tornaram os produtos tecnológicos de predominância na Web (TAPSCOTT, 1998). O quadro 2 apresenta uma linha de tempo contendo a evolução dos navegadores Web desenvolvidos.

Navegador	Derivado de	Principais versões
WordWideWeb/Nexus	-	0.1 (1991). Descontinuado em 1993.
ViolaWWW	-	1 (1992), 3 (1994). Descontinuado em 1994.
Cello	-	Sem versão definida (1993), 1 (1994). Descontinuado em 1994.
Lynx	-	2 (1993), 2.3 (1994), 2.4 (1996), 2.7 (1997), 2.8 (1998), 2.8.2 (1999), 2.8.4 (2001), 2.8.5 (2004), 2.8.6 (2006), 2.8.7 (2009).
Mosaic	-	1 (1993), 2 (1995), 3 (1996/1997). Descontinuado em 1997.
AMosaic/IBrowse	Mosaic	1 (1993), IBrowse 1 (1996), 2.3 (2003), 2.4 (2006). Descontinuado em 2007.
Internet Explorer	Mosaic	1 (1995), 2 (1995), 3 (1996), 4 (1997), 5 (1999), 5.5 (2000), 6 (2001), 7 (2006), 8 (2009).
Internet Explorer para Mac	Internet Explorer	2 (1996), 3 (1997), 4 (1998), 5 (2000), 5.2 (2002). Descontinuado em 2007.
SimulBrowse/NetCaptor	-	3 (1998), 5 (1999), 6 (2000), 7 (2002), 7.5 (2003). Descontinuado em 2007.
MyIE/MyIE2/Maxthon	-	Sem versão definida (2000), 1 (2004), 1.5 (2005), 2 (2007), 2.5 (2009).

Navegador	Derivado de	Principais versões
IEOpera/Avant Browser	-	7 (2002), 8 (2003), 9 (2004), 10 (2004), 11 (2005), 11.5 (2007).
AOL Browser/ AOL Explorer	-	1 (2005), 1.5 (2006). Descontinuado em 2006.
Mosaic Netscape/ Netscape Navigator	Mosaic	1 (1994), 2 (1996), 3 (1996), 4 (1997), 4.5 (1998), 4.8 (2002). Descontinuado em 2002.
Mozilla	Netscape Navigator	Sem versão definida (1998), 0.6 (2000), 1 (2002), 1.5 (2003). Descontinuado em 2006.
Netscape/ Netscape Browser/ Netscape Navigator	Mozilla	6 (2000), 7 (2002), 8 (2005), 9 (2007). Descontinuado em 2008.
SeaMonkey	Mozilla	1 (2006), 1.1 (2007), 2 (2009).
Firefox	Mozilla	Sem versão definida (2002), 1 (2004), 1.5 (2005), 2 (2006), 3 (2008), 3.5 (2009), 3.6 (2009).
Flock	Firefox	0.5 (2005), 0.7 (2006), 1 (2007), 2 (2008), 2.5 (2009).
Camino	Mozilla	0.5 (2002), 1 (2006), 1.5 (2007), 2 (2009).
K-Meleon	Mozilla	0.5 (2001), 0.8 (2003), 1 (2006), 1.1 (2007), 1.5 (2008).
Galeon	Mozilla	1 (2001), 2 (2005). Descontinuado em 2006.
Epiphany	Galeon	1 (2003), 2 (2006), 2.28 (2009).
WebRunner/HotJava	-	Sem versão definida (1994), 1 (1997), 3 (1999). Descontinuado em 1999.
SlipKnot	-	1 (1994), 1.4 (1995). Descontinuado em 1997.
Konqueror	-	2 (2000), 3 (2002), 3.5 (2005), 4 (2008).
Safari	Konqueror	1 (2003), 2 (2005), 3 (2007), 4 (2009).
OmniWeb	Safari (A partir da versão 5)	1 (1995), 2 (1996), 3 (1999), 4 (2001), 5 (2004), 5.5 (2006), 5.10 (2009).
Google Chrome	-	0.2 (2008), 1 (2008), 2 (2009), 3 (2009), 4 (2010).
Shiira	-	1 (2005), 2 (2007).
iCab	Safari (A partir da versão 4)	1 (1999), 2 (2000), 2.5 (2001), 3 (2006), 4 (2008).
w3m	-	Sem versão definida (1995), 0.2 (2001), 0.3 (2002), 0.4 (2003), 0.5 (2003), 0.5.2 (2007).
Tamaya/Amaya	-	Sem versão definida (1996), 1 (1997), 2 (1999), 3 (2000), 4 (2000), 5 (2001), 6 (2002), 7 (2002), 8 (2003), 9 (2005), 9.5 (2006), 10 (2008), 11 (2008).
Opera	-	Sem versão definida (1994), 1 (1994), 2 (1996), 3 (1997/1998), 3.5 (1998), 4 (2000), 5 (2000), 6 (2001), 7 (2003), 7.5 (2004), 8 (2005), 9 (2006), 9.5 (2008), 9.6 (2008), 10 (2009), 10.1 (2009), 10.5 (2010).
Opera Mini	Opera	1 (2005), 2 (2006), 3 (2006), 4 (2007), 4.1 (2008), 4.2 (2008), 5 (2010).
Arachne	-	1 (1996), 1.7 (2001), 1.9 (2006).
AWeb	-	1 (1996), 2.1 (1996), 3 (1997), 3.4 (2001), 3.5b (2004).
Links	-	0.8 (1999), 0.9 (2000), 2 (2002).
ELinks	Links	0.3 (2002), 0.9 (2003), 0.10 (2004), 0.11 (2005/2006).
Gzilla	-	0.1 (1997), 0.1.7 (1998). Descontinuado em 1999.
Dillo	Gzilla	0 (1999), 0.1 (2000), 0.5 (2001), 0.7 (2003), 0.8 (2004), 0.8.6 (2006), 2 (2008).
NetSurf	-	Sem versão definida (2002), 1 (2007), 1.1 (2007), 1.2 (2008), 2 (2009), 2.1 (2009).

Quadro 2 - Linha de tempo dos navegadores Web (Adaptado de ADEVERIA, 2010)

No mesmo ano foi criada uma organização denominada *World Wide Web Consortium* (também conhecida pela sigla W3C) com o intuito de elaborar e analisar

recomendações para a evolução da Web a fim de atingir o seu potencial pleno e manter a sua interoperabilidade, isto é, a heterogeneidade de plataformas dotadas de acesso a ela. A organização desenvolve recomendações, e não padrões de uso obrigatório, como forma de manter a abertura do processo de desenvolvimento da Web. Estas recomendações, da mesma forma, são de livre acesso. No consórcio são permitidos membros dos setores governamental, educacional e comercial. Empresas deste último participam como forma de sugerir e acompanhar o desenrolar de novas aplicações, colaborando competitivamente para o desenvolvimento da Web e das infra-estruturas que a suportam (DERTOUZOS, 1997). As sedes do W3C ficam no *Massachusetts Institute of Technology* (MIT), nos Estados Unidos, no *European Research Consortium for Informatics and Mathematics* (ERCIM) na Europa e na *Keio University* no Japão. Na comemoração dos 10 anos do W3C, foi publicada uma ilustração que contém os fatos mais relevantes para a Web até o ano de 2004, adaptada no quadro 3.

Web pré-W3C e histórico anterior da Web	
Antes de 1979	<ul style="list-style-type: none"> • 1945: O artigo de Vannevar Bush na Atlantic Monthly descreve um dispositivo foto-elétrico-mecânico chamado Memex, para extensão de memória, que poderia criar e seguir <i>links</i> entre documentos em microficha. • 1960: J. C. R. Liddider publica “Man-Computer Symbiosis”. • 1962: Douglas Englebart publica “Augmenting Human Intellect: A Conceptual Framework”. • 1965: Ted Nelson cunha o termo “Hipertexto” em “A File Structure for the Complex, the Changing, and the Indeterminate”. 20ª Conferência Nacional, New York, <i>Association for Computing Machinery</i>. • 1968: Douglas Englebart demonstra o <i>Online System</i> (NLS) • 1969: <i>Advanced Research Projects Agency</i> licencia a ARPANET para conduzir pesquisa em redes. • 1971: Ray Tomlinson da BBN cria programa de <i>e-mail</i> para distribuir mensagens através de uma rede distribuída. • 1972: Tomlinson expande programa para usuários ARPANET, usando o símbolo “@” como parte do endereço. • 1974: Vint Cerf e Bob Kahn publicam “A Protocol for Packet Network Interconnection”, que especifica em detalhes o projeto de um Protocolo para Controle de Transmissão (<i>Transmission Control Protocol – TCP</i>). • 1978: Parte do TCP publicado separadamente como Protocolo Internet (<i>Internet Protocol – IP</i>).
Antes de 1989	<ul style="list-style-type: none"> • 1980: Como consultor no CERN, Tim Berners-Lee escreve um programa, “Enquire-Within-Upon-Everything”, que permite criar <i>links</i> entre nós arbitrários. • 1984: Paul Mockapetris introduz o <i>Domain Name Systems</i> (DNS).
1989	<ul style="list-style-type: none"> • Março: Tim Berners-Lee circula “Information Management: A Proposal” para comentários no CERN.
1990	<ul style="list-style-type: none"> • Maio: Tim Berners-Lee publica a versão 2 de “Information Management: A Proposal”. • Final de 1990: Inicia o desenvolvimento do primeiro navegador (chamado “WorldWideWeb”, editor, servidor e navegador em modo de linha. Culmina na primeira comunicação cliente-servidor Web na Internet em dezembro de 1990.
1991	<ul style="list-style-type: none"> • Dezembro: Conferência de Hipertexto em 1991 em San Antonio, Texas (EUA). O artigo

	<p>de Tim Berners-Lee sobre a Web é aceito apenas para a sessão de pôsteres.</p> <ul style="list-style-type: none"> • Servidores Web: 1
1992	<ul style="list-style-type: none"> • Dezembro: Primeiro servidor Web fora da Europa estabelecido na <i>Stanford University</i>. • Servidores Web: 10
1993	<ul style="list-style-type: none"> • Janeiro: Número de navegadores aumenta; incluindo Midas, Erwise, Viola e Samba. • Março: NCSA lança a primeira versão alfa do Mosaic para X Windows. • Abril: CERN concorda em permitir que qualquer um use o protocolo e código Web livremente. • Junho: Dale Dougherty da O'Reilly organiza o <i>WWW Wizards Workshop</i> em Cambridge, Massachusetts, EUA. • Novembro: Em uma conferência em Newcastle, Reino Unido, Tim Berners-Lee discute sobre o futuro da Web com David Gifford do MIT, que sugere que Tim entre em contato com Michael Dertouzos. • Servidores Web: 50
1994	<ul style="list-style-type: none"> • Mark Andreessen e colegas deixam a NCSA para formar a Mosaic Communications Corp., que depois se tornou Netscape. • Sistemas <i>dial-up</i> tradicionais (CompuServe, AOL, Prodigy) começam a prover acesso à Internet.
W3C	
1994	<ul style="list-style-type: none"> • Fevereiro: Tim Berners-Lee encontra Michael Dertouzos em Zurich para discutir a possibilidade de constituir uma nova organização no MIT. • Abril: Alan Kotok, então na DEC, visita o CERN para discutir a criação do Consórcio. • 1º de outubro: W3C criado. <p>Ano de início das atividades W3C atuais</p> <ul style="list-style-type: none"> • Conferências Web: Genebra (Junho) e Chicago (Outubro) • Servidores Web: 623
1995	<ul style="list-style-type: none"> • Abril: INRIA se torna a sede do W3C na Europa. • Junho: W3C sedia o primeiro Workshop, em Classificação de Conteúdo, criando o PICS. • Atividades W3C: Gráficos, <i>Hypertext Markup Language</i> (HTML) e estilo. • Conferências Web: Darmstadt (Abril) e Boston (Dezembro). • Servidores Web: 10.022.
1996	<ul style="list-style-type: none"> • Junho: Em resposta ao "Peabody Meeting", W3C forma o Process ERB. • Setembro: A <i>Keio University</i> se torna a sede do W3C na Ásia. • Atividades W3C: Matemática, <i>Extensible Markup Language</i> (XML). • Conferência Web: Paris (a maior conferência até a data da fonte). • Servidores Web: 100.000.
1997	<ul style="list-style-type: none"> • Novembro: W3C lança programas Offices. • Atividades W3C: <i>Document Object Model</i> (DOM), política de patente, privacidade, multimídia sincronizada, <i>Web Accessibility Initiative</i> (WAI). • Conferência Web: Santa Clara. • Servidores Web: 603.367.
1998	<ul style="list-style-type: none"> • O W3C cria o Conselho Consultivo (<i>Advisory Board – AB</i>). • Atividades W3C: Internacionalização. • Conferência Web: Brisbane. • Servidores Web: Aproximadamente 1,6 milhão.
1999	<ul style="list-style-type: none"> • Atividades W3C: Navegador de voz. • Conferência Web: Toronto. • Servidores Web: Aproximadamente 3,7 milhões
2000	<ul style="list-style-type: none"> • Atividades W3C: URI. • Conferência Web: Amsterdã. • Servidores Web: Aproximadamente 9,5 milhões.
2001	<ul style="list-style-type: none"> • Julho: W3C libera a licença <i>Technical Architecture Group</i> (TAG). • Atividades W3C: Independência de dispositivos, Web semântica, gerenciamento de chave XML, garantia de qualidade. • Conferência Web: Hong Kong. • Servidores Web: Aproximadamente 26 milhões.
2002	<ul style="list-style-type: none"> • Atividades W3C: Interação multimodal, <i>Web services</i>.

	<ul style="list-style-type: none"> • Conferência Web: Honolulu. • Servidores Web: Aproximadamente 36 milhões.
2003	<ul style="list-style-type: none"> • Janeiro: ERCIM substitui o INRIA como sede do W3C na Europa. • Fevereiro: W3C adota a <i>Royalty-Free Patent Policy</i> (Política de patente de direito livre). • Atividades W3C: XForms. • Conferência Web: Budapeste. • Servidores Web: Aproximadamente 35 milhões.
2004	<ul style="list-style-type: none"> • Atividades W3C: Formatos de documentos compostos. • Conferência Web: Nova Iorque. • Servidores Web: Mais de 46 milhões.

Quadro 3 - Linha de tempo de evolução da Web e do W3C (Adaptado de WORLD WIDE WEB CONSORTIUM, 2004c)

O W3C descreve a arquitetura da Web da seguinte maneira: “[...] é um espaço informacional no qual os itens de interesse, denominados recursos, são identificados por identificadores globais chamados *Uniform Resource Identifiers* (URI)” (WORLD WIDE WEB CONSORTIUM, 2004a, tradução nossa). Berners-Lee (2000, p. 76, tradução nossa) complementa com o propósito da Web: “Ser um meio de hipertexto único, universal e acessível para compartilhar informação”. Define-se a estrutura da Web como um espaço informacional, ou seja, que contém todos os itens de interesse de informação (páginas, imagens, arquivos em geral), ao redor uns dos outros.

A Web, portanto, engloba o conjunto de tecnologias que permite que usuários (humanos ou não) criem, publiquem ou acessem recursos de forma interligada. Algumas das características destas tecnologias são o uso de *links* para interligação dos recursos; a alocação de um identificador único (URI) e não ambíguo para cada recurso, de forma a manter a unicidade de cada ponto de acesso em toda a Web; o uso de programas específicos para navegação e visualização de páginas e outros arquivos; além do emprego de protocolos e linguagens de programação específicos para associar texto e outros formatos de arquivos entre si.

Entretanto, um objetivo pretendido desde o advento da Web somente seria explorado mais recentemente. A colaboração imediata entre usuários nas páginas Web, funcionalidade almejada por Berners-Lee para implementação nos navegadores viria a se tornar não apenas mais um serviço ou aplicação, mas uma filosofia que atualmente direciona os caminhos da Web, abordada com maiores detalhes no capítulo seguinte.

3.1 A Web social

Os rumos da sociedade foram decisivamente atingidos pela existência da Internet e da World Wide Web. O ambiente criado pelo compartilhamento de informação e conhecimento viria a personificar o ciberespaço, conforme imaginado pelo filósofo Pierre Lévy em uma descrição mais filosófica do que a utilizada no capítulo anterior: “Espaço móvel das interações entre conhecimentos e conhecedores de coletivos inteligentes desterritorializados” (LÉVY, 2007, p. 29). Trata-se, portanto, de um meio criado para permitir relações entre as pessoas, conter uma inteligência coletiva, de forma distribuída, passível de valorização, coordenável e acima de tudo utilizável para o benefício da humanidade. Não por coincidência, são características ressaltadas por Vannevar Bush ainda em 1945.

Em sua análise sobre a nova geração de seres humanos nascidos na sociedade da informação, Tapscott (1998) resgata o modelo de difusões de idéias de Michel Foucault, denominado “*Web of Power*”, anterior ao desenvolvimento da World Wide Web. Segundo este modelo, a discussão sobre qualquer assunto ganha em conhecimento (e conseqüentemente em poder) na medida em que novos indivíduos participam dela. De forma similar, Tapscott sustenta que na Web os participantes obtêm conhecimentos pela interação com outros usuários.

Como base para a construção de um ciberespaço, Lévy destaca a interação sem fronteiras, o reconhecimento e o enriquecimento mútuo que tal interação tem o poder de causar entre os envolvidos. Para permitir a coordenação destas relações em tempo real, o filósofo aponta o uso obrigatório de TICs digitais⁶: “as redes de comunicação e as memórias digitais englobarão em breve a maioria das representações e mensagens em circulação no planeta” (LÉVY, 2007, p. 103). A pergunta a se fazer é se este “em breve” já não chegou.

O autor menciona claramente o hipertexto como uma estrutura de criação, navegação e relação social de conhecimento, o que permite associar o ciberespaço à Web. Trata-se de uma construção simples, mas ao mesmo tempo capaz de

⁶ Por tecnologias digitais entenda-se aquelas que utilizam processamento e transmissão de sinais digitais, obedecendo à lógica binária para a qual existem apenas dois valores (0 e 1) que, combinados, servem para representar qualquer informação. Os computadores são os maiores exemplos destas tecnologias.

representações complexas de conhecimento, similares a conexões cerebrais (DERTOUZOS, 1997). Para um ser humano, determinado odor pode remeter a uma lembrança de um lugar totalmente dissociado de seu contexto atual. Similarmente, o hipertexto pode ser o atalho entre conhecimentos distintos, para carregar percepções, sentimentos, lembranças, referências e saberes em um ciberespaço, de forma desorganizada e sem classificação. O hipertexto armazena essas representações para um número ilimitado de indivíduos, permitindo a combinação e a reutilização delas em escala não mensurável, sendo um dos ideais por trás do funcionamento da Web e uma das fontes de sua vitalidade, antagonizando com a sua existência baseada em tecnologias artificiais. A utilização do hipertexto permite que a leitura de qualquer registro de informação carregue um potencial de conexão com outros registros de informação (LÉVY, 1999). Por isso, Castells (1999) define a Web como organizada por informação, e não apenas por localização.

A definição da teia de elementos ligados por hipertexto como uma rede viva, em constante crescimento e cada vez mais impactante sobre o cotidiano, depende diretamente da intensa interação de seus usuários entre si e com a própria Web. Da mesma forma como comportamentos sociais são replicados na rede, novos são forjados a partir dela. Hoje não é mais possível pensar o uso do ambiente dito como virtual sem considerar as possíveis implicações éticas, legais e sociais na vida real da sociedade da informação descrita no capítulo anterior em diferentes contextos. Conforme citado por Berners-Lee (2000), a Web é verdadeiramente uma criação mais social do que técnica, sendo composta por conteúdo e conexões entre ele, e não apenas por *software*. Seu crescimento se deu como uma forma de movimentação social, com o crescente número de usuários e suas respectivas contribuições.

A visão da Web como um hipertexto sem dimensões definidas permite uma forma de acesso único a uma imensidão de arquivos espalhados. Seu crescimento se dá de forma dinâmica e caótica, sem organização, tornado inalcançáveis as idéias de descrição finita e da totalidade de domínio de uma mesma idéia, visto que as opiniões e abordagens sobre ela se multiplicam em diversos pontos do sistema em profusão e desordem. Entretanto, por outro lado, a existência de conexões hipertextuais previne que determinada informação possa ser vista desligada de um contexto sob o qual se encontra conectada, levantando a idéia de uso dos *hyperlinks*

como agentes de seleção organizadora, de estruturação e de filtragem (LÉVY, 1999).

Em um ambiente criado para não oferecer restrições etárias para uso, cria-se também um espaço onde o exercício de pensamento crítico e de julgamento das informações tornadas disponíveis é livre (TAPSCOTT, 1998). Isso leva à quebra de estruturas tradicionais de autoridade e hierarquia da sociedade, sendo capaz de democratizar acesso e opiniões, mas também levantando dúvidas e receios em gerações criadas antes do advento da Web em relação aos seus desdobramentos.

O próprio desenvolvimento das tecnologias sobre as quais a Web funciona se deu de forma diferenciada. O que levou à criação do W3C foi a necessidade de integrar em um fórum único entidades diversas como governos, a academia e o mercado corporativo. Com interesses tão distintos, a evolução das recomendações da Web não teve como não se dar em um clima competitivo. Porém, diferentemente do que se pode imaginar, o fluxo financeiro de investimentos e da real possibilidade de retorno deles em virtude das diversas oportunidades do ambiente aceleraram o consenso de novas iniciativas ou o descarte daquelas em que o consenso não podia ser obtido (BERNERS-LEE, 2000).

Em um âmbito mais geral, o crescimento do número das conexões hipertextuais na Web ocorre por um mecanismo de confiança e colaboração de seus usuários. Um *link* pode ser utilizado como uma forma de citação a um conteúdo que se deseje relacionar como referência. Já o conteúdo diretamente publicado por alguém o torna exemplo de autoria, com a respectiva responsabilidade pela informação. Isso apenas exemplifica o papel da Web como representação de relações humanas. Entretanto, mais recentemente, a característica colaborativa, que permite a interação entre diversos indivíduos em tempo real e sem respeitar fronteiras, da Web levou à criação de um novo modelo de autoria coletiva, segundo o qual todo conteúdo é livre para ser expandido, divulgado e julgado pelos seus próprios usuários.

Sob um ponto de vista filosófico, Lévy (1999) aponta três fatores que orientaram o crescimento inicial do ciberespaço (entendido especificamente aqui como a Web): a interconexão, vertente sob a qual existe a necessidade humana de fugir do isolamento, e se caracterizando tecnicamente como o objetivo da comunicação universal, onde cada ser humano se encontra permanentemente em contato com o meio de comunicação através de vários dispositivos; a criação de

comunidades virtuais, agregando pessoas com interesses em comum e gerando novas formas de expressar opinião pública; e a inteligência coletiva, que preza pelo compartilhamento constante de conhecimento.

Exemplos desta criação coletiva são os *softwares* de código aberto (em inglês, *open source*). Toda aplicação computacional é composta de algoritmos, mencionados anteriormente, também chamados de código da aplicação. Eles contêm a inteligência de quem criou a aplicação e normalmente são explorados como vantagem competitiva, sendo, portanto, mantidos em sigilo. Com o advento da Web, porém, alguns indivíduos acreditaram que a chave para um melhor produto poderia residir no emprego do conhecimento de usuários nele interessados. E assim passaram a divulgar na rede os códigos desses produtos, de forma que qualquer um pudesse acessá-los, realizar melhorias e mesmo submetê-las para que outras pessoas pudessem delas usufruir. Como resultados deste padrão, existem hoje aplicativos como o navegador Web Mozilla Firefox (sucessor do Mosaic e do Netscape Navigator, cujo código foi aberto em janeiro de 1998) e o servidor Web mais utilizado mundialmente, o Apache (surgido a partir de melhorias de código aberto de um programa servidor desenvolvido e posteriormente abandonado pelo NCSA).

A partir do desenvolvimento de *software*, o padrão colaborativo se estendeu para a criação de conteúdo. Isso reflete ainda mais o caráter de confiança das relações na Web, para aqueles que aceitam divulgar seu conhecimento ou expandir o que foi tornado disponível por outro, sem a necessidade de reconhecimento individual. Este é o assunto abordado com maior ênfase no próximo capítulo.

Com a Web notam-se avanços na aproximação das pessoas, agregadas em comunidades com interesses comuns e vínculos de socialização da informação, não apenas para compartilhamento de conhecimento, mas também na execução de atividades efetivamente laborais.

Existem ganhos também em conveniência. Sítios (conjunto de documentos inter-relacionados e acessáveis a partir de um mesmo endereço Web) de comércio eletrônico foram estabelecidos na Web, não apenas para divulgar as características de seus bens ou serviços (normalmente explorando seus benefícios frente à concorrência) e auxiliar o consumidor, no conforto da sua sala, em sua decisão de adquiri-los, mas também para oferecer a venda direta deles. Surge uma nova forma de divulgação e propaganda, que com o desenvolvimento da Web passou da escala

meramente informacional para a transacional, oferecendo maiores experiências de interação e de relacionamento com seus usuários. Isso sem contar com avanços na forma de apresentação e a abrangência de outros serviços e meios de comunicação, paulatinamente absorvidos ou integrados à Web. Notícias, vídeos e música sob demanda, acesso a exposições de arte, a transmissão de rádio, TV e de ligações telefônicas *on-line* são algumas das novas possibilidades.

Campos específicos como a educação também exibem benefícios. As conexões hipertextuais que caracterizam a própria Web contêm informação que, se devidamente analisada, pode servir para descobrir novo conhecimento, da mesma forma que as referências em um artigo acadêmico. A divulgação direta da informação pode facilitar desde o acompanhamento escolar de jovens pelos seus responsáveis, até a composição de bancos de conhecimento e o desenvolvimento de estruturas de aprendizado a distância, normalmente caracterizadas por ambientes não apenas de acesso à informação, mas também de colaboração. Estes, por sua vez, podem levar à revisão dos processos educacionais e institucionais que por ora se apresentam defasados diante da evolução da sociedade, mantendo os estudantes engajados em seu uso e difundindo valores democráticos (TAPSCOTT, 2010a; 2010b; 2010c).

Para instituições governamentais existe a possibilidade de transparência para divulgação de seus atos, visando a facilitar a fiscalização do cumprimento de suas tarefas. Outras formas mais avançadas de interação com o público dizem respeito a receber críticas, reclamações, sugestões e elogios. Além disso, a tomada de preços e o emprego de licitações eletrônicas para aquisição de bens e serviços permitem a redução de custos e a realização de processos com maior lisura. Em suma, a Web provê diversas facilidades para incrementar a eficiência governamental.

As empresas também podem ser beneficiadas com ganhos de produtividade e eficiência. O compartilhamento de informação *on-line* permite o fortalecimento dos relacionamentos corporativos internos, entre empregados e entre filiais, e externos, com clientes ou fornecedores. Como exemplo do vínculo interno, trabalhadores de uma mesma empresa lotados em filiais espalhadas pelo globo, ao participarem de comunidades práticas, são capazes tanto de transmitir experiências quanto de cooperarem no cumprimento de tarefas ou na geração de novo conhecimento para um fim comum. Trata-se de um estímulo ao aprendizado, à descentralização, à colaboração e à abolição de fronteiras. Além disso, a empresa pode criar uma Web

própria e de uso restrito (normalmente denominada de Intranet) para comunicação de informação e gestão de conhecimento. Ou mesmo abrir este ambiente para acesso externo, agregando maior valor a portais que não sejam apenas de comércio eletrônico ou de prestação de serviços.

Entretanto, nem só de benefícios é feita a Web. Da mesma forma como reflete ou ajuda a desenvolver características sociais positivas, o ambiente apresenta e majora pontos negativos da vida em sociedade. Efeitos como a divulgação livre de pornografia (acessível para um público indevido, o infantil), o emprego da censura sobre o conteúdo disponível para determinado país, a falta de privacidade gerada pelo descontrole sobre o que pode ser publicado a respeito de qualquer um ou mesmo pelo emprego de tecnologias desenvolvidas com motivações comerciais questionáveis, más práticas publicitárias como o envio em massa de mensagens não autorizadas (*spam*) e o desrespeito à autoria (que apesar da possibilidade de criação coletiva citada acima, não pode ser negada a ninguém que a desejar) são alguns dos exemplos da grande escala, da falta de coordenação e da diversidade de motivações e experiências de seus usuários (MANNING, RAGHAVAN, SCHÜTZE, 2009).

Em relação ao acesso à informação via Web, tão importante como permitir a publicação descentralizada de conteúdo, o que sugere poder representativo e democratização da informação, é necessário prover o devido acesso ao ambiente como forma não só de minimizar a exclusão social, mas impedir o seu crescimento, além da própria orientação de como utilizar o ambiente para usufruir de seus benefícios, não relegando à simples existência da tecnologia o papel educativo. Como inclusão social, entenda-se o processo de inserção de cidadãos na sociedade da informação descrita no capítulo anterior, de forma a usufruir dos benefícios do emprego da informação para melhoria da qualidade de vida (PROGRAMA SOCIEDADE DA INFORMAÇÃO, 2000).

O acesso à informação na Web influencia a redução da exclusão social com o combate à exclusão digital, série de fatores sócio-econômicos e culturais que fazem com que existam tanto pessoas sem acesso à Web quanto aquelas que tenham acesso mas não saibam exprimir suas necessidades informacionais e/ou não saibam utilizar as informações obtidas de forma útil (ROBREDO, 2005). Da mesma forma como é importante não apenas evidenciar o papel da tecnologia na melhoria das condições de vida em detrimento das mudanças culturais que ela proporciona, é

necessário prover mais que o acesso à infra-estrutura tecnológica, mas principalmente o uso do conhecimento contido na rede. Para tanto, se torna necessário o estímulo ao acesso, à fluência, à motivação e à integração destes usuários às ferramentas digitais. A mesma Web criada para aproximar pessoas deve ter como preocupação não isolar aquelas sem acesso e permitir que todas possam usufruir e contribuir para a inteligência coletiva. Isso é um problema que continua em aberto (LÉVY, 1999).

Entre os fatores causadores de exclusão digital destacam-se os mais óbvios, tais como falta de condições financeiras para adquirir ferramentas e serviços para acesso à Web, educacionais como analfabetismo, inclusive funcional (onde o indivíduo, apesar de conhecer letras e números, é incapaz de interpretar textos) e a ausência de competências como o analfabetismo digital (incapacidade de utilizar as tecnologias à contento). Entretanto, existem outros menos explorados, mas também preocupantes, como restrições tecnológicas a regiões com menor desenvolvimento, censura por parte de organizações governamentais e também a dificuldade de acesso à informação por pessoas portadoras de necessidades especiais.

A acessibilidade se tornou uma vertente com amplo estudo na Web atualmente. O W3C criou um projeto chamado *Web Accessibility Initiative* (Iniciativa de Acessibilidade Web) com o intuito de criar recomendações para a representação e organização de conteúdo na Web possibilitando formas distintas de leitura por parte de navegadores especialmente criados para este fim. Por exemplo, a leitura em voz alta da informação contida em uma página para um deficiente visual. No Brasil, o Decreto 5.296, de 2 de dezembro de 2004 regulamentou normas e critérios para a promoção da acessibilidade de pessoas portadoras de deficiência para, entre outros pontos, o acesso à informação e à comunicação (BRASIL, 2004). Assim, todo conteúdo Web da administração governamental deve ser publicado buscando garantir o acesso a este público.

Outros problemas de acesso à informação podem ser provenientes da própria (des)organização da informação em sítios Web. Isso quer dizer que, por mais que um usuário tenha acesso à Web e saiba utilizar suas ferramentas, ele pode não encontrar a informação desejada por um problema de usabilidade da página. A organização de informação em sítios Web de forma a torná-la mais útil e amigável se tornou alvo de outra vertente da AI denominada de Arquitetura da Informação para a Web (MORVILLE; ROSENFELD, 2006; WYLLYS, 2000).

A superação dos obstáculos ao acesso traça o futuro da Web direcionando para uma aproximação da representatividade de seus usuários com a da sociedade em geral. Hoje já existem vários serviços oferecidos no ambiente real disponíveis também no ambiente virtual, como bancos, comércio, informações sobre governo e aprendizado eletrônico (ROBREDO, 2005), além do próprio uso da Web como veículo de comunicação. Um exemplo foi a eleição de Barack Obama para a presidência do Estados Unidos da América, utilizando uma estratégia de campanha baseada amplamente em meios Web.

Os próximos passos já estão sendo dados. Um relatório da Nielsen Company, que realiza pesquisas sobre audiência e publicidade na Web, destaca que cada vez mais os sítios Web buscam enriquecer a experiência conectada (*on-line*) de seus usuários. Estes, por sua vez, apresentam cada vez maior interesse em transmissão de vídeos e nas chamadas mídias ou redes sociais (sítios como Facebook, MySpace, Orkut e Twitter), onde podem construir redes de relacionamento com outros indivíduos a fim de compartilhar interesses ou algum tipo de informação, e que em fevereiro de 2009 ultrapassaram o correio eletrônico (*e-mail*) em alcance mundial. Além disso, utilizam cada vez mais dispositivos móveis para acesso à rede, em virtude da melhoria da tecnologia disponível (telefones celulares inteligentes, ou *smartphones*, e a própria velocidade de acesso) e da diminuição de seus custos (THE NIELSEN COMPANY, 2009).

Outra questão abordada pelo relatório diz respeito à mudança constante de interesses dos usuários. Anteriormente, estes utilizavam prioritariamente sítios com acesso a diferentes tipos de conteúdo (também denominados portais) de forma a centralizar a busca pela informação desejada. O movimento atual aponta para o uso de sítios de conteúdo específico, que em tese podem vir a apresentar visões mais aprofundadas sobre o assunto desejado.

Dois problemas mencionados quando da abordagem da explosão informacional persistem na Web: a dificuldade de encontrar informação efetivamente útil em meio à grande quantidade de informação publicada na Web e a concentração de grande parte dos acessos em um número limitado de registros informacionais.

O primeiro problema pode ser entendido pela expansão da cibercultura prevista por Lévy. Ela se dá em meio ao crescimento caótico de suas conexões, porém isso gera um conflito de interesses entre aquele que torna disponível informação e aquele que busca a informação. Grande quantidade de informação

inútil ou replicada está disponível na Web, tornando difícil a localização de conteúdo efetivamente empregável para execução de tarefas ou geração de novo conhecimento, assim como também de autoria e confiabilidade reconhecíveis.

Para compreender o segundo problema é necessário abordar o conflito conhecido como “*long tail*” X “*short tail*” (cauda longa X cauda curta). Identificado por Anderson (2004), o fenômeno da cauda longa refere-se à mudança na oferta de produtos por sítios de comércio eletrônico. Com a possibilidade de anunciar digitalmente esses produtos através da Web, as lojas suplantam o problema de espaço físico para divulgação e oferta e aumentam substancialmente não apenas a quantidade de produtos em oferta, mas principalmente a abrangência e a diversidade cultural do público para os quais os produtos são oferecidos, além de possibilitar melhor relação custo/benefício considerando o preço cobrado de acordo com a demanda pelos consumidores. Uma vez igualados os custos de distribuição dos produtos mais vendidos com os dos menos vendidos em meio digital, empresas de comércio eletrônico como a Amazon.com tornam virtualmente ilimitada a oferta de produtos nas mais diversas áreas e verificam que existe demanda para cerca de 99% destes produtos, com vendas que somadas superam as dos produtos mais vendidos, normalmente os únicos a serem ofertados em lojas convencionais. Embora grande parte das vendas se concentre em poucos produtos (a cauda curta), a oferta de uma grande quantidade de produtos com vendas individuais menores (a cauda longa), pode representar um número agregado de vendas tão grande ou maior do que os produtos da cauda curta.

Esse conflito se repete com a oferta de informação na Web e nada mais é do que uma atualização do processo de seleção natural mencionado por Saracevic (1996). Substituindo a grande quantidade de publicações científicas pela quantidade incomensurável de sítios Web publicados e a pequena parcela de publicações efetivamente citadas pela relativa pequena quantidade de sítios que concentram grande parte dos acessos, estes últimos apresentam-se como a cauda curta. Os demais sítios, que apresentam grande parte da quantidade, mas pequena parte dos acessos, compõem a cauda longa. Ignorar a cauda longa pode significar a desconsideração de conteúdo importante para suprir as necessidades informacionais dos usuários e a própria abrangência da Web como meio de comunicação.

Similar à teoria de Anderson (2004), quanto maior a quantidade de conteúdo disponível, maior é a necessidade de tornar disponíveis ferramentas úteis de busca para os usuários. Além das ferramentas convencionais, modelos baseados em recomendações de outros usuários e filtragem colaborativa baseada no perfil de acesso dos usuários permitem um efeito de customização em massa, isto é, mesmo em meio a uma grande quantidade de informação e de usuários, é possível obter resultados de busca diferenciados para cada usuário. Essas são necessidades e possibilidades que ajudam a explicar a renovação do interesse da RI nas atividades de busca de informação na Web.

3.2 Recuperação da Informação na Web

Em um esforço de melhor abordar os aspectos social e cultural da Web, bem como de seu alcance, é possível citar outra definição proposta pelo professor Kevin Hughes, do Colégio Comunitário de Honolulu em 1993, e que ficou conhecida por ser utilizada em diversas publicações que abordaram o advento da Web: “uma iniciativa de recuperação da informação hipermídia num amplo espectro de áreas, que pretende dar acesso a um largo universo de documentos” (HUGHES, 1993 apud ROBREDO, 2005, p. 247). Por hipermídia deve-se compreender o uso de texto, imagens, sons, vídeos e a crescente gama de formatos disponíveis para acesso via Web.

O próprio alcance da Web, quando comparada com outros ambientes “convencionais” para recuperação da informação, merece destaque especial no texto de Robredo (2005, p. 336):

[...] a *Web* permite ter acesso a listas de discussão e grupos especializados, que apresentam informações confiáveis e autorizadas que não se encontram nas bibliotecas.

Enfim, em termos de interdisciplinaridade, a *Web* é uma ferramenta única para recuperar juntas páginas relacionadas com os tópicos pesquisados, produzidas por especialistas de diversas áreas, cuja localização nas bibliotecas, especializadas ou não, seria praticamente impossível com base em catálogos tradicionais, cuja abrangência se limita ao acervo do recinto da entidade.

Porém, o autor ainda destaca que a idéia de acesso livre e universal às informações na Web também é extremamente frágil. Primeiramente por motivos de

exclusão digital mencionados anteriormente, o que leva à inexistência do registro de conhecimento destes excluídos no ambiente. Em segundo lugar por existirem bases de dados especializadas somente disponíveis a um público que possa arcar com os custos de suas assinaturas, além de informações não corretamente estruturadas, de acordo com contextos para aplicações específicas, e que por isso se tornam inacessíveis por processos de recuperação da informação para essas aplicações.

Superados esses obstáculos, a Web (entendida como a realização técnica do ciberespaço) “será em breve o principal equipamento coletivo internacional de memória, pensamento e comunicação” (LÉVY, 1999, p. 167). Uma vez que a Web se torna cada vez mais de uso difundido para registro da informação, transformada em bem de valor para o desenvolvimento de pessoas, de empresas e da sociedade em geral, maior também tem se tornado o desafio de recuperar informação útil. Da mesma forma como privilegia o compartilhamento de informação e conhecimento, a Web traz dificuldades para organizar e recuperar este conteúdo, considerando o real objetivo que é responder a necessidades informacionais de seus usuários.

Portanto, a Web pode ser vista como uma base de dados muito grande, não estruturada mas onipresente. Isso desencadeia a necessidade por ferramentas eficientes para gerenciar, recuperar e filtrar informação dessa base de dados (BAEZA-YATES; RIBEIRO-NETO, 1999, p. 367, tradução nossa).

Dentre as características da Web que influenciam a RI constam a grande escala de abrangência, o constante crescimento, a grande quantidade de conteúdo não textual, o fato de os documentos não serem auto-contidos⁷, a presença de documentos pouco informativos, a inexistência de contexto comum aos autores e o grande acesso de utilizadores considerados leigos nas técnicas de RI (CALADO, 2009).

Baeza-Yates e Ribeiro-Neto (1999) destacam como desafios à RI na Web, considerando a análise técnica dos dados para fins computacionais, a distribuição de dados em diferentes computadores e plataformas, o alto percentual de volatilidade dos dados, o grande volume de dados, a presença de dados não estruturados e redundantes, bem como a qualidade e a heterogeneidade dos dados.

⁷ O autor sustenta que a informação na Web é composta de documentos, estrutura HTML, estrutura hipertextual, características do endereço Web e metadados. Isso faz com que determinado registro informacional só tenha significado após a combinação de vários documentos interligados. Da mesma forma, os próprios *hyperlinks* contêm relações temáticas e de confiança entre autores dos documentos.

Do ponto de vista dos usuários, os autores destacam a dificuldade na especificação de consultas e de interpretação dos resultados apresentados pelo sistema de busca.

Manning, Raghavan e Schütze (2009) destacam não apenas a grande escala e quase completa falta de coordenação na criação e no crescimento da Web, mas também a diversidade de experiências anteriores e motivações de seus participantes. Isso pode levar à geração de conteúdo duplicado de forma legítima ou não, contendo verdades, mentiras, contradições e suposições, o que faz a RI ser diferente e de mais difícil emprego nesse ambiente.

Todos esses fatores são extremamente desafiadores aos paradigmas tradicionais da RI discutidos no capítulo anterior. Por exemplo, a publicação de conteúdo de forma descentralizada e também sem controle central de autoria, responsáveis pelo crescimento exponencial da Web, tornam inviável a indexação e mesmo a recuperação de todo esse conteúdo (MANNING; RAGHAVAN; SCHÜTZE, 2009). Resulta que a pesquisa na área não apresentou tendências essencialmente novas que pudessem vir a atender plenamente as necessidades de usuários da sociedade interligada na Web em relação à RI.

Inicialmente, métodos tradicionalmente utilizados em outros meios de comunicação foram empregados para tentar facilitar a tarefa do usuário de encontrar a informação desejada. Com isso, notou-se a existência de publicações com o intuito de fornecer guias de navegação para sítios Web. Um tipo de páginas amarelas, onde os usuários podiam encontrar classificados sob distintas categorias os sítios então existentes. Por exemplo, para encontrar todos os sítios de busca na Web ou todas as páginas de notícia, bastava recorrer à sessão correspondente, verificar quais os endereços aplicáveis e digitá-los em seu navegador. Hoje pitorescos, estes guias apregoavam descrever “toda a Web”, inocentemente desconhecendo ou ignorando o crescimento exponencial da Web e, portanto, a sua imediata desatualização.

Atualmente, é possível dizer que a Web é o melhor guia para si própria (LÉVY, 1999). Para tanto, os usuários dispõem de sítios que proporcionam ferramentas de busca, utilizando algoritmos específicos para responder às palavras-chave inseridas com um conjunto de resultados ou através da navegação entre páginas de um mesmo sítio ou de sítios diferentes, com as conexões hipertextuais.

O emprego de diretórios de navegação em tese normalmente facilitaria a percepção de relevância para seus usuários. Isso porque neles os documentos Web

passam por um processo prévio de classificação, agrupando-os em categorias navegáveis em taxonomias hierárquicas, que podem ser inclusive especializadas em uma determinada área do conhecimento. O problema reside no fato que a classificação demanda interação humana ou técnicas automatizadas para ser feita, contrastando com o grande volume de informação na Web. Da mesma forma como existem problemas óbvios com a limitação do número de pessoas para executar tal tarefa, não existem métodos automáticos considerados plenamente satisfatórios. Além disso, o próprio processo de classificação vem a refletir a interpretação das pessoas que o realizam ou das pessoas que desenvolveram o método automatizado, e não necessariamente a dos usuários que têm necessidade de encontrar o conteúdo classificado.

Exemplos de recuperação da informação através da navegação se apresentam normalmente sob dois modelos. A navegação direta consiste na seqüência de visitação de páginas de acordo com uma ordem de classificação segundo a qual o usuário acredita que vá encontrar a informação desejada, que é o seu objetivo definido inicialmente. Obviamente, a ordem de classificação imaginada pelo usuário pode ser muito distinta da efetivamente utilizada pelo administrador do conteúdo no sítio, relação estudada pela Arquitetura da Informação para Web mencionada no item anterior e abordada com maiores detalhes a seguir. Além disso, o próprio objetivo do usuário pode sofrer modificações na medida em que acessa novo conteúdo, uma vez que ele pode obter respostas parciais ou complementares que interferem ou influenciam a sua necessidade inicial.

O segundo modelo é um tipo de navegação indireta. Neste caso, o usuário não define inicialmente o seu objetivo de forma completa ou o tem modificado ou esquecido, em virtude de outros interesses surgidos no intervalo de tempo em que ocorre o processo de navegação. Um exemplo da indefinição inicial pode ocorrer quando se está interessado em notícias sobre economia em geral e, ao acessar um sítio de um portal de informações econômicas, notícias de diferentes subáreas se encontram disponíveis, levando a navegação por conteúdos não vislumbrados inicialmente como necessidade. Já o caso de mudança de interesse pode ocorrer quando, ao visualizar uma propaganda de um produto qualquer na página de notícias econômicas, conhecer mais sobre esse conteúdo se torna a necessidade momentânea do usuário, levando-o a acessar a nova página, respondendo a um objetivo totalmente distinto.

Ao analisar os motores de busca, assim chamadas as estruturas de *hardware* e *software* que compõem os serviços de busca na Web, Baeza-Yates e Ribeiro-Neto (1999) destacam algumas diferenças entre sistemas de RI tradicionais e os direcionados para a Web. A principal delas reside no fato de que a resposta às consultas não pode ser feita acessando diretamente toda a informação disponível, visto que, dado o grande volume de dados, tal tarefa seria inviável. Desta forma, ainda mais importante se torna a escolha e o desenvolvimento de algoritmos de indexação e de busca.

Ferramentas de indexação amplamente utilizadas são os *web crawlers*. Do verbo rastejar, em inglês, *Crawlers* são agentes de *software* que fazem varredura na Web, informando a um servidor central sobre novas páginas ou atualizações em páginas existentes. Eles são executados localmente em servidores Web de serviços de indexação enviando requisições de dados para outros servidores, lidando com dados desestruturados e por vezes incompletos para indexação. Negativamente, isso representa aumento no tráfego de dados na Internet e também sobrecarga no processamento dos servidores acessados. Além disso, considerando a questão da volatilidade dos dados, em que eles podem ser alterados ou excluídos constantemente, mesmo a execução periódica dos *crawlers* não garante a confiabilidade dos dados indexados, correndo o risco de majorar os problemas de tráfego e acesso acima citados. Como solução, alguns estudos são desenvolvidos no sentido de padronizar as regras e formatos para intercâmbio de informações entre páginas alocadas em servidores e agentes utilizados para indexação, abordados ainda neste tópico.

As interfaces de usuário merecem destaque ao abordar a RI em serviços de busca na Web, prioritariamente em dois momentos, na interface de consulta e na apresentação de resultados, objetivando melhorar a experiência de uso (usabilidade) e a compreensão das ferramentas pelos usuários. Estes normalmente não sabem a forma como serviços de busca processam os termos por eles digitados. Por exemplo, serviços de busca podem não considerar palavras curtas ou comuns como preposições (as chamadas *stopwords*), podem fazer a busca pela raiz ortográfica da palavra digitada (*stemming*), ou diferenciar entre maiúsculas e minúsculas (BAEZA-YATES; RIBEIRO-NETO, 1999). Além disso, podem prever a construção de consultas complexas ou permitir o uso de linguagens de comando para criar uma expressão de busca, utilizando conectores lógicos como “E” (*AND*) e “OU” (*OR*).

Com isso, é comum o problema de os termos de busca utilizados pelos usuários já não apresentarem a sua real necessidade de informação.

A apresentação de resultados deve prever também o menor esforço possível para o usuário, facilitando a sua tarefa de RI. Por isso, deve ser possível a reorganização dos resultados apresentados segundo critérios como número de resultados por página, data, relevância (segundo algum conceito do próprio serviço) e até mesmo alfabeticamente. Os próprios resultados devem conter as informações minimamente necessárias para caracterizar os registros, como título, descrição indexada, o endereço, o tamanho do documento e a data de indexação (BAEZA-YATES; RIBEIRO-NETO, 1999). Outras opções possíveis seriam a de refinamento ou expansão dos resultados. O refinamento é a aplicação de um novo filtro ou critério de seleção sobre o conjunto previamente obtido. A expansão é a busca por termos ou itens de alguma forma relacionados ao conjunto ou resultado inicialmente apresentado.

A reorganização dos resultados é fundamental para a apresentação, pois a forma de criação de uma espécie de *ranking* de resultados pode definir, dentro do conjunto, quais efetivamente serão visualizados pelos usuários, de acordo com a sua disponibilidade de esforço e tempo. Algoritmos de *ranking* se tornaram um trunfo para os serviços de busca e podem ser criados utilizando lógica booleana, modelos de vetores e mais recentemente, baseados em teorias de análises de *links* (abordadas a seguir), pois dado o grande volume de informação na Web, o número de respostas para uma pesquisa simples se torna muito grande, o que inclusive torna difícil a mensuração de revocação (BAEZA-YATES; RIBEIRO-NETO, 1999). Os autores apontam ainda a necessidade de desenvolvimento de melhores algoritmos de *ranking* que englobem tanto conteúdo quanto estrutura da informação.

Existem ainda ferramentas que integram vários serviços de busca em uma única interface de entrada de dados para consulta e de apresentação de resultados. Os assim chamado *metasearchers* (ou metabuscadores) são desenvolvidos com o intuito de facilitar a tarefa de interação de usuários com vários serviços, sendo adaptados especificamente para trocar informação com estes, desempenhando o papel de agentes de *software*.

Entre os serviços de busca na Web claramente se vê o destaque do sítio Google como o de maior uso nos Estados Unidos da América (THE NIELSEN COMPANY, 2010) e provavelmente em todo o mundo, excetuando-se em países

onde ele não atua ou sofre restrições de acesso. O serviço foi considerado revolucionário quando de seu lançamento ao utilizar um algoritmo denominado PageRank, que funciona basicamente calculando o número de *links* que apontam para cada registro indexado pela busca e assim define o grau de importância deste registro na apresentação dos resultados. Baseando o seu julgamento de relevância no número de vezes em que um registro é indicado por outro documento na Web, por inúmeras vezes não resolve o problema de RI de um usuário, pois o aspecto que lhe é relevante pode ser diferente. Além disso, em pesquisas com pequeno número de parâmetros, ou com a especificação destes sem maiores cuidados pelo usuário, o que caracteriza a maioria das situações, o Google se caracteriza por normalmente apresentar um grande número de resultados, inviabilizando o exame minucioso do conjunto retornado.

Mesmo estando longe de ser considerado um serviço infalível, por ser o de maior uso no mercado o Google se tornou referência para o desenvolvimento de outros motores de busca. Entretanto, apesar de algumas iniciativas inovadoras, as ferramentas comerciais de maior alcance tendem a utilizar o mesmo caminho que ele, interessados em sobrepujá-lo e não necessariamente em efetuar melhor RI (CALADO, 2009).

O diferencial do algoritmo implementado pelo Google remete a um exemplo de desenvolvimento de inovações em teorias de RI para a Web, a análise de *links* (também conhecida como *Link Analysis*). Baseada em antecedentes de análises de citações, Manning, Raghavan e Schütze (2009) sustentam que o estudo dos relacionamentos entre documentos Web contidos em *links* pode ser empregado não apenas para organizar os resultados de busca, mas também para explicitar referência entre um e outro documento. Da mesma forma como é possível vir a consultar as obras citadas em um determinado artigo, também é possível ter interesse em acessar documentos Web referenciados por uma página visitada.

Entretanto, novamente as próprias características do conteúdo disponível na Web criam empecilhos mesmo para inovações como a análise de *links*. O problema no caso reside no fato de que a conexão entre documentos na Web via *links* não é necessariamente perene (os *links* podem ser modificados ou excluídos constantemente) nem balanceado, o que personifica o dilema cauda longa x cauda curta citado no item anterior. O fato de existirem algumas páginas com grande volume de *links* para elas direcionados influencia a análise dos *links* de forma a crer

que essas páginas são de fato mais relevantes do que as outras, o que não necessariamente é verdadeiro. Manning, Raghavan e Schütze (2009) ilustram a estrutura e a concentração de *links* entre documentos na Web em um formato de laço, composto basicamente por três estruturas, onde a maioria dos documentos Web se enquadra: IN, SCC e OUT (em português, respectivamente, entrada, componentes fortemente conectados – *strongly connected components* – e saída). As páginas classificadas como IN, apontam para páginas do tipo SCC. Páginas SCC possuem *links* para outras páginas SCC e para páginas OUT. As exceções são apresentadas em formato de tubo (*tubes*), páginas IN que apontam diretamente para páginas OUT, ou cachos (*tendrils*).

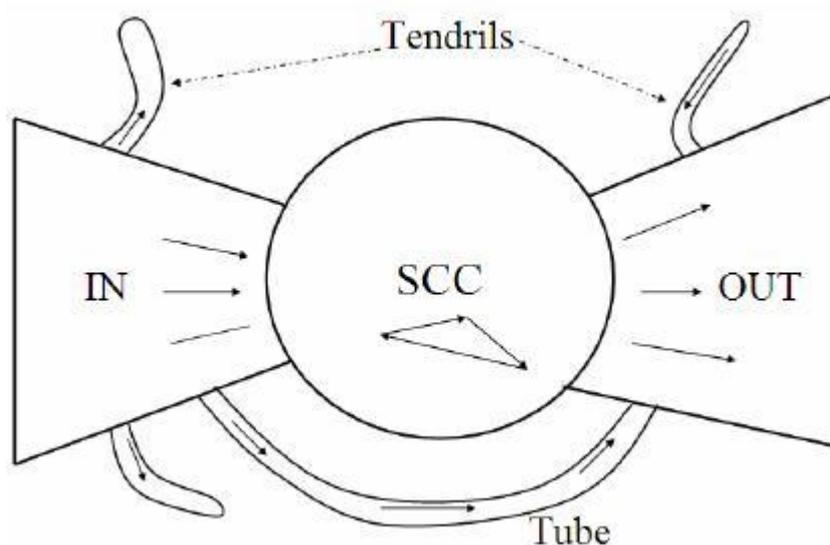


Figura 1 - A estrutura em laço de documentos na Web (MANNING; RAGHAVAN; SCHÜTZE, 2009, p. 427)

A própria forma de estruturação e apresentação da informação empregada pelos sítios Web representa um desafio à RI pelos usuários. Robredo (2005, p. 316) indica os dois enfoques normalmente dados a este problema, o primeiro vinculado à visão mais tradicional da AI e da CI, e o segundo voltado para a Arquitetura da Informação para a Web:

Assim, quando se fala de arquitetura da informação em ambientes *Web* cabe distinguir dois focos de atenção: um, a estruturação dos documentos a serem publicados na Internet e de seu conteúdo, com vistas à exploração de suas informações de um ponto de vista temático, e o outro, a facilidade de uso e eficiência na satisfação das necessidades e exigências do usuário, ou seja, a usabilidade do *site*.

A abordagem da AI tradicional é ressaltada por Calado (2009) ao destacar a necessidade de inserção de conhecimento (significados, eventos, entidades, opiniões, comunidades) nos modelo de estruturação dos documentos Web, seja por via da extração da informação ou pela integração com serviços de colaboração de usuários na Web, de forma a proverem fácil acesso e recuperação. Problemas na captura de semântica dos documentos em um conjunto de palavras são um risco perene. Pode ocorrer, por exemplo, a polissemia (mais de um significado para um mesmo termo) e dificuldade no tratamento de palavras sinônimas (BAEZA-YATES; RIBEIRO-NETO, 1999).

Outros métodos mais recentes visam a classificação da informação e do relacionamento entre seus registros na Web. *Tags* no padrão HTML permitem definir tanto o conteúdo, quanto a estrutura e até mesmo o estilo de apresentação do texto. Embora à primeira vista isso seja uma característica conveniente, essa múltipla utilidade impede que elas sejam utilizadas para classificar, interpretar e recuperar informação de forma automática por programas de computador. Além disso, a complexidade crescente com o maior uso das páginas Web demandava a criação de cada vez mais *tags*, ampliando o problema.

A linguagem de marcação XML permite a criação de *tags* personalizadas para melhor caracterizar a informação. Em XML, as *tags* são criadas exclusivamente para a representação da estrutura e do conteúdo da informação. A representação do estilo da apresentação da informação ficou a cargo de outro padrão, o CSS (*Cascading Style Sheets*), também regulamentado pelo W3C e que é utilizado juntamente com o XML e com versões mais recentes do HTML. Um exemplo de registro de informação em XML pode ser da seguinte forma:

```
<dissertacao>
  <instituicao>Universidade de Brasília</instituicao>
  <autor>Fábio Augusto Guimarães Teixeira</autor>
  <titulo>A Recuperação da Informação e a colaboração de usuários na Web</titulo>
  <orientador>Cláudio Gottschalg Duque</orientador>
</dissertacao>
```

Quadro 4 - Exemplo de registro de informação em formato XML

Neste exemplo, a *tag* <dissertacao> representa um grupo de dados que contém outras quatro *tags* (<instituicao>, <autor>, <titulo> e <orientador>), criando

uma relação de pertinência que pode ser compreensível a um programa de computador que busque especificamente a informação de instituição, autor, título ou orientador de dissertações. Isso é a representação da estrutura da informação. O conteúdo em si é o que se encontra descrito entre a abertura e o fechamento de cada *tag*, como a instituição, o autor, o título e orientador. A descrição das *tags*, mesmo sendo o resultado de uma tarefa de personalização, deve ser tornada disponível para que outras pessoas, além do próprio criador, e dispositivos tenham acesso de forma a permitir o intercâmbio de informação.

A possibilidade de compreensão da informação registrada na Web tanto por humanos quanto por máquinas (não apenas computadores, mas diferentes tipos de dispositivos, como celulares, veículos, eletrodomésticos, etc.) é um dos ideais de Tim Berners-Lee para a Web desde a sua criação. O funcionamento dos agentes automatizados encarregados de compartilhar informação pelo ambiente, descritos no capítulo anterior, só é possível com o registro de informações semânticas em formatos por eles compreensíveis. Exemplos de registro de informações semânticas são os metadados: “[...] informação sobre informação. [...] Metadados consistem em um conjunto de propriedades de um documento” (BERNERS-LEE, 2000, p. 181, tradução nossa). Em páginas HTML, os metadados são normalmente registrados através de *tags* específicas em uma área denominada de cabeçalho.

O aumento da importância e da abrangência do uso desses metadados aponta em direção à Web semântica, nome dado à evolução da Web que atende os princípios de acesso e compreensão universal de informação para homens e máquinas. Por isso, padrões como o RDF (*Resource Description Framework* – Estrutura de Descrição de Recurso) e o OWL (*Web Ontology Language* – Linguagem de Ontologia Web), que permitem o registro e compartilhamento de metadados, foram concebidos e se tornaram recomendações do W3C⁸.

A estrutura RDF permite descrever informação sobre recursos Web e sobre os sistemas que os utilizam através de uma relação em que existe um sujeito, um objeto e um predicado (ou propriedade) através de XML (WORLD WIDE WEB CONSORTIUM, 2004e). Já o padrão OWL permite a descrição de ontologias,

⁸ Para saber mais sobre as recomendações RDF e OWL é possível visitar no sítio do W3C, respectivamente, as páginas “<http://www.w3.org/RDF/>” e “<http://www.w3.org/TR/owl-features/>”, que contêm uma visão geral e *links* para documentos relacionados. Observar que o padrão OWL possui a versão 2 disponível.

“vocabulários formalizados de termos, normalmente cobrindo um domínio específico e compartilhados por uma comunidade de usuários” (WORLD WIDE WEB CONSORTIUM, 2009). São especificadas classes, propriedades, indivíduos e valores de dados através dos relacionamentos entre si em um mesma ontologia, utilizando a própria estrutura RDF.

De forma resumida, a Web semântica se caracteriza como uma iniciativa de descrever os recursos de informação disponíveis na Web e os relacionamentos destes entre si, de acordo com diferentes contextos (representados por ontologias). Além de ser uma iniciativa para organização da informação, a Web semântica também tem como objetivo tornar disponível este conteúdo organizado para tarefas de recuperação da informação.

Apesar dessas novas iniciativas, o problema de RI na Web não apenas permanece, mas se fortalece com o passar dos dias. Em 1999, Baeza-Yates e Ribeiro-Neto propuseram que, em face de tantos desafios e da necessidade de desenvolvimento de modelos especiais de RI para a Web, seria mais fácil ensinar os usuários a utilizar corretamente os serviços de busca do que efetivamente tentar adivinhar as necessidades de usuários. Embora incômoda, se verifica essa afirmação na maioria dos casos até hoje.

Inovações para a melhoria da RI podem residir em estudos do emprego da própria colaboração entre usuários da Web como meio de facilitar a organização e a recuperação da informação. Este é o caminho escolhido para o desenvolvimento deste estudo.

4 Colaboração de usuários na Web

No capítulo anterior, foi descrito que quando da idealização da Web, Berners-Lee imaginou que os navegadores deveriam ser dotados de capacidade não apenas de visualização, mas também de edição. Segundo sua visão, as páginas Web poderiam ser livremente editadas por seus visitantes, tornando ainda mais descentralizado o processo de publicação e autoria. Entretanto, as empresas de *software* que lançaram versões comerciais dos navegadores abandonaram essa funcionalidade, que implicava em maiores gastos de desenvolvimento e mesmo de adaptação dos servidores Web.

A funcionalidade de edição almejada por Berners-Lee para os navegadores, além do projeto da Web semântica citada anteriormente para compartilhamento de informação entre agentes de *software*, são características básicas do ambiente colaborativo para o qual a Web foi desenhada:

Quando eu propus a Web em 1989, a força motriz que eu tinha em mente era comunicação através de conhecimento compartilhado e o “mercado” direcionado para isso era colaboração entre pessoas no trabalho e em casa. Ao construir uma Web de hipertexto, um grupo de pessoas de qualquer tamanho poderia facilmente se expressar, rapidamente adquirir e transmitir conhecimento, superar mal-entendidos e reduzir duplicações de esforço. Isso daria às pessoas em um grupo uma nova força para construir algo juntas (BERNERS-LEE, 2000, p. 162, tradução nossa).

Como meio de comunicação colaborativo, pessoas e dispositivos poderiam construir modelos de compreensão que, da mesma forma como capturam, mantêm ou evoluem o conhecimento de indivíduos que deixam o grupo e o tornam disponível para novos incluídos, imediatamente capazes de colaborar. O conhecimento não apenas se torna disponível, mas também constantemente evoluído, com informação em vários níveis e escalas, de rascunhos a versões finais, de local a global, de forma transparente e independente de localização. Além de um plano tecnológico, a Web apresentava uma base social e uma filosofia básica: ser uma representação do mundo (BERNERS-LEE, 2000).

A denominação Web 2.0 surgiu em uma conferência entre as empresas O'Reilly e MediaLive International, na qual se fazia um balanço após a crise que

abalou diversas empresas denominadas “.com”⁹. Concluíram que não só a Web permanecia muito importante, como também novas aplicações surgiam com algumas características em comum (O’REILLY, 2005). Segundo o autor, essas características seriam:

- Considerar toda a Web como uma plataforma (assim como existem plataformas baseadas em sistemas operacionais como o Windows);
- Atrelar-se à inteligência coletiva;
- Considerar a importância dos dados e de seu gerenciamento;
- Definir o fim dos ciclos de lançamentos de *software* (que passariam a ser desenvolvidos e lançados rotineiramente);
- Utilizar modelos de programação leves (de fácil implementação por outros envolvidos ou interessados);
- Desenvolver *software* que ultrapassa o nível de um dispositivo único (no caso, o computador convencional, atingindo dispositivos, móveis ou não, que passam a ter acesso às redes de comunicação de informação); e
- Proporcionar ricas experiências de usuários (que permitam que aplicações na Web, com custos de utilização eventualmente menores, se assemelhem a aplicações executadas localmente nos dispositivos e com melhores interfaces).

Essas seriam ainda aplicações que podem ser liberadas para reuso e alterações, *software* que se torna melhor quanto mais pessoas o utilizam (a chamada arquitetura de participação), onde o comportamento dos usuários não é predeterminado e existe confiança depositada na interação com e entre eles.

Na Web colaborativa são os usuários que adicionam valor ao conteúdo, mesmo quando isso for apenas um efeito colateral do uso dessas aplicações. Isso pode ocorrer apenas por um motivo de interesse pessoal, e não o de voluntariamente compartilhar colaborações. Barbosa, Sepúlveda e Costa (2009, p. 21-22) opinam que “o conjunto de tecnologias conhecidas como Web 2.0 está disponível para que as pessoas possam obter acesso à informação e se

⁹ O chamado “estouro da bolha”, no qual inúmeras empresas de tecnologia para serviços Web que receberam vultosos investimentos ficaram sem acesso a novos investimentos, em vista da falta de retornos, de modelos de negócio não sustentáveis e da concepção de que os negócios no ambiente estavam superdimensionados, e por isso fecharam as portas.

comunicarem com outras pessoas”. Definição mais abrangente é apresentada por Tapscott e Williams (2006, p. 19, tradução nossa):

Existem vários nomes para essa nova Web: a Web 2.0, a Web viva, a Hypernet, a Web ativa, a Web de leitura/escrita. Chame do que você quiser – o sentimento é o mesmo. Nós estamos todos participando no crescimento de uma plataforma global, onipresente para computação e colaboração que está redesenhando quase todos os aspectos das questões humanas. Enquanto a velha Web era sobre sítios Web, cliques e visualizações rápidas, a nova Web é sobre comunidades, participação e compartilhamento. Com a multiplicação de usuários e poder computacional e a proliferação de ferramentas de fácil uso, a Internet está evoluindo para um computador global, vivo e interligado que todos podem programar. Mesmo o simples ato de participar em uma comunidade *on-line* é uma contribuição para os novos bens comuns digitais – mesmo que alguém esteja construindo um negócio na Amazon ou produzindo um vídeo para o YouTube, criando uma comunidade sobre a sua coleção de fotos no Flickr ou editando o verbete de astronomia na Wikipedia.

Antigamente pesquisadores se mostravam surpresos com a quantidade de informação utilizada pelas pessoas, o público em geral ou pesquisadores, e que na verdade eram provenientes de outras pessoas e não de fontes especializadas, demonstrando a importância de consultar as primeiras como fontes de informação (BATES, 2002). Hoje, o desenvolvimento de ambientes que permitam a interação humana se tornou uma necessidade, e o grau de facilidade com que isso ocorre define o valor do ambiente como forma de transmissão de conhecimento.

A colaboração na Web permite registrar diretamente as conexões percebidas por um usuário, ao qual é dado o mesmo valor de participação que outros, quando em contato com determinada informação. Isso evita não apenas perder essas conexões, mas estimula a construção de conhecimento, sem limitações de idéias ou soluções. Supondo a motivação benéfica da colaboração, a compreensão compartilhada permite o engrandecimento das discussões, partindo da premissa de que delas não resultam verdades absolutas, mas em constante modificação.

A Web colaborativa é composta por aplicações e arquiteturas voltadas para a participação ativa, em contraponto à passividade dos usuários. Os sítios Web passam a se tornar comunidades de relacionamento *on-line*, agrupando usuários de acordo com afinidades, mesmo temporárias ou superficiais. Para isso, os sítios podem ser compostos apenas por uma estrutura básica, onde o mais importante é o desenho para conter informação. Os usuários então a inserem e a relacionam, agregando valor que pode ser utilizado para denotar ou inferir conhecimento, através

de análise da própria ferramenta ou mesmo da combinação de várias outras ferramentas.

A interface e as políticas de uso desses serviços devem prover a rica experiência de usuários de forma a facilitar o uso e a reafirmar a vocação da Web para aproximar as pessoas, fomentando o relacionamento entre elas, suprimindo a necessidade por espaços físicos de convivência, em número cada vez mais reduzido no mundo não virtual.

A inteligência coletiva, conforme descrita anteriormente, se encontra registrada no conhecimento agregado das escolhas descentralizadas, no julgamento de grupos de participantes independentes, nos serviços Web de coleta de participações de usuários e por último nas novas formas de estruturação e classificação de conteúdo, baseadas em *tagging* e que resultam em folksonomias, esquemas de classificação (taxonomias) orgânicos que organizam conteúdo na Web (TAPSCOTT; WILLIAMS, 2006).

O advento de tecnologias de colaboração na Web serve de suporte para mudanças estruturais não apenas na forma de comunicação em sociedade, mas também na economia. O uso da colaboração em empresas, quando dirigido por políticas específicas, atua como agente de inovação e melhoria de desempenho (BARBOSA; SEPÚLVEDA; DA COSTA, 2009). Tapscott e Williams (2006) denominam esta nova economia, cujos princípios são a abertura e a transparência para novas idéias, o compartilhamento de pessoas e empresas para inovação e crescimento, e a tomada de decisões em escala global, de *Wikinomics*. Trata-se de um sistema econômico baseado em avanços da tecnologia e no engajamento dos usuários, além de representativo em escala global, cujos benefícios para as empresas podem ser obtidos a partir de passos como:

- Aceitar sugestões dos principais usuários;
- Construir massa crítica;
- Prover uma estrutura para colaboração;
- Não apressar a obtenção de estruturas e governança corretas;
- Certificar-se de que todos os participantes podem obter algum valor;
- Manter normas comunitárias;
- Permitir a evolução do processo; e
- Estimular a mente colaborativa.

Baseando-se no baixo custo para o emprego da colaboração de pessoas, no cada vez maior número de pessoas envolvidas na produção de conhecimento e de produtos com características únicas e inovações importantes descobertas pelo acaso, os autores ainda definem alguns modelos de colaboração em massa que desafiam os padrões de negócios tradicionais:

- **Compartilhamento:** o grande número de voluntários envolvidos no desenvolvimento de projetos, como *software* de código aberto e a Wikipedia, que se contrapõem àqueles criados por empresas tradicionalmente fechadas;
- **“Ideagoras”:** a abertura e a transparência de empresas para aceitar idéias e inovações provenientes de pessoas externas à sua força de trabalho;
- **“Prosumers”:** a participação ativa dos usuários como consumidores, avaliando, criticando, exigindo ou mesmo realizando melhorias em produtos a partir do acesso à informação;
- **Novas Alexandrias:** a construção de novos repositórios de conhecimento científico, cultural e tecnológico baseada no compartilhamento e na colaboração, de forma mais rápida e com menos custo;
- **Plataformas para participação:** tornando disponíveis ferramentas para que comunidades de parceiros possam agregar valor informacional a projetos e produtos;
- **Chão de fábrica global:** a globalização sob o ponto de vista da indústria, onde produtos podem ser fabricados em diferentes lugares do planeta, ou compostos por componentes fabricados dessa forma; e
- **O espaço de trabalho Wiki:** a colaboração em equipes de trabalho considerando aspectos de meritocracia em substituição a cadeias hierárquicas.

A falta de hierarquia, característica da Web, se torna ainda mais aprofundada na Web colaborativa. Em um ambiente de permanente compartilhamento, é inevitável a recombinação de informação. A partir de uma fonte única, a informação pode ser retransmitida de diferentes formas sem qualquer tipo de controle, pelos próprios usuários ou mesmo por serviços que combinam contextos distintos. Isso

não apenas influencia a quantidade de informação redundante gerada, mas também pode afetar a sua confiabilidade e a sua qualidade.

A preocupação se estende ao mencionar que os relacionamentos existentes de usuários entre si e com determinados objetivos ou projetos na Web colaborativa podem ser tanto passageiros quanto superficiais. A informação, neste caso, pode se tornar de baixa qualidade e/ou não corretamente aplicada ao contexto que se supõe, gerando uma forma de inteligência coletiva medíocre. Novos usuários dessa informação, se não capazes de julgar a sua qualidade e corrigi-la, tendem apenas a piorar a situação.

O nível de participação de usuários em ferramentas colaborativas é uma questão a ser discutida. Nielsen (2006) sustenta que a grande maioria dos usuários são apenas observadores, isto é, apenas lêem o conteúdo sem realizar novas contribuições; uma pequena parte é composta por usuários que contribuem esporadicamente; e existe um número ínfimo de usuários que efetivamente realizam a grande maioria das contribuições. Por exemplo, na Wikipedia, esta relação é de, respectivamente, 99,8%, 0,2% e 0,003%, do quantitativo de usuários. Segundo o autor, o problema é permanente. Por isso, propõe algumas formas de estimular a participação dos usuários a fim de minimizá-lo: tornar mais fácil a colaboração, com esquemas de classificação do tipo *ranking* ao invés de descrições completas em linguagem natural; utilizar a participação implícita, onde o usuário não precisa efetivamente escrever ou classificar algo, mas tem o seu comportamento (de compra, de visualização, etc.) utilizado como forma de prover recomendação para os demais; tornar disponíveis modelos iniciais contendo alguma informação, permitindo aos usuários editar, ao invés de criar, conteúdo; criar algum tipo de recompensa, como reputação, para os usuários; e promover usuários cujas contribuições tenham qualidade reconhecida pelo sistema ou por outros usuários (NIELSEN, 2006).

A participação em projetos colaborativos também pode afetar a confidencialidade das informações que os usuários detêm. Como não há controle sobre a publicação, fica a critério de cada um definir o que pode ou não informar. Por mais abertas e transparentes que possam ser as organizações com as quais possuem vínculos, possivelmente existem determinadas informações de cunho estratégico e restrito. Ao mesmo tempo, como participantes ativos de uma comunidade ou de um projeto, os usuários permitem que os demais tenham acesso

a suas opiniões e ao seu conhecimento, comprometendo a privacidade de que dispõem enquanto são apenas usuários passivos.

Existem ainda riscos que a Web colaborativa traz à autoria, como tradicionalmente conhecida. A coletivização na construção de conceitos como processo inacabado e em constante evolução pode pressupor que seus autores concedam suas propriedades intelectuais em nome de um bem maior, tornado disponível para toda a comunidade. Por outro lado, a não existência de informações sobre autoria também pode dificultar o julgamento de confiabilidade, pertinência e até mesmo de relevância da informação para determinado usuário ou grupo de usuários.

Uma forma de observar os benefícios, bem como debater sobre os riscos, da colaboração de usuários na Web é através da descrição das principais tecnologias existentes para este fim, o que foi realizado a seguir.

4.1 *Tecnologias de colaboração entre usuários na Web*

Para McAfee (2006) existem duas categorias sob as quais se enquadram as tecnologias de informação utilizadas para comunicação: canais de comunicação e plataformas de comunicação. Entre os canais estão incluídas as mensagens eletrônicas (*e-mail*) e as mensagens instantâneas (cujas aplicações famosas são, por exemplo, o Microsoft Messenger e o Google Talk), permitindo livre distribuição da informação para os usuários diretamente envolvidos. Já as plataformas englobam Intranets, páginas Web corporativas e portais de informação. As plataformas se distinguem dos canais por terem o controle de publicação de informação centralizado em um grupo mais restrito de pessoas. Porém, uma vez liberada a circulação da informação, ela se torna disponível a todos os seus usuários, atingindo maior escala, o que não acontece com os canais.

Ambas as categorias apresentam benefícios e restrições de uso. Os canais não podem ser acessados e/ou buscados por qualquer usuário, da mesma forma como as visitas realizadas por pessoas às plataformas não deixam rastros visíveis para os demais, isso sem contar que o próprio controle de publicação restringe a circulação da informação e o crescimento da rede de comunicação. Considerando

benefícios das duas categorias, novas plataformas para “geração, compartilhamento e refinamento de informação” (MCAFEE, 2006, p. 23, tradução nossa) foram desenvolvidas e se tornaram populares na Web, caracterizando as tecnologias da Web colaborativa.

O autor ainda descreve as tecnologias que compõem o chamado paradigma Enterprise 2.0, termo criado para designar as ferramentas de colaboração entre usuários que podem ser utilizadas por empresas para tornar disponíveis as práticas e os produtos de conhecimento de seus funcionários, através do acrônimo SLATES: *Search* (busca), *Links*, *Authoring* (autoria), *Tags*, *Extensions* (extensões) e *Signals* (sinalizadores).

A importância das tecnologias de busca reflete para a colaboração de usuários o mesmo desafio imposto à RI em geral, referindo-se não só a cada vez mais crescente dificuldade de classificar e ordenar toda a informação tornada disponível na Web, mas ao próprio fato de que se torna cada vez mais difícil encontrar a informação desejada. Na Web colaborativa, isso ocorre em ritmo ainda mais acelerado, e com destaque para os serviços de busca:

Para que qualquer informação tenha valor, seus usuários devem ser capazes de encontrar o que estão procurando. Leiautes de páginas Intranet e ajudas de navegação podem ajudar, mas usuários estão crescentemente ignorando estes em favor de buscas por palavras-chave (MCAFEE, 2006, p. 23, tradução nossa).

Uma das ferramentas utilizadas na Web para enfrentar este problema é a teoria de análise de *links* (*Link Analysis*). Apresentada no capítulo anterior, ela possui uma relação estreita com a colaboração de usuários, registrada por Kleinberg (1999, p. 604-605, tradução nossa):

A estrutura em rede de um ambiente com *hyperlinks* pode ser uma rica fonte de informação sobre o conteúdo do ambiente, desde que tenhamos meios efetivos para entendê-lo. [...] Em particular, nós focamos o uso de *links* para analisar a coleção de páginas relevantes para um tópico de ampla pesquisa e para descobrir as páginas com mais “dominância” nesses tópicos. [...] A WWW é um *corpus*¹⁰ hipertextual de enorme complexidade e continua a se expandir em uma taxa fenomenal. Além disso, ela pode ser vista como uma forma intrincada de hipermídia populista, na qual milhões de participantes *on-line*, com objetivos diversos e freqüentemente conflitantes, estão continuamente criando conteúdo com *hyperlinks*. Dessa forma, enquanto indivíduos podem impor ordem em um nível extremamente local, a sua organização global é completamente não planejada – estrutura de alto nível pode emergir somente através de uma análise posterior.

¹⁰ Conjunto de elementos utilizado como base para algum tipo de análise.

Um exemplo da análise de *links* é a empregada pelo Google em seu já mencionado algoritmo PageRank. Ao mencionar um assunto e criar um *link* para uma outra página em seu próprio conteúdo, o usuário está colaborando para que o algoritmo de busca informe a outros usuários que essa relação foi descrita. Em Intranets, a criação de *links* é bem mais restrita do que na Internet, em virtude do controle de publicação normalmente mantido por um grupo pré-estabelecido.

O'Reilly (2005) comparou a criação de sites e *links* na Internet às conexões cerebrais. Quando um usuário insere nova informação, outras pessoas descobrem o conteúdo e constroem ligações para ele. Quanto mais ligações, mais vezes a informação pode ser encontrada e utilizada, crescendo em importância de forma orgânica, ou seja, vinculada ao contexto de cada usuário.

Tecnologias de autoria são as que permitem que pessoas criem e compartilhem conteúdo entre os demais usuários de uma rede de comunicação. Exemplos claros desse tipo de tecnologia são os *weblogs* (ou simplesmente *blogs*) e as *wikis*.

Blogs são ferramentas através das quais uma pessoa posta (cria e publica) conteúdo por conta própria, apresentando opinião ou notícias segundo sua responsabilidade, e onde outras pessoas podem realizar comentários, que podem ser censurados ou não pelo autor da postagem. Trata-se, portanto, de uma iniciativa de autoria individual, diferentemente da autoria de grupo permitida pelas *wikis* (MCAFEE, 2006). O conjunto de *blogs* e o grande número de suas respectivas postagens e comentários recebidos dá nome à chamada "blogosfera", ambiente que reúne grupos de discussão em conversas dos mais diferentes níveis. Os *blogs* desempenham ainda um papel fundamental na disseminação de informação pois são meios ágeis de publicação e que podem também interferir na análise de *links* abordada anteriormente. Enquanto algoritmos de busca realizam verificações quantitativas de *links* criados para determinada página, *blogs* podem conter verificações qualificativas de seu conteúdo (O'REILLY, 2005).

Uma pesquisa realizada pelo Pew Research Center recentemente divulgada afirma que o número de norte-americanos adolescentes e jovens entre 18 e 29 anos que freqüentam *blogs* vem diminuindo desde 2006, enquanto que adultos acima de 30 anos têm passado a usar mais as ferramentas (BBC BRASIL, 2010). Segundo a pesquisa, alguns dos motivos para que os jovens perdessem o interesse nos *blogs* é a falta de rapidez e diversão nas leituras, além do surgimento de outras redes

sociais que permitem mensagens curtas, motivadas pelo crescimento do número de acessos à Web via celular.

As *wikis* mereceriam um trabalho a parte para descrever todas as suas implicações sociais, benefícios e riscos para o compartilhamento de informação e conhecimento. Neste momento, cabe a descrição de seu papel como ferramenta de autoria na qual o usuário se torna responsável pela contribuição e pela sua veracidade, com o objetivo de que ela seja útil para outras pessoas e se mantenha publicada com o passar do tempo, como está ou servindo de base para outros aprofundamentos. Da mesma forma como estão amparadas na confiança dos usuários, as *wikis* se utilizam deles para o controle de possíveis incorreções. Essas ferramentas atuam sobre a retenção do conhecimento, um processo da gestão do conhecimento, por “facilitar a pesquisa futura sobre diversos temas a partir da categorização dos conteúdos, por exemplo, através das perguntas mais freqüentes, manuais, tutoriais, planilhas e artigos” (SCHONS, 2008, p. 85). Segundo o autor, servem como repositório e memória de conhecimento organizacional por permitir a socialização do conhecimento e recuperação futura, inclusive de versões anteriores de páginas alteradas.

A Wikipedia é um exemplo do poder das *wikis*. Desenvolvida a partir de uma plataforma *wiki*, permite que usuários voluntários do mundo inteiro criem, editem, consultem e debatam sobre verbetes nas mais variadas áreas. A enciclopédia, assim como sua gestora (*Wikimedia Foundation*), não tem fins lucrativos, não cobrando ou remunerando os seus usuários pelo uso e participação. Com isso, todo o seu conteúdo é considerado de livre uso, transcrição, modificação e ampliação, desde que preservados os direitos para que outros usuários possam fazer o mesmo, registrando os autores para fins históricos. Existe algum tipo de controle, com usuários que cumprem o papel de editores, escolhidos devido ao maior número de suas contribuições, para moderar discussões a respeito dos verbetes e verificar atualizações possivelmente incorretas. Em maio de 2010, a Wikipedia contava com mais de 3.286.000 verbetes em língua inglesa, mais de 1.065.000 verbetes em alemão, mais de 945.000 em francês, mais de 685.000 em italiano, mais de 674.000 em japonês, mais de 595.000 em espanhol e mais de 570.000 em português, além de diversas outras línguas (WIKIMEDIA FOUNDATION, 2010).

Um artigo publicado na revista *Nature* em 2005 causou polêmica ao afirmar que, após o exame de uma amostra por especialistas nos assuntos, os verbetes da

Wikipedia se encontravam em nível próximo de exatidão aos da Enciclopédia Britânica (GILES, 2005). Não necessariamente pelo número de verbetes da Wikipedia considerados corretos, mas pelo número aproximado de inconsistências encontradas nos verbetes.

Apesar do destaque, Jimmy Wales, um dos criadores da enciclopédia, em entrevista à revista *Veja* declarou que “A Wikipédia é um excelente ponto de partida para uma pesquisa, mas não é o melhor ponto final”, sendo necessário buscar fontes de confirmação ou ampliação do conceito abordado (FAVARO, 2009). Mesmo assim, destaca não só a diversidade de usuários, mas também a diversão que o projeto proporciona a eles, o que estimula a velocidade de correção e de crescimento. Outro destaque da entrevista é a afirmação de que a enciclopédia possui maior conteúdo em tecnologia, ciências exatas e cultura popular, em detrimento de ciências humanas e poesias, por exemplo, o que denota áreas de maior interesse de seus usuários.

As aplicações de autoria estão se tornando cada vez mais difundidas, o que possibilita que usuários comuns produzam informação de maneira mais intuitiva e mais amigável, permitindo que o produtor informacional se atenha à mensagem e não ao meio, tanto para uso individual quanto em grupo. Aplicações normalmente encapsuladas em pacotes de produtividade vendidos para instalação e uso em computadores convencionais, como processadores de texto, planilhas eletrônicas e apresentações eletrônicas, são exemplos de serviços tornados disponíveis diretamente na Web, oferecendo o compartilhamento de documentos e a edição destes em qualquer dispositivo com conexão ao ambiente.

Uma plataforma de autoria que serviu de base para o desenvolvimento de vários outros serviços similares é o YouTube. Surgido como um meio de publicação e distribuição de vídeos postados pelos próprios usuários, o sítio atualmente conta com um elevado número de participantes e participações, modificando a forma como o compartilhamento deste tipo de conteúdo ocorre na Web. É utilizado para divulgação não apenas de vídeos pessoais, mas também filmes de propaganda, corporativos, entre outros. Um problema constante, porém, envolve a publicação de conteúdo não autorizado.

Discussões envolvendo a questão de autoria em ambientes colaborativos são abordadas por iniciativas como as licenças *Creative Commons* (ARAYA; VIDOTTI, 2009). Mediante o conceito de recombinação de informação, característica da Web

colaborativa, as autoras destacam o crescimento e a onipresença de produtores de informação e a criação de licenças flexíveis para atender a este novo público, bem como a minimização do risco na divulgação de conteúdo intelectual. O projeto sem fins lucrativos *Creative Commons* se apresenta como um conjunto de licenças para diversos formatos que explica claramente e de acordo com a legislação vigente quais as possibilidades e restrições de uso, modificação e compartilhamento aplicadas a determinada informação.

Outras ferramentas para controle de autoria em meio ao ambiente de compartilhamento são as licenças GNU. O projeto GNU começou como uma iniciativa para desenvolvimento de *software* livre, gerando posteriormente a fundação *Free Software Foundation*. Como forma de proteger o conteúdo autoral de cada voluntário no projeto e ao mesmo tempo torná-lo disponível para uso de forma transparente e melhoria de outros, foi criada a licença GNU *General Public License* (GNU GPL). Posteriormente, a *Free Software Foundation* lançou a GNU *Free Documentation License* (GNU FDL), nos mesmos moldes da licença para *software*, mas aplicada para documentos e textos.

O emprego dessas licenças gerou novos conceitos de autoria: o *copyleft* e o “*some rights reserved*” (alguns direitos reservados), que contrastam com as idéias de *copyright* e “*all rights reserved*” (todos os direitos reservados). Da mesma forma como contribui em um serviço, o usuário é livre para fazê-lo novamente em outro. Para exemplificar, a figura que serviu de base para o quadro 2 é de propriedade de uma pessoa (ADEVERIA, 2010) que a tornou disponível na Web, permitindo o seu uso gratuito desde que citada a fonte, de acordo com licenças GNU FDL e Creative Commons.

Apesar do uso cada vez mais difundido de buscas para recuperação da informação, estruturas de organização como as *tags* continuam sendo criadas com o intuito de categorizar conteúdo. As tecnologias de *tagging* (ato de criar e compartilhar *tags* aplicadas a algum conteúdo informacional) são utilizadas por pessoas em sistemas de categorização criadas por elas mesmas, sem níveis definidos de organização e que representam a forma como cada um caracteriza aquele conteúdo utilizando palavras-chaves. Essas tecnologias não evitam problemas de redundância (*tags* distintas aplicadas a um mesmo conteúdo ou uma única *tag* aplicada a conteúdos completamente distintos), mas também registram

quais páginas são visitadas e com que intuito ou visão por um grupo de usuários (MCAFEE, 2006).

O emprego de técnicas de *tagging* deu origem a conceitos colaborativos vinculados à necessidade de recuperar informação a partir de experiências anteriores de outros usuários: “*social bookmarking*”, composto pelo conjunto de documentos favoritos publicamente indicados ou visitados por usuários em rede na Web; e folksonomia¹¹. Derivando do termo taxonomia, estrutura hierárquica e pré-definida de classificação de informação, as folksonomias se apresentam como um sistema orgânico de descrição e interação de *tags*:

Folksonomia é o resultado do livre ato pessoal de criar *tags* para informação e objetos (qualquer coisa com uma URL) para a própria recuperação. A criação de *tags* é feita em um ambiente social (normalmente compartilhado e aberto a outros). Folksonomia é criada a partir da criação de *tags* pela pessoa consumindo a informação.

O valor nesta criação externa de *tags* é derivada de pessoas usando o seu próprio vocabulário e incluindo significado explícito, que pode vir de compreensão inferida da informação/objeto. As pessoas não estão necessariamente categorizando, mas sim provendo formas de conectar itens (colocando ganchos) para fornecer o significado de acordo com seu próprio entendimento (VANDER WAL, 2007, tradução nossa).

Existem ainda ferramentas de extensões que automatizam de alguma forma o *social bookmarking* (MCAFEE, 2006). Um exemplo mencionado é a estrutura de recomendações de sites de compras como a Amazon, onde a visualização de determinado produto apresenta também uma lista de outros produtos ou categorias relacionadas, a partir do comportamento de compra e visualização de outros usuários. Assim como as *tags*, são dispositivos empregados para atrelar as aplicações a seus usuários e que fazem com que serviços como o eBay (leilões de produtos pela Internet) e a Amazon, que possuem um grande número de usuários, tenham vantagem competitiva sobre serviços com número inferior de usuários (O'REILLY, 2005).

As tecnologias de sinalização são as que permitem que alertas sobre conteúdo publicado sejam tornados disponíveis a qualquer pessoa que deseje acompanhar as novidades de determinado site ou serviço. A tecnologia RSS (*Really Simple Syndication*) foi desenvolvida com este intuito. Um usuário que assina os sumários RSS de um sítio passa a receber os resumos de conteúdos publicados.

¹¹ Um sítio bastante conhecido por aplicar ambos os conceitos é o Delicious (“<http://del.icio.us>” ou “<http://delicious.com>”).

Com isso, não precisa acessá-lo diretamente para saber se existem atualizações, só o realizando quando deseja saber mais sobre o conteúdo informado pelo resumo.

Dois técnicas de sinalização que permitiram, por exemplo, o desenvolvimento de ferramentas de *blogs* são os *permalinks* (ou *links* permanentes) e os mecanismos de *trackbacks*. Com os primeiros, puderam ser construídos elos diretamente com um *post* específico de um *blog*, facilitando o acesso e o compartilhamento da informação ali contido. Já os segundos permitem que os usuários sejam alertados quando for criado um *link* para o conteúdo sob sua responsabilidade em outra página. Estes mecanismos não criam *links* bidirecionais, o que tornaria o relacionamento entre páginas na Web muito mais organizável e facilitando a recuperação da informação, mas sim apenas dois *links* simétricos e em direções opostas (O'REILLY, 2005).

Um exemplo mais recente de emprego de tecnologias de sinalização é o Twitter (que também permite empregar RSS), um serviço onde cada usuário pode postar mensagens de até 140 caracteres. Mais do que a informação compartilhada por cada usuário, o foco do serviço é permitir que cada usuário acompanhe o que for publicado por pessoas por ele selecionadas, de quem se torna seguidor. Com isso, ele é sinalizado automaticamente por atualizações realizadas por aqueles que "segue". Cada usuário possui, portanto, uma visualização única. Neste tipo de serviço, existe uma criação de rede de relacionamentos entre um usuário e seus seguidores, e destes com seus respectivos seguidores. A criação de redes sociais é provavelmente o campo de maior profusão de ferramentas na Web colaborativa.

Redes sociais na Web colaborativa são criadas com os mais diversos fins: relacionar pessoas com algum vínculo, permitir que elas mantenham contato entre si tanto para fins pessoais quanto profissionais, proporcionar a criação de comunidades agregando usuários com algum interesse em comum para discussão ou simplesmente compartilhamento de informação e conhecimento. Exemplos de sítios famosos são o próprio Twitter, o Orkut (utilizado principalmente por brasileiros), o Facebook, o MySpace (criado inicialmente para a divulgação de conteúdo musical), o LinkedIn (para contatos profissionais) entre tantos outros que surgem constantemente.

Por permitirem o relacionamento entre usuários em um ambiente virtual, estes serviços apenas refletem os benefícios e os riscos da convivência em sociedade. Da mesma forma como estimulam o compartilhamento de informação e conhecimento,

podem servir também para disseminar conteúdo com baixa qualidade, redundante, superficial, inverídico e até mesmo de cunho discriminatório.

Considerando a Web como meio de comunicação, as tecnologias de colaboração possibilitam cada vez mais a participação ativa de seus usuários, através de interfaces de fácil utilização e serviços que lhes sejam de alguma forma úteis. Em meio ao maior interesse dado à RI com o advento da Internet e da Web, é possível supor que um dos motivos que impulsiona o uso destas ferramentas colaborativas é a própria necessidade informacional de cada usuário, independente do nível de relevância (determinado assunto pode ser considerado menos ou mais relevante por um olhar externo, mas isso é um fator que só pode ser definido pelo próprio usuário). Ele pode considerar mais fácil o acesso à informação desejada através da sua rede de contatos explicitada em um serviço na Web, ou em um repositório editado livremente por outros usuários, ou em uma página de opiniões de alguém que considera uma boa fonte de informação.

As tecnologias de colaboração de usuários na Web viriam, desta forma, a representar um tipo de ferramenta de RI. Nos capítulos seguintes, foi descrita a investigação na qual se buscou, em iniciativas colaborativas, características da RI e até mesmo possíveis inovações que poderiam ser aproveitadas pela RI em estudos futuros.

5 A pesquisa por iniciativas de colaboração de usuários na Web

Foram previamente abordados os conceitos de Recuperação da Informação (RI) tradicionais que, influenciados pelo advento de novas TICs como a Internet e a Web, foram contextualizados com a sociedade da informação. Posteriormente, foi realizada a conceituação da colaboração entre usuários na Web e a descrição de algumas das tecnologias que provêm serviços desse tipo.

O passo seguinte foi verificar a possível existência de relacionamento entre a colaboração de usuários na Web e os conceitos de RI, através da análise de trabalhos acadêmicos sobre iniciativas de estudo ou aplicação de ferramentas colaborativas. De forma a padronizar a consulta e facilitar a visualização dos resultados, foi escolhida uma única base de dados de indexação para esse tipo de publicação, a da ferramenta *Web of Science*, descrita a seguir.

5.1 A base de dados *Web of Science*

A base de dados *Web of Science* é integrante da aplicação *ISI (Institute for Scientific Information) Web of Knowledge*, de propriedade do grupo *Thomson Reuters*, conhecido mundialmente pelo foco em notícias e informação para uso cotidiano, profissional e acadêmico. A *Web of Knowledge* é descrita pela empresa proprietária como uma plataforma (aqui entendida como agregadora de várias ferramentas) para pesquisa em ciências, ciências sociais, artes e humanidades, incluindo ferramentas de análise e medição (THOMSON REUTERS, 2010). A empresa opina que a ferramenta compreende diversas necessidades dos pesquisadores, como o uso de funcionalidades de análise, interface unificada e a navegação através das referências explícitas e também implícitas (obtidas através de algoritmos próprios) entre publicações.

Seu diferencial reside no fato de indexar apenas publicações acadêmicas em periódicos, anais de eventos, sítios Web, etc. que contenham tanto referências quanto sejam referenciadas por outras publicações, como uma forma de simbolizar as associações de idéias científicas e permitir a análise dessas associações pelos

usuários. Os critérios de seleção exigem padrões que devem ser seguidos pelas publicações, como convenções bibliotecárias e revisão por pares, conteúdo editorial enriquecedor ou inovador, diversidade internacional e a análise de citações (BLUNDELL, 2005). O acesso é realizado através do endereço Web “<http://apps.isiknowledge.com>”¹².

A *Web of Science* é composta por 7 sub-bases de dados: *Science Citation Index Expanded* (SCI-Expanded), *Social Sciences Citation Index* (SSCI), *Arts & Humanities Citation Index* (A&HCI), *Conference Proceedings Citation Index – Science* (CPCI-S), *Conference Proceedings Citation Index – Social Sciences & Humanities* (CPCI-SSH), *Index Chemicus* (IC) e *Current Chemical Reactions* (CCR-Expanded). As três primeiras bases (SCI-Expanded, SSCI e A&HCI) contêm estudos com referências citadas por autores de artigos; a quarta e a quinta (CPCI-S e CPCI-SSH) indexam as publicações de conferências, simpósios, seminários, colóquios, *workshops* e convenções de diversas áreas, sem a necessidade de citações em outros trabalhos, como forma de disseminar o novo conhecimento; e as duas últimas (IC e CCR-Expanded) são bases temáticas que permitem criar e encontrar compostos e reações químicos (THOMSON REUTERS, 2009).

A base SCI-Expanded indexa de forma multidisciplinar publicações científicas em mais de 6.650 periódicos em 150 áreas, como a Ciência da Computação. A base SSCI indexa publicações de ciências sociais em mais de 1.950 periódicos em 50 áreas distintas, como Ciência da Informação e Biblioteconomia e a Lingüística, além de itens relevantes selecionados de 3.300 periódicos científicos e técnicos. A base A&HCI indexa publicações de artes e humanidades em 1.160 periódicos completos, além de itens relevantes selecionados individualmente em mais de 6.800 periódicos de ciências e ciências sociais (THOMSON REUTERS, 2009).

Após acessar a *Web of Knowledge*, a opção oferecida por padrão é a de pesquisa em todas as ferramentas oferecidas segundo o contrato de uso. A plataforma dispõe de outras ferramentas além da *Web of Science*. No caso específico da CAPES, estão disponíveis o *Derwent Innovations Index*, com informações sobre patentes, e o *Journal Citation Reports*, com métricas para avaliar periódicos de diversas áreas em relação a itens como total de citações, fatores de

¹² Este acesso não é gratuito. A pesquisa só foi possível através de acesso via servidor da Universidade de Brasília (UnB) utilizando as permissões de uso contratadas pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) junto à proprietária.

impacto, etc. Por pertinência ao desejado pela pesquisa (buscar publicações acadêmicas referentes a trabalhos envolvendo colaboração de usuários na Web), foi utilizada apenas a busca na ferramenta *Web of Science*, conforme figura a seguir.

The screenshot shows the ISI Web of Knowledge search interface. At the top, there is a green header with the logo and navigation links. Below it is a navigation bar with 'All Databases', 'Select a Database', 'Web of Science', and 'Additional Resources'. The main search area is titled 'Web of Science' and contains three search fields: 'Search for:' with a dropdown menu (options: Topic, Author, Publication Name), an 'AND' operator dropdown, and a search button. Below the search fields are 'Current Limits' and 'Timespan' options. A sidebar on the right contains promotional text for 'Looking for ISI Proceedings?' and 'Discover Web of Science'.

Figura 2 - Página da ferramenta ISI *Web of Knowledge* para busca na base de dados *Web of Science*

5.2 Parâmetros utilizados para pesquisa

Na busca por artigos na base *Web of Science* foram utilizados os termos “Web”, “user” (usuário) e “collaboration” (colaboração). Como a indexação (sobre a qual a busca é efetivamente realizada) das publicações na *Web of Science* é feita apenas em inglês, independente da linguagem utilizada no conteúdo, os termos se encontram nesta língua. Isso poderia ser visto como um problema considerando a não representatividade de produção indexada em outras linguagens. Porém, a maior parte da produção científica na área é realizada em inglês, inclusive por acadêmicos que falam nativamente outras línguas.

A ferramenta permite que a busca seja realizada nos seguintes campos de dados, inclusive com combinação de campos (por exemplo, com uma palavra-chave para tópico e outra para autor):

- *Topic* (Tópico);
- *Title* (Título);

- *Author* (Autor);
- *Group author* (Autor do grupo);
- *Editor* (Editor);
- *Publication name* (Nome da publicação);
- *Year published* (Ano de publicação);
- *Address* (Endereço);
- *Language* (Idioma);
- *Document type* (Tipo de documento);
- *Funding agency* (Agente de financiamento); e
- *Grant number* (Número de concessão).

Visando à maior abrangência da pesquisa, optou-se pela busca das palavras no campo “Tópico” (*Topic*). A busca por este campo é realizada sobre as palavras-chave associadas aos trabalhos por indexação fornecida pelos próprios autores ou pela *Web of Knowledge*, que possui ferramenta (denominada *KeyWord Plus*) de análise automatizada para termos freqüentemente utilizados nos títulos dos trabalhos referenciados pela publicação a ser indexada (THOMSON REUTERS, 2009).

Outra opção para personalizar a busca fornecida pela ferramenta é a possibilidade de definir um intervalo de tempo em que as publicações foram realizadas. Desde intervalos curtos (publicações realizadas na última semana, nas duas últimas semanas, nas quatro últimas semanas ou no ano vigente) até intervalos longos (últimos cinco anos, ou intervalo de anos definido pelo usuário e até mesmo todos os anos disponíveis – a partir de 1945). Para preservar a viabilidade temporal e de esforço para estudar todos os artigos obtidos na pesquisa e, ao mesmo tempo, abranger o período de maior desenvolvimento das ferramentas de colaboração de usuários na Web, optou-se pela utilização do filtro de limitação para as publicações realizadas nos 5 anos anteriores à pesquisa (*Latest 5 years*).

Conforme mencionado anteriormente, a *Web of Science* é composta de sete sub-bases de indexação. Dessas, estavam disponíveis a *SCI-Expanded*, a *SSCI* e a *A&HCI*, de acordo com o acesso via CAPES/UnB. Considerando não apenas a possibilidade de expandir o número de artigos encontrados, mas principalmente o interesse que o assunto pode despertar em áreas tão distintas quanto a ciência em

geral, as ciências sociais e as artes e humanidades, nenhuma das três bases foi descartada.

Web of Science®

Search for:

Web user collaboration in Topic
Example: oil spill mediterranean*

AND in Author
Example: O'Brian C OR OBrian C**
 Need help finding papers by an author? Use Author Finder.

AND in Publication Name
Example: Cancer OR Journal of Cancer Research and Clinical Oncology*

Add Another Field >>

Search Clear

Current Limits: [\[Hide Limits and Settings\]](#) (To save these permanently, sign in or register.)

Timespan:

Latest 5 years (updated 2010-03-13)

From 1945-1954 to 2010 (default is all years)

Citation Databases:

Science Citation Index Expanded (SCI-EXPANDED)--1945-present

Social Sciences Citation Index (SSCI)--1956-present

Arts & Humanities Citation Index (A&HCI)--1975-present

Figura 3 - Parâmetros utilizados para busca na base de dados *Web of Science*

A consulta utilizando os parâmetros indicados no dia 21 de março de 2010 (segundo a ferramenta naquela data a última atualização da base de dados fora efetuada no dia 20 de março de 2010) retornou 103 registros dos 6.608.602 registros disponíveis como informado pela ferramenta. Ao usuário é dada a possibilidade de refinar o resultado, selecionando um dos seguintes critérios, oferecidos de acordo com o conjunto de resultados já obtido.

- Áreas de assunto;
- Tipos de documento (artigos, anais de eventos, etc.);
- Autores;
- Título da publicação;
- Fonte dos trabalhos;
- Anos de publicação;
- Instituições de autoria;
- Agentes de financiamento;
- Idiomas; e
- Países de publicação.

O usuário pode classificar a apresentação dos resultados pela data de publicação, pela quantidade de vezes em que o trabalho foi citado, pela relevância – baseada em um sistema de *ranking* que considera quantas vezes os termos de busca foram encontrados em cada registro (THOMSON REUTERS, 2009), pelo primeiro autor, pelo título da publicação fonte do trabalho e pelo ano de publicação. Outras possibilidades oferecidas são a análise dos resultados, criando *rankings* de autores, países, agentes de financiamento, número de concessão, tipo de documento, nome da instituição, idioma, ano de publicação, título da publicação, fonte do trabalho e área de assunto; e a geração de relatórios de citação, com o detalhamento por ano do número de trabalhos publicados no conjunto de resultados e do número de citações que receberam.

Nos resultados apresentados, a base *SCI-EXPANDED* esteve representada com 90 trabalhos, a *SSCI* com 30 e a *A&HCI* com apenas 1 trabalho no resultado final. Ao todo, os 103 registros obtidos foram classificados pela própria ferramenta como sendo de 54 áreas de assunto distintas. O resultado da busca foi refinado para apresentar apenas trabalhos classificados em áreas de assunto diretamente relacionadas ao estudo da informação. Áreas cujo foco principal não fosse a informação (como “Ciências e Serviços de Cuidados de Saúde”, com abordagem mais voltada para a área da Saúde) ou que apresentassem maior ênfase na automatização de processos informacionais (por exemplo, “Ciência da Computação, Inteligência Artificial”) foram descartadas para preservar o interesse no processo humano de colaboração para estudo da informação. A seguir é apresentada tabela contendo as áreas de assunto obtidas, a quantidade de registros e as áreas que foram efetivamente selecionadas no refinamento do resultado:

Tabela 1 - Áreas de assuntos de artigos recuperados na base *Web of Science*

Área de assunto	Quantidade de registros	Percentual	Área selecionada
Ciência da Computação, Sistemas de Informação (<i>Computer Science, Information Systems</i>)	17	16,50%	Sim
Ciência da Computação, Teoria e Métodos (<i>Computer Science, Theory & Methods</i>)	15	14,56%	Sim
Ciência da Computação, Engenharia de Software (<i>Computer Science, Software Engineering</i>)	12	11,65%	Sim
Ciência da Informação e Biblioteconomia (<i>Information Science & Library Science</i>)	12	11,65%	Sim

Área de assunto	Quantidade de registros	Percentual	Área selecionada
Ciência da Computação, Inteligência Artificial (<i>Computer Science, Artificial Intelligence</i>)	9	8,74%	Não
Engenharia, Elétrica e Eletrônica (<i>Engineering, Electrical & Electronic</i>)	8	7,77%	Sim
Ciências e Serviços de Cuidados de Saúde (<i>Health Care Sciences & Services</i>)	7	6,80%	Não
Telecomunicações (<i>Telecommunications</i>)	6	5,83%	Sim
Ciência da Computação, Aplicações Interdisciplinares (<i>Computer Science, Interdisciplinary Applications</i>)	5	4,85%	Sim
Ciência da Computação, Cibernética (<i>Computer Science, Cybernetics</i>)	5	4,85%	Não
Informática Médica (<i>Medical Informatics</i>)	5	4,85%	Não
Ciência Nuclear e Tecnologia (<i>Nuclear Science & Technology</i>)	4	3,88%	Não
Ergonomia (<i>Ergonomics</i>)	4	3,88%	Não
Pesquisa em Educação e Educacional (<i>Education & Educational Research</i>)	4	3,88%	Não
Pesquisa Operacional e Ciência de Gerenciamento (<i>Operations Research & Management Science</i>)	4	3,88%	Sim
Sistemas de Automação e Controle (<i>Automation & Control Systems</i>)	4	3,88%	Não
Biotecnologia e Microbiologia Aplicada (<i>Biotechnology & Applied Microbiology</i>)	3	2,91%	Não
Ciências Ambientais (<i>Environmental Sciences</i>)	3	2,91%	Não
Engenharia, Manufatura (<i>Engineering, Manufacturing</i>)	3	2,91%	Não
Métodos de Pesquisa Bioquímica (<i>Biochemical Research Methods</i>)	3	2,91%	Não
Psicologia, Multidisciplinar (<i>Psychology, Multidisciplinary</i>)	3	2,91%	Não
Biologia Matemática e Computacional (<i>Mathematical & Computational Biology</i>)	2	1,94%	Não
Bioquímica e Biologia Molecular (<i>Biochemistry & Molecular Biology</i>)	2	1,94%	Não
Ciência da Computação, Hardware e Arquitetura (<i>Computer Science, Hardware & Architecture</i>)	2	1,94%	Não
Ciências Sociais, Interdisciplinar (<i>Social Sciences, Interdisciplinary</i>)	2	1,94%	Sim
Engenharia, Civil (<i>Engineering, Civil</i>)	2	1,94%	Não
Engenharia, Multidisciplinar (<i>Engineering, Multidisciplinary</i>)	2	1,94%	Não
Gerenciamento (<i>Management</i>)	2	1,94%	Sim
Agricultura, Multidisciplinar (<i>Agriculture, Multidisciplinary</i>)	1	0,97%	Não
Astronomia e Astrofísica (<i>Astronomy & Astrophysics</i>)	1	0,97%	Não
Ciências Multidisciplinares (<i>Multidisciplinary Sciences</i>)	1	0,97%	Não
Ciências Veterinárias (<i>Veterinary Sciences</i>)	1	0,97%	Não
Comunicação (<i>Communication</i>)	1	0,97%	Sim
Conservação da Biodiversidade (<i>Biodiversity Conservation</i>)	1	0,97%	Não
Dermatologia	1	0,97%	Não
Educação, Disciplinas Científicas (<i>Education, Scientific Disciplines</i>)	1	0,97%	Não
Enfermagem (<i>Nursing</i>)	1	0,97%	Não
Engenharia, Ambiental (<i>Engineering, Environmental</i>)	1	0,97%	Não
Engenharia, Biomédica (<i>Engineering, Biomedical</i>)	1	0,97%	Não
Engenharia, Oceânica (<i>Engineering, Ocean</i>)	1	0,97%	Não
Engenharia, Química (<i>Engineering, Chemical</i>)	1	0,97%	Não
Estudos Ambientais (<i>Environmental Studies</i>)	1	0,97%	Não
Física, Nuclear (<i>Physics, Nuclear</i>)	1	0,97%	Não
Genética e Hereditariedade (<i>Genetics & Heredity</i>)	1	0,97%	Não
Geoquímica e Geofísica (<i>Geochemistry & Geophysics</i>)	1	0,97%	Não

Área de assunto	Quantidade de registros	Percentual	Área selecionada
Instrumentos e Instrumentação (<i>Instruments & Instrumentation</i>)	1	0,97%	Não
Medicina Integrativa e Complementar (<i>Integrative & Complementary Medicine</i>)	1	0,97%	Não
Negócios (<i>Business</i>)	1	0,97%	Não
Psicologia, Clínica (<i>Psychology, Clinical</i>)	1	0,97%	Não
Psicologia, Experimental (<i>Psychology, Experimental</i>)	1	0,97%	Não
Psiquiatria (<i>Psychiatry</i>)	1	0,97%	Não
Recursos Hídricos (<i>Water Resources</i>)	1	0,97%	Não
Saúde Pública, Ambiental e Ocupacional (<i>Public, Environmental & Occupational Health</i>)	1	0,97%	Não
Tecnologia de Construção e Edificação (<i>Construction & Building Technology</i>)	1	0,97%	Não

Fonte: THOMSON REUTERS, 2009, tradução nossa

Notas: O somatório de registros por área de assunto (176) supera o total de registros recuperados (103) pois um mesmo registro pode estar classificado em mais de uma área de assunto. Percentual calculado a partir do número de registros por área de assunto dividido pelo total de registros recuperados (103) e não pelo somatório de registros por área de assunto (176). Desta forma, pelo mesmo motivo da nota acima, o somatório dos percentuais supera os 100%.

Dados percentuais arredondados

O número total de trabalhos obtidos considerando o refinamento por áreas de assunto foi de 58 publicações. A análise, realizada com o auxílio das funcionalidades da própria ferramenta *Web of Science*, das características do conjunto de resultados é detalhada no Apêndice A. A seguir, a tabela de trabalhos selecionados para execução da pesquisa:

Tabela 2 - Artigos recuperados na base *Web of Science*, organizados pelo número de citações

Item	Título	Autores	Ano de publicação	Citações
1	<i>Many eyes: A site for visualization at internet scale</i>	Viegas, F. et al.	2007	26
2	<i>Virtual reality for collaborative e-learning</i>	Monahan, T.; McArdle, G.; Bertolotto, M.	2008	10
3	<i>Knowledge-intensive collaborative design modeling and support Part II: System implementation and application</i>	Zha, X.; Du, H.	2006	7
4	<i>Unifying human and software services in Web-scale collaborations</i>	Schall, D.; Truong, H.; Dustdar, S.	2008	5
5	<i>Marketing - Harnessing the power of the oh-so-social web</i>	Bernoff, J.; Li, C.	2008	4
6	<i>Emerging technologies changing our service delivery models</i>	Kajewski, M.	2007	4
7	<i>Structured collaborative workflow design</i>	Held, M.; Blochinger, W.	2009	3
8	<i>Groups formation and operations in the Web 2.0 environment and social networks</i>	Lai, L.; Turban, E.	2008	3

Item	Título	Autores	Ano de publicação	Citações
9	<i>VLab: collaborative Grid services and portals to support computational material science</i>	Nacar, M. et al.	2007	3
10	<i>Meta-searches in peer-to-peer networks</i>	Lehikoinen, J. et al.	2006	3
11	<i>Activity management as a Web service</i>	Cozzi, A. et al.	2006	3
12	<i>Design and implementation of Web services middleware to support fair non-repudiable interactions</i>	Cook, N.; Robinson, P.; Shrivastava, S.	2006	3
13	<i>Flexible collaboration over XML documents</i>	Ignat, C.; Norrie, M.	2006	3
14	<i>Understanding affective commitment, collectivist culture, and social influence in relation to knowledge sharing in technology mediated learning</i>	Hwang, Y.; Kim, D.	2007	2
15	<i>Developing subject-related web sites collaboratively: The Virtual Business Information Center</i>	Abels, E.; White, M.; Kim, S.	2007	2
16	<i>Using common sense to recognize cultural differences</i>	Anacleto, J. et al.	2006	2
17	<i>Friends night out - A working prototype of a blended lifestyle service enabled through IMS</i>	Chiang, T. et al.	2006	2
18	<i>A system for international telemedicine through integrated synchronous/asynchronous collaboration</i>	Nishantha, D. et al.	2006	2
19	<i>Collaborative attack on Internet users' anonymity</i>	Puzis, R. et al.	2009	1
20	<i>Science 2.0: Tools catalogue and consequences for scientific activity</i>	Cabezas-Clavijo, A.; Torres-Salinas, D.; Delgado-Lopez-Cozar, E.	2009	1
21	<i>Performance evaluation for implementations of a network of proxy caches</i>	Kumar, C.	2009	1
22	<i>WebClass: Adding rigor to manual labeling of traffic anomalies</i>	Ringberg, H.; Soule, A.; Rexford, J.	2008	1
23	<i>Virtual integration of existing web databases for the genotypic selection of cereal cultivars</i>	Bergamaschi, S.; Sala, A.	2006	1
24	<i>Opportunistic collaboration - Unlocking the archives of the Birmingham Institute of Art and Design</i>	Everitt, S.	2005	1
25	<i>Principles and Tools for Collaborative Entity-Based Intelligence Analysis</i>	Bier, E.; Card, S.; Bodnar, J.	2010	0
26	<i>Ciclope Robot: Web-Based System to Remote Program an Embedded Real-Time System</i>	Lopez, D. et al.	2009	0
27	<i>Service-Oriented Policy Management for Web-Application Frameworks</i>	Feeney, K.; Lewis, D.; O'Sullivan, D.	2009	0
28	<i>Developing a collective intelligence application for special education</i>	Gregg, D.	2009	0
29	<i>Content-Centered Collaboration Spaces in the Cloud</i>	Erickson, J. et al.	2009	0
30	<i>An investigation of coopetitive pedagogic design for knowledge creation in Web-based learning</i>	Fu, F.; Wu, Y.; Ho, H.	2009	0
31	<i>Research evaluation and citation analysis: key issues and implications</i>	Herther, N.	2009	0
32	<i>Information portal development and practice at Shanghai Jiao Tong University Library</i>	Jin, Y.; Peng, J.	2009	0
33	<i>Towards Environment-to-Environment (E2E) multimedia communication systems</i>	Singh, V. et al.	2009	0

Item	Título	Autores	Ano de publicação	Citações
34	<i>Discovering discovery tools Evaluating vendors and implementing Web 2.0 environments</i>	James, D.; Garrett, M.; Krevit, L.	2009	0
35	<i>Cultivating Personalized Museum Tours Online and On-Site</i>	Wang, Y. et al.	2009	0
36	<i>Engineering E-Collaboration Processes to Obtain Innovative End-User Feedback on Advanced Web-Based Information Systems</i>	Bragge, J.; Merisalo-Rantanen, H.	2009	0
37	<i>Arizona Literature Mapper: An Integrated Approach to Monitor and Analyze Global Bioterrorism Research Literature</i>	Dang, Y. et al.	2009	0
38	<i>A novel method for personalized music recommendation</i>	Lu, C.; Tseng, V.	2009	0
39	<i>The Web Archives Workbench (WAW) Tool Suite: Taking an Archival Approach to the Preservation of Web Content</i>	Hswe, P. et al.	2009	0
40	<i>Social tagging, online communication, and Peircean semiotics: a conceptual framework</i>	Huang, A.; Chuang, T.	2009	0
41	<i>Service Broker Architecture: Location Business Case and Mashups</i>	Loreto, S. et al.	2009	0
42	<i>The VuFind implementation at Villanova University</i>	Houser, J.	2009	0
43	<i>Using Web 2.0 for scientific applications and scientific communities</i>	Pierce, M. et al.	2009	0
44	<i>Applying adaptive swarm intelligence technology with structuration in web-based collaborative learning</i>	Huang, Y.; Liu, C.	2009	0
45	<i>Identity Management and Web Services as Service Ecosystem Drivers in Converged Networks</i>	Yelmo, J.; Trapero, R.; del Alamo, J.	2009	0
46	<i>The DBin platform: A complete environment for Semantic Web Communities</i>	Tummarello, G.; Morbidoni, C.	2008	0
47	<i>Voyagers and Voyeurs: Supporting Asynchronous Collaborative Visualization</i>	Heer, J.; Viegas, F.; Wattenberg, M.	2009	0
48	<i>Supporting teachers in collaborative student modeling: A framework and an implementation</i>	Gaudioso, E. et al.	2009	0
49	<i>Collaborative Web Search Based on User Interest Similarity</i>	Godoy, D.; Amandi, A.	2008	0
50	<i>Collaborative web agent based on friend network</i>	Kim, B. et al.	2008	0
51	<i>GramyaVikas: A distributed collaboration model for rural development planning</i>	Adinarayana, J. et al.	2008	0
52	<i>InOrder: enhancing Google via stigmergic query refinement</i>	Camp, G.; Ulieru, M.	2007	0
53	<i>A web-based user friendly simulator of optical fiber communication systems</i>	Razzak, M.; Hamam, H.	2007	0
54	<i>Supporting customised collaboration over shared document repositories</i>	Ignat, C.; Norrie, M.	2006	0
55	<i>Design and implementation of a mobile class Web site using intelligent user interface</i>	Hong, Y.; Jun, W.; Kwak, B.	2006	0
56	<i>Collaborative virtual learning model for Web intelligence</i>	Fiaidhi, J.; Mohammed, S.	2006	0
57	<i>Sharing protected web resources using distributed role-based modeling</i>	Encheva, S.; Tumin, S.	2006	0
58	<i>A cooperative model for wide area content delivery applications</i>	Rashkovits, R.; Gal, A.	2005	0

Fonte: THOMSON REUTERS, 2009, tradução nossa

Utilizando as funcionalidades da *Web of Knowledge*, foram gerados dois gráficos contendo a distribuição nos cinco anos anteriores à pesquisa da publicação dos artigos selecionados, bem como a quantidade de citações do conjunto. O ano de 2010, em ambos os gráficos, estava representado até o período da última atualização da base de dados *Web of Science*, isto é, 20 de março de 2010.

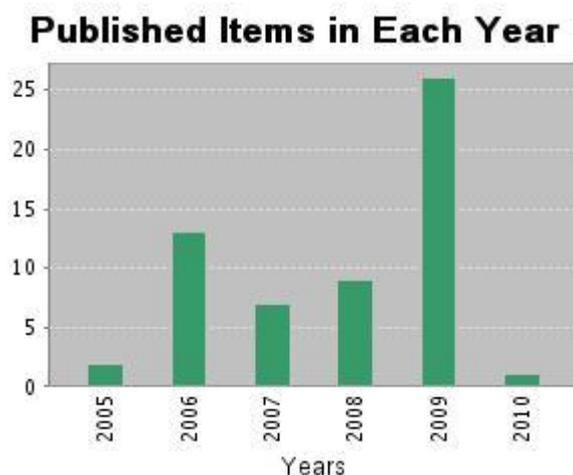


Figura 4 - Publicações por ano dos artigos recuperados na base *Web of Science* (THOMSON REUTERS, 2009)

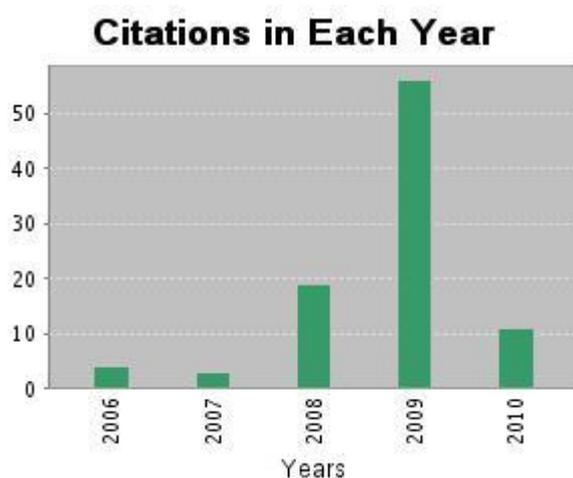


Figura 5 - Citações por ano dos artigos recuperados na base *Web of Science* (THOMSON REUTERS, 2009)

Os artigos selecionados no conjunto de resultados da busca após o refinamento foram analisados, a fim de confirmar a existência de características da RI influenciada pelo advento da Internet e da Web e do desenvolvimento da sociedade da informação, nos trabalhos que descrevem ferramentas ou pesquisas de colaboração de usuários na Web. Para permitir esta análise, foram selecionados

20 tópicos que resumem a RI sob ambas as abordagens, de acordo com a descrição realizada nos capítulos anteriores deste estudo.

Além das características de RI, buscou-se em cada artigo idéias e práticas aplicadas às iniciativas descritas e que possam ser exploradas por outros estudos da RI, como forma de estimular a inovação na área. O capítulo seguinte contém as discussões referentes à confrontação entre as generalizações conceituais e os resultados efetivamente obtidos na pesquisa.

6 Discussão

Na revisão sobre RI realizada nos capítulos anteriores, foram destacados conceitos que definem e caracterizam o escopo de estudo da área, desde o seu advento e considerando os novos tempos de Internet, Web e da sociedade da informação. Resumindo a abordagem foram extraídos do referencial teórico 20 tópicos contendo generalizações de conceitos e de assuntos de interesse da RI, fundamentando a análise objetiva dos artigos recuperados na base de dados *Web of Science*. Estes itens foram classificados em cinco grupos de acordo com a correlação entre eles:

- A) Conceitos básicos de RI
 - 1. Necessidades de informação
 - 2. Acesso à informação
 - 3. Subjetividade dos usuários
 - 4. Uso da informação
- B) Sistemas de RI
 - 5. Sistemas informatizados
 - 6. Grande volume de informação
 - 7. Armazenamento
 - 8. Escalabilidade de usuários
 - 9. Avaliação de sistemas
- C) Estruturas de sistemas
 - 10. Estruturação de dados
 - 11. Diversos formatos
 - 12. Indexação automatizada
 - 13. *Ranking* de resultados
- D) Interfaces de sistemas
 - 14. Interfaces de usuário
 - 15. Apresentação da informação
 - 16. Personalização de resultados
- E) Sociedade da informação
 - 17. Recomendações/Classificações de usuários
 - 18. Análise de *links* como agentes de contexto

19. Acessos concentrados em poucos registros

20. Inclusão social

O grupo A (“Conceitos básicos de RI”) foi composto pelos principais objetivos e preocupações de estudo da RI, relacionados às necessidades informacionais, ao acesso e ao uso da informação e aos aspectos subjetivos dos usuários interessados em informação.

A presença dos sistemas de RI foi investigada pelo grupo B (“Sistemas de RI”). Foram observadas características como a descrição de sistemas de RI, metodologias de avaliação aplicadas a eles, além da preocupação com o gerenciamento de grande volume de informação, com estruturas de armazenamento de informação para uso futuro e o crescimento do número de usuários, também conhecida como a escalabilidade dos sistemas.

O grupo C (“Estruturas de sistemas”) englobou itens de interesse quando da análise da estruturação e do funcionamento de sistemas. Foram verificadas características como a existência de modelos para estruturação de dados, a abrangência de formatos de informação diversos ao texto convencional, bem como a implementação de funcionalidades de automação de indexação e de ordenação (*ranking*) de resultados.

Um aspecto de interesse de investigação nos sistemas de RI diz respeito às interfaces com usuários. O grupo D (“Interfaces de sistemas”) investigou tópicos como a descrição de interfaces de usuários, preocupações com a apresentação da informação e a existência de funcionalidades de personalização de resultados de buscas por informação.

O grupo E (“Sociedade da informação”) abordou possíveis benefícios e preocupações surgidas com o maior valor dado à informação na sociedade atual. Aspectos como o uso de recomendações e/ou classificações realizadas por usuários e de análise de *links* para contextualização de informação se encontram como técnicas com potenciais de desenvolvimento para a RI. Por outro lado, o problema representado pela concentração de acessos a poucos registros (o dilema cauda longa x cauda curta apresentado anteriormente) e a preocupação em utilizar a RI como forma de combater a exclusão social também foram observadas pelo estudo.

Os artigos recuperados na pesquisa utilizando a base de dados *Web of Science* foram analisados individualmente de forma a verificar quais desses tópicos

se encontravam abordados ou descritos pelos autores em cada texto, correlação descrita no quadro 5. As duas primeiras colunas informam o número de ordem e qual a característica relacionada. As demais colunas apresentam o número de ordem dos artigos estudados, organizados decrescentemente pelo número de citações realizadas a eles por outras publicações, de acordo com a tabela 2. A descrição da referência bibliográfica, da classificação do tipo de artigo e dos possíveis pontos de interesse e inovação em RI de cada publicação analisada está presente no Apêndice B.

Dos 58 artigos inicialmente obtidos na consulta à base *Web of Science*, foram efetivamente analisados 39 trabalhos (67,24%¹³). Considerando que o escopo pretendido pela pesquisa era o de ferramentas, análises ou iniciativas que demonstrassem, estudassem ou sugerissem o uso da colaboração de usuários na Web, 12 artigos apresentaram-se como não sendo pertinentes ao estudo, mesmo apresentando as três palavras chaves utilizadas (“Web”, “user” e “collaboration”). Dessa forma, é possível afirmar que pelo menos 20,69% dos resultados obtidos na busca não refletiam a real necessidade de informação da pesquisa, seja por indexação incorreta dos artigos ou mesmo falta de maior definição dos parâmetros para busca. Trata-se de experiência prática dos desafios da RI na definição e representação das necessidades de informação pelo usuário que realiza a busca e na representação da informação pela indexação de cada artigo na base de dados.

Além dos artigos não pertinentes à abrangência do estudo, sete outros trabalhos (12,07%) não foram analisados porque não estavam acessíveis para consulta. O acesso aos 39 artigos analisados foi realizado das seguintes maneiras: através de *link* direto na descrição dos artigos na ferramenta *Web of Science*, em periódicos assinados pela CAPES e disponíveis para acesso via conexão a servidor da UnB, pesquisando em serviços Web de busca especializados em publicações acadêmicas (normalmente direcionando para os periódicos assinados pela CAPES) e pesquisando em serviços Web de busca comuns (por vezes determinado artigo não presente em uma base de periódicos estava disponível nas páginas pessoais dos autores).

¹³ Todos os percentuais apresentados com 2 casas decimais são valores arredondados, utilizando como divisor o número 39 (total de artigos analisados), à exceção dos percentuais de artigos analisados, que usa como divisor o número 58 (total de artigos obtidos na consulta à base *Web of Science*)

Art.	Característica	1	2	3	4	5	6	7	8	9	10	11	13	16	17	20	24	25	27	28	29
A) Conceitos básicos de RI																					
1	Necessidades de informação		X	X	X		X	X	X		X	X					X	X		X	X
2	Acesso à informação	X	X	X	X		X	X	X	X		X	X		X	X	X	X		X	X
3	Subjetividade dos usuários	X			X	X	X	X	X	X		X					X	X		X	
4	Uso da informação	X	X					X	X		X	X	X	X	X	X	X	X	X	X	X
B) Sistemas de RI																					
5	Sistemas informatizados	X	X	X	X			X		X	X	X		X	X			X	X	X	X
6	Grande volume de informação									X											
7	Armazenamento			X				X		X										X	
8	Escalabilidade de usuários			X	X	X		X	X	X						X				X	X
9	Avaliação de sistemas		X					X				X		X				X		X	
C) Estruturas de sistemas																					
10	Estruturação de dados	X		X	X	X	X	X		X	X	X	X			X	X	X	X		X
11	Diversos formatos	X	X			X	X	X	X			X			X	X	X	X	X		X
12	Indexação automatizada				X						X	X		X		X		X			X
13	Ranking de resultados				X						X			X		X			X		
D) Interfaces de sistemas																					
14	Interfaces de usuário	X	X	X	X		X	X	X	X		X	X	X	X		X	X	X	X	X
15	Apresentação da informação	X	X				X					X		X			X	X		X	
16	Personalização de resultados	X		X			X	X	X		X	X		X	X	X		X	X	X	X
E) Sociedade da informação																					
17	Recomendações/Classificações de usuários		X			X		X	X		X					X		X	X	X	X
18	Análise de links como agentes de contexto			X												X		X			
19	Acessos concentrados em poucos registros								X												
20	Inclusão social																			X	

Descrição de método/ferramenta
 Revisão de literatura
 Estudo de caso
 Relatório de pesquisa

Quadro 5 - Características de RI localizadas nos artigo analisados

Art.	Característica	30	32	33	34	35	37	38	39	40	42	43	44	46	47	48	49	50	51	54	
A) Conceitos básicos de RI																					
1	Necessidades de informação		X	X	X	X		X	X	X		X				X	X	X	X		
2	Acesso à informação	X	X	X	X	X	X			X		X	X	X	X					X	
3	Subjetividade dos usuários	X		X		X		X		X	X		X	X			X	X	X		
4	Uso da informação	X	X	X	X	X	X			X			X		X	X			X	X	
B) Sistemas de RI																					
5	Sistemas informatizados		X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	
6	Grande volume de informação								X			X								X	
7	Armazenamento				X															X	
8	Escalabilidade de usuários			X			X											X			
9	Avaliação de sistemas					X	X	X	X				X	X	X	X	X	X	X		
C) Estruturas de sistemas																					
10	Estruturação de dados		X	X	X	X	X	X	X	X	X	X	X	X	X	X					X
11	Diversos formatos	X	X	X	X	X		X		X	X	X	X	X	X			X	X		
12	Indexação automatizada		X	X			X	X	X	X	X		X		X		X		X		
13	Ranking de resultados			X	X	X		X		X	X			X	X	X	X	X			
D) Interfaces de sistemas																					
14	Interfaces de usuário	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X		X	X	
15	Apresentação da informação		X		X	X	X			X	X			X	X		X				
16	Personalização de resultados		X	X	X	X		X		X	X	X		X			X	X	X		
E) Sociedade da informação																					
17	Recomendações/Classificações de usuários					X		X	X	X	X	X		X	X		X	X			
18	Análise de <i>links</i> como agentes de contexto						X		X	X							X	X			
19	Acessos concentrados em poucos registros							X													
20	Inclusão social	X																	X		

■ Descrição de método/ferramenta ■ Revisão de literatura
■ Estudo de caso ■ Relatório de pesquisa

Quadro 5 - Características de RI localizadas nos artigos analisados (Continuação)

Os sete artigos não acessados estavam incluídos em anais de eventos, com visualização restrita por seus editores ou organizadores. Desses artigos, um estava disponível para acesso via periódicos controlados pela Oxford University Press até 2009, quando os anais do evento em questão foram assumidos pela própria organizadora. Este trabalho apareceu em 18º posição no resultado da pesquisa organizado pelo *ranking* de citações da *Web of Science*. Os outros seis foram apresentados entre as piores classificações (52, 53, 55, 56, 57 e 58) do mesmo *ranking*, o que permite supor que a restrição ao acesso destes estudos contribuiu negativamente para o uso deles como referências por outros autores da comunidade científica.

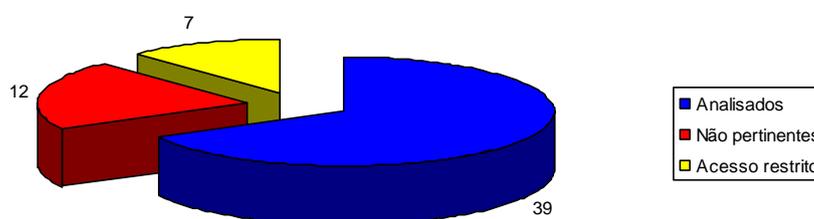


Figura 6 - Situação de análise dos artigos recuperados na base *Web of Science*

Os artigos analisados foram classificados como sendo de quatro tipos: descrição de método/ferramenta (27 artigos, ou 69,23%), revisão de literatura (seis artigos, ou 15,38%), estudos de caso (cinco artigos, ou 12,82%) e relatório de pesquisa (um artigo, ou 2,56%). Como o interesse era identificar a aplicação ou a sugestão de aplicação das características de RI nos artigos que abordam a colaboração de usuários na Web, essa classificação foi realizada com o intuito de facilitar a análise de cada artigo de acordo com o seu tipo. Citações de características de RI se concentravam em publicações descritivas de métodos/ferramentas no desenvolvimento das iniciativas que apresentavam, e não no referencial teórico. Em artigos de revisão de literatura os autores as apresentaram principalmente em capítulos de discussão. Nos estudos de caso elas apareceram dispersas em diferentes partes dos textos. Já no relatório de pesquisa elas são encontradas principalmente na fundamentação da pesquisa. Apesar desta ênfase em partes específicas, todos os artigos foram inteiramente analisados.

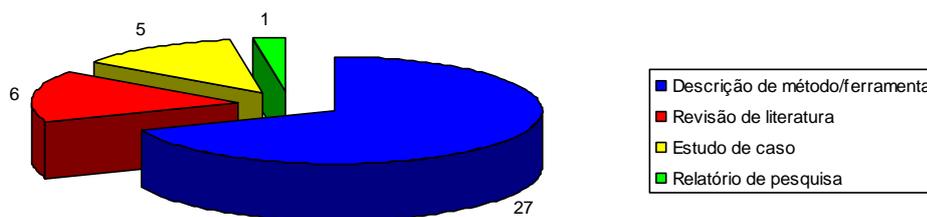


Figura 7 - Quantidade de artigos analisados por tipo

Entre os grupos de características da RI apresentados anteriormente, observa-se que o grupo A é o que concentra as definições básicas do objetivo de estudo da RI. Não houve nenhum artigo analisado que não apresentasse ao menos um dos itens do grupo. Ou seja, o grupo foi representado em 100% das publicações. Além disso, todas as características deste grupo foram localizadas em uma grande quantidade de artigos, sendo a mais presente “Acesso à informação” (em 28 artigos, 71,79%) e a menos presente “Subjetividade dos usuários” (em 22 artigos, 56,41%). Os itens “Uso da informação” e “Necessidade de informação” foram encontrados em 27 (69,23%) e 24 (61,54%) artigos, respectivamente.

Artigo	Necessidades de informação	Acesso à informação	Subjetividade dos usuários	Uso da informação
1		X	X	X
2	X	X		X
3	X	X		
4	X	X	X	
5			X	
6	X	X	X	
7	X	X	X	X
8	X	X	X	X
9		X	X	
10	X			X
11	X	X	X	X
13		X		X
16				X
17		X		X
20		X		X
24	X	X	X	X
25	X	X	X	X
27				X
28	X	X	X	X
29	X	X		X
30		X	X	X

32	X	X		X
33	X	X	X	X
34	X	X		X
35	X	X	X	X
37		X		X
38	X		X	
39	X			
40	X	X	X	X
42			X	
43	X	X		
44		X	X	X
46		X	X	
47		X		X
48	X			X
49	X		X	
50	X		X	
51	X	X	X	X
54				X

Quadro 6 - Características de RI do grupo A por artigo analisado

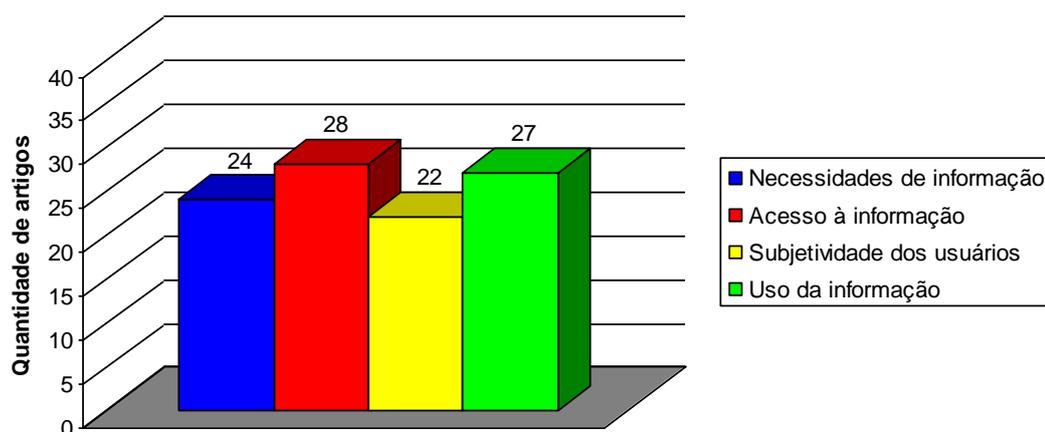


Figura 8 - Características de RI do grupo A encontradas nos artigos analisados

É possível estabelecer a preocupação dos autores dos artigos em aumentar a disponibilidade de informação normalmente armazenada em alguma base de dados. Isso se dava através do desenvolvimento de novos sistemas, novas técnicas de divulgação ou simplesmente aumentando a abrangência do acesso destas informações na Web, incluindo o uso de novos dispositivos, principalmente móveis. Muito além de apenas buscar maior visibilidade, essas iniciativas tenderam a ser concebidas levando em consideração o contexto do público-alvo (leitores) dos artigos, o que remete às três outras características do grupo A.

A relevância dada pelos autores ao contexto de aplicação das iniciativas de colaboração na Web foi fundamentada com o estudo e definição de quais necessidades informacionais dos usuários suas abordagens iriam atender e com o estudo e a definição do efetivo uso que eles poderiam fazer da informação obtida a partir delas. Foram explicitadas ainda as possíveis variações de necessidade e de uso de informação causadas pela subjetividade de representação e interpretação dos próprios usuários.

As características classificadas sob o grupo B são as relativas aos sistemas de RI. A presença do item “Sistemas informatizados” ocorreu em 30 artigos (76,92%). Observou-se uma relação deste item com os artigos do tipo descrição de método/ferramenta. Apenas um dos 27 artigos assim classificados não mencionou a existência de sistema na abordagem, restringindo-se à especificação de metodologia sem a aplicação em um sistema informatizado. As demais características aparecem em proporções menores: “Avaliação de sistemas” em 17 artigos (43,59%), “Escalabilidade de usuários” em 12 (30,77%), “Armazenamento” em sete (17,95%) e “Grande volume de informação” em quatro (10,26%).

Artigo	Sistemas informatizados	Grande volume de informação	Armazenamento	Escalabilidade de usuários	Avaliação de sistemas
1	X				
2	X				X
3	X		X	X	
4	X			X	
5				X	
6					
7	X		X	X	X
8				X	
9	X	X	X	X	
10	X				
11	X				X
13					
16	X				X
17	X				
20				X	
24					
25	X				X
27	X				
28	X		X	X	X
29	X			X	
30					
32	X				

33	X			X	
34	X		X		
35	X				X
37	X			X	X
38	X				X
39	X	X			X
40					
42	X				
43		X			
44	X				X
46	X				X
47	X				X
48	X				X
49	X				X
50	X			X	X
51	X	X	X		X
54	X		X		

Quadro 7 - Características de RI do grupo B por artigo analisado

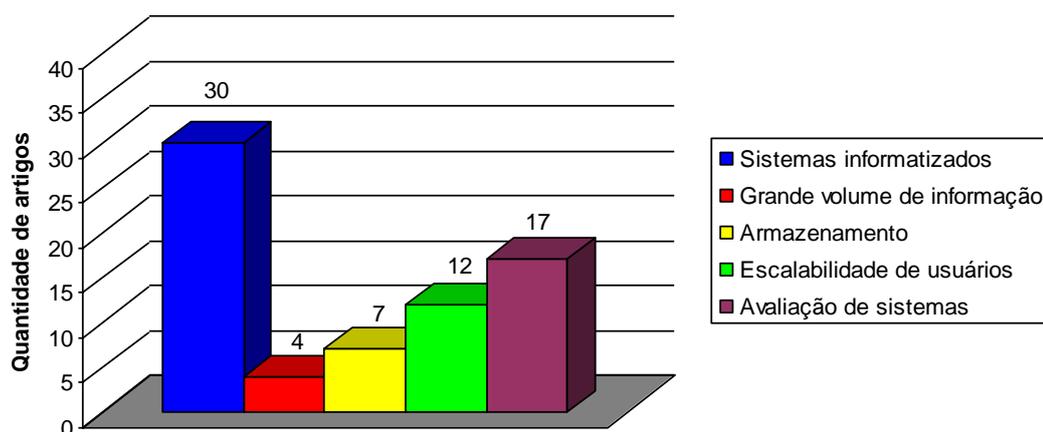


Figura 9 - Características de RI do grupo B encontradas nos artigos analisados

A descrição de sistemas informatizados ocorreu em grande parte dos artigos, mas verificou-se que a freqüência não se repetiu para os demais itens. É especialmente notável que haja o detalhamento do desenvolvimento, do funcionamento, ou mesmo da sugestão de sistemas de informação, ignorando as respectivas metodologias para avaliação de efetividade para os usuários. Outro ponto a considerar é que, mesmo em se tratando de iniciativas de colaboração na Web, baseadas na participação e interação cada vez maior dos usuários com os sistemas, preocupações com o crescimento do número de usuários (escalabilidade) foram abordadas por menos de um terço dos artigos analisados.

Técnicas de tratamento do maior volume de informação disponível e questões ligadas ao armazenamento de informação foram mencionadas em número ainda menor. Com o acentuado desenvolvimento das tecnologias de processamento nos últimos tempos, é possível supor que os autores tiveram maiores preocupações em estudar aspectos qualitativos dos sistemas, como o contexto dos usuários, do que propriamente estes fatores relacionados a aspectos quantitativos de performance.

O grupo C, referente a características de estruturação e funcionalidades de sistemas, apresentou os seguintes valores para cada item: “Estruturação de dados” em 30 artigos (76,92%), “Diversos formatos” em 25 (64,10%), “Indexação automatizada” em 18 (46,15%) e “*Ranking* de resultados” em 15 (38,46%).

Artigo	Estruturação de dados	Diversos formatos	Indexação automatizada	<i>Ranking</i> de resultados
1	X	X		
2		X		
3	X			
4	X		X	X
5	X			
6	X	X		
7	X	X		
8		X		
9	X			
10	X		X	X
11	X	X	X	
13	X			
16			X	X
17		X		
20	X	X	X	X
24	X	X		
25	X	X	X	
27	X	X		X
28				
29	X	X	X	
30		X		
32	X	X	X	
33	X	X	X	X
34	X	X		X
35	X	X		X
37	X		X	
38	X	X	X	X
39	X		X	
40	X		X	
42	X	X	X	X
43	X	X		
44	X	X	X	

46	X	X		X
47	X	X	X	X
48	X			X
49			X	X
50		X		X
51		X	X	
54	X			

Quadro 8 - Características de RI do grupo C por artigo analisado

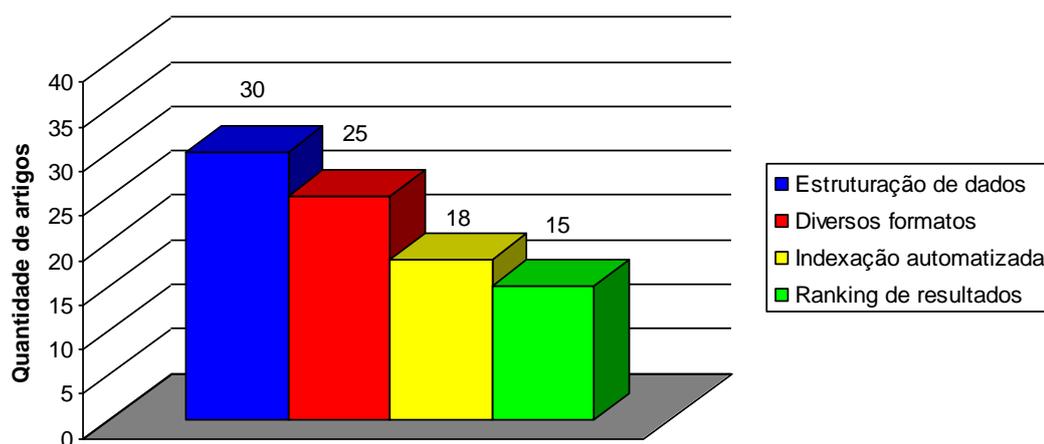


Figura 10 - Características de RI do grupo C encontradas nos artigos analisados

Os autores apresentaram interesse na maioria dos artigos em prever o uso ou descrever padrões de dados para facilitar o armazenamento e a posterior recuperação, normalmente utilizando formatos flexíveis baseados em XML. A presença dessa característica refletia também a preocupação ou a necessidade de os sistemas de informação manterem conteúdo não apenas em texto, mas em formatos diversos, como figuras, vídeos e sons, expandindo as opções de representação da informação para os seus usuários.

Embora citadas com relevante freqüência, funcionalidades como a geração e/ou leitura de índices automatizados para criar pontos de acesso para os usuários e algoritmos de ordenação de resultados de buscas em sistemas de informação foram mencionadas em menos artigos que os outros dois itens do grupo. É possível considerar que tais funcionalidades apresentavam-se até certo tempo atrás como promessas de desenvolvimento para a RI. Atualmente, se encontram presentes em várias iniciativas, mas não se tornaram unanimidade pelo fato da especialização necessária (do sistema, do contexto de sua atuação e da equipe por ele

responsável) para implementação, podendo vir a impactar o custo de desenvolvimento de forma a ser considerada não vantajosa a sua utilização.

A abordagem das interfaces de sistemas no grupo D teve a citação ou descrição de “Interfaces de usuários” como a característica que, individualmente, apresentou a maior quantidade de anotações nos artigos analisados dentre todos os itens avaliados na pesquisa, em 33 artigos (84,62%). Os demais tópicos do grupo, “Personalização de resultados” e “Apresentação da informação”, ocorreram em 26 (66,67%) e 16 (41,03%) artigos, respectivamente.

Artigo	Interfaces de usuário	Apresentação da informação	Personalização de resultados
1	X	X	X
2	X	X	
3	X		X
4	X		
5			
6	X	X	X
7	X		X
8	X		X
9	X		
10			X
11	X	X	X
13	X		
16	X	X	X
17	X		X
20			X
24	X	X	
25	X	X	X
27	X		X
28	X	X	X
29	X		X
30	X		
32	X	X	X
33	X		X
34	X	X	X
35	X	X	X
37	X	X	
38	X		X
39	X		
40			X
42	X	X	X
43	X		X
44			
46	X	X	X
47	X	X	
48	X		

49	X	X	X
50			X
51	X		X
54	X		

Quadro 9 - Características de RI do grupo D por artigo analisado

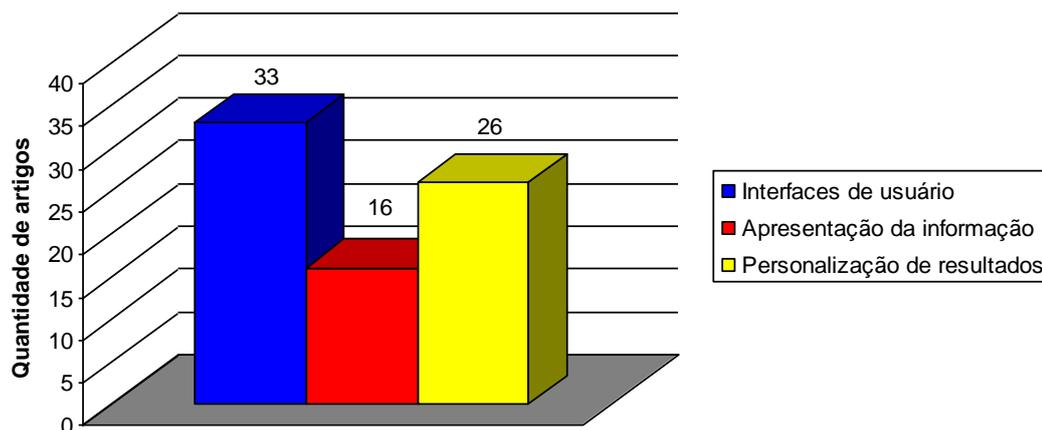


Figura 11 - Características de RI do grupo D encontradas nos artigos analisados

O interesse em interfaces de usuários se tornou quase uma constante em estudos práticos de sistemas de informação, justificando a preocupação dos autores de grande maioria dos artigos de descrição de métodos/ferramentas e de todos os estudos de caso. Além disso, verificou-se que metade dos artigos de revisão de literatura apresentou também algum tipo de análise teórica sobre esse assunto. Por outro lado, alguns poucos artigos de descrição de ferramentas informatizadas não deram nenhum tipo de destaque a este ponto atualmente considerado fundamental para a efetividade dos sistemas de informação.

Funcionalidades de personalização de resultados de acesso à informação para usuários (ou grupos de usuários) específicos podem ser consideradas como exemplos de possibilidades de melhoria para a apresentação da informação. Ao dividir a análise em dois tópicos, foi possível visualizar o destaque dado às iniciativas de contextualização das interfaces, maior do que às demais preocupações com a Arquitetura de Informação para a Web (vinculada à organização da informação para facilitar o seu uso). Se considerada apenas como uma única característica para análise, ela estaria presente em 30 dos artigos analisados.

O grupo E, vinculando os benefícios e preocupações de interesse da RI na relação entre informação e sociedade da informação, apresentou os seguintes

valores para suas características: “Recomendações/Classificações de usuários” em 20 artigos (51,28%), “Análise de *links* como agentes de contexto” em oito (20,51%), “Inclusão social” em três (7,69%) e “Acesso concentrado em poucos registros” em dois (5,13%).

Artigo	Recomendações/ Classificações de usuários	Análise de <i>links</i> como agentes de contexto	Acessos concentrados em poucos registros	Inclusão social
1				
2	X			
3		X		
4				
5	X			
6				
7	X			
8	X		X	
9				
10	X			
11				
13				
16				
17				
20	X	X		
24				
25	X	X		
27	X			
28	X			X
29	X			
30				X
32				
33				
34				
35	X			
37		X		
38	X		X	
39	X	X		
40	X	X		
42	X			
43	X			
44				
46	X			
47	X			
48				
49	X	X		
50	X	X		
51				X
54				

Quadro 10 - Características de RI do grupo E por artigo analisado

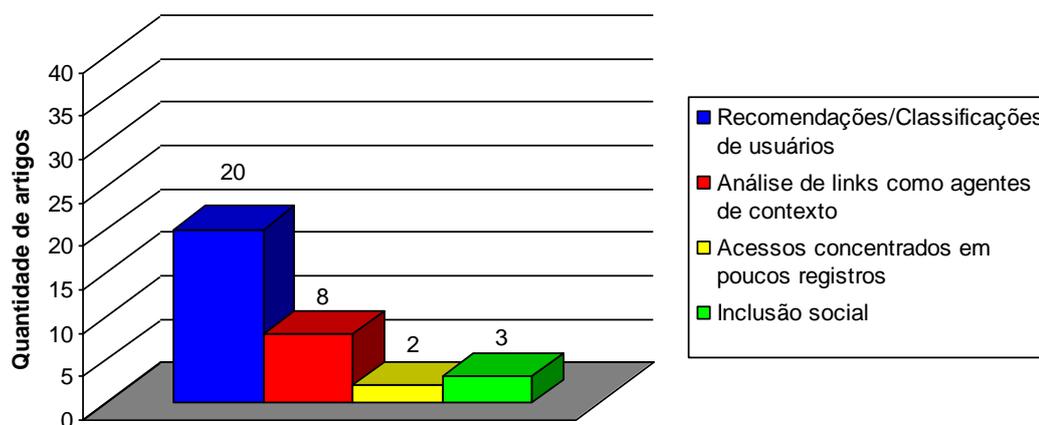


Figura 12 - Características de RI do grupo E encontradas nos artigos analisados

Mais da metade dos artigos apresentou ou sugeriu o reuso de recomendações/classificações obtidas (explícita ou implicitamente) a partir de experiências anteriores de utilização para empreender melhorias nos resultados de busca por informação de outros usuários. Trata-se do emprego efetivo da colaboração de usuários não apenas para armazenamento de informação, mas também em técnicas de processamento de buscas. Entretanto, a análise de *links*, outra abordagem com potencial para o desenvolvimento da RI a partir da contextualização do grande volume de informação na Web, foi citada em um número pequeno de estudos, demonstrando que tal funcionalidade ainda se encontra pouco explorada ou foi considerada de difícil implementação pelos autores.

Se por um lado a citação de possíveis benefícios do compartilhamento da informação variou entre números elevado e baixo, a abordagem dos problemas surgidos no cenário foi muito menor. Preocupações com a concentração de acessos em poucos registros informacionais e com medidas de inclusão social foram descritas em apenas cinco dos artigos analisados. De uma forma geral, observou-se nos artigos o interesse maior dos autores em citar benefícios potenciais ou até mesmo efetivos de suas abordagens para uma visão restrita e otimista de um determinado contexto, em detrimento de análises mais aprofundadas sobre a ocorrência abrangente destes problemas não só nas iniciativas já existentes, mas também para evitar o surgimento deles nos seus próprios desenvolvimentos. Além de continuarem sendo necessárias abordagens que procurem entender e diminuir

essas dificuldades, elas podem ser majoradas caso desconsideradas em novas pesquisas. De qualquer forma, mesmo os poucos artigos que têm essas características representam iniciativas interessantes para o desenvolvimento e emprego da RI na sociedade da informação.

Uma medida utilizada para ponderar a representatividade de cada grupo em relação aos demais foi obter a média da quantidade de vezes em que as características de cada grupo foram encontradas nos artigos analisados. Por exemplo, as quatro características do grupo A somaram 101 citações nos artigos analisados, uma média de 25,25 citações para o grupo nos 39 artigos. Obedecendo a este cálculo, os grupos ordenados pela média obtida são: Grupo A (“Conceitos básicos de RI”), Grupo D (“Interfaces de sistemas”), Grupo C (“Estruturas de sistemas”), Grupo B (“Sistemas de RI”) e Grupo E (“Sociedade da informação”).

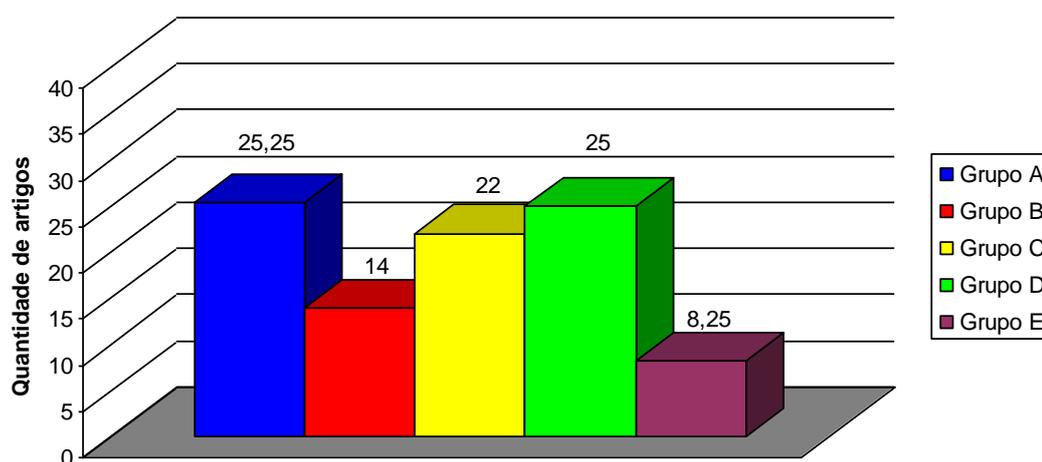


Figura 13 - Média de citações por grupo nos artigos analisados

Além de uma média de 25,25 citações nos artigos analisados, a maior entre os grupos, o grupo A foi o único em que ao menos uma de suas características foi encontrada em cada um dos artigos analisados, conforme mencionado anteriormente. Como é o grupo que engloba os conceitos básicos de RI, é possível afirmar que todos os artigos do conjunto continham texto de interesse para a área, independente do enfoque inicialmente pretendido por seus autores. Isso ensejou a não desconsideração de nenhum dos artigos analisados para a etapa de discussão dos resultados.

A média de 14 citações do grupo B nos artigos analisados reflete a contradição entre a grande quantidade de artigos que abordam sistemas de

informações e o menor interesse dispensado à avaliação destes sistemas, ao suporte para maior escalabilidade de usuários e, principalmente, aos aspectos quantitativos dos sistemas. Os papéis menores destinados às avaliações de sistemas e à escalabilidade de usuários, ao contrário do pouco destaque para as questões de performance e armazenamento explicado pelos avanços tecnológicos, podem ser entendidos como deficiências dos autores em abordar a análise da efetividade, atual e futura, dos sistemas.

O grupo C obteve uma média de 22 citações nos artigos analisados. A presença média acima da metade do número de artigos revela que as características de estruturação de dados, formatos diversos de conteúdo e funcionalidades de indexação automatizada e algoritmos de ordenação de resultados, de forma geral, tiveram interesse relevante para os autores dos artigos.

Com a média de 25 citações nos artigos analisados, o grupo D ficou atrás apenas do grupo A neste critério de classificação. O alto valor de representatividade demonstra a importância já estabelecida das interfaces de usuários em sistemas de informação, principalmente em iniciativas de colaboração na Web, fundamentadas na necessidade de prover rica experiência e interação para estímulo de seus usuários.

Além de apresentar, entre todos os grupos, o menor valor de citações para a característica mais encontrada, o grupo E conteve também os dois itens que, individualmente, obtiveram a menor quantidade de anotações nos artigos analisados entre todas as 20 características selecionadas. Isso se refletiu no grupo com menor média de citações (8,25) nos artigos analisados. A abordagem de benefícios para a RI baseados no compartilhamento de informação e conhecimento na sociedade atual foram citados de forma mediana nos artigos, sugerindo haver espaço para maior desenvolvimento destes pontos em pesquisas em colaboração de usuários na Web. Notório foi o pouco destaque à necessidade de considerar as dificuldades causadas pelo mesmo cenário, não apenas como sugestão para novos estudos e desenvolvimentos, mas também como variáveis a serem avaliadas em quaisquer outras iniciativas.

Analisando os resultados da pesquisa considerando o número de características encontradas em cada artigo, verificou-se que a média de itens observados por artigo é de aproximadamente 9,41 características (ou cerca de 47,05% das 20 características escolhidas para análise). Nos dois artigos que

apresentaram menos características, foram observados quatro itens (20%), enquanto que no artigo que mais características apresentou, foi obtido o número de 14 itens (70%).

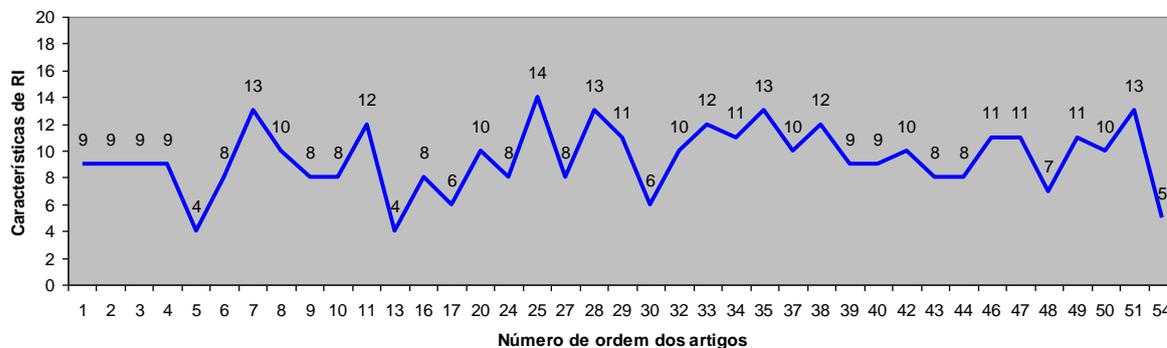


Figura 14 - Quantidade de características de RI encontradas por artigo analisado

A discussão a respeito das características do grupo A indicou a pertinência de todos os artigos analisados para o estudo, independente da quantidade de características de RI encontradas em cada artigo. Observou-se que fatores que causaram impacto no gráfico acima estão relacionados não apenas à abordagem de cada autor, mas também ao tipo de artigo (os que menos apresentaram itens eram revisões de literatura) e até mesmo o próprio tamanho do trabalho.

Além dos tópicos selecionados para análise dos artigos, estes também foram observados com o objetivo de extrair possíveis contribuições para o avanço da pesquisa em RI, não apenas vinculado à colaboração de usuários. Por vezes, alguns destes pontos de interesse chegaram a ser mencionados no referencial teórico deste estudo. Porém, não puderam ser considerados como características de RI, mas tão somente como possibilidades de desenvolvimento da área, o que se confirmou ao encontrá-los na análise dos artigos.

O detalhamento dos pontos de interesse para RI por artigo está presente no Apêndice B. Verificou-se a existência de relacionamento entre os diversos itens, permitindo ordená-los sob três áreas de atuação para o desenvolvimento da RI: usuários, contextualização e sistemas. É possível observar que alguns tópicos se enquadram em mais de uma área.

Usuários	<ol style="list-style-type: none"> 1. Acesso estimulado pela colaboração; 2. Aprendizagem de usuários por interação social; 3. Autoridade específica para grupos/usuários sobre recursos; 4. Confiança em usuários de um mesmo grupo; 5. Controle de conteúdo pelos próprios usuários; 6. Diversas formas/modelos de colaboração; 7. Estruturação e definição de objetivos para os usuários; 8. Interação e diversão para melhor desempenho dos usuários; 9. Modelos de confiabilidade e regras de filtragem de informação; 10. Realidade virtual; 11. Recomendações para colaboração em outras atividades ou com outros usuários; 12. Redes sociais (Relacionamento e recomendações); 13. Significância da informação reconhecida por colaboradores externos; e 14. Usuários com necessidades especiais (Público-alvo).
Contextualização	<ol style="list-style-type: none"> 1. Acesso por dispositivos móveis; 2. Base compartilhada para usuários externos; 3. Compartilhamento e permissão de acesso a versões intermediárias de produtos de informação; 4. Conteúdo criado por usuários; 5. Contexto obtido a partir de dispositivos móveis; 6. Estratégias de atualização de documentos compartilhados em repositórios para colaboração de usuários; 7. Ferramentas colaborativas para personalização e divulgação de informação e <i>feedback</i> de usuários; 8. Fontes de informação distintas utilizadas por uma aplicação/interface única; 9. Índices e informações (inclusive históricos de buscas) compartilhados em máquinas em rede (P2P¹⁴); 10. Informações das ontologias acessadas em ferramentas tradicionais, como <i>e-mail</i>; 11. Recombinação de informação; 12. Redes sociais (Relacionamento e recomendações); e 13. Sistema distintos de consulta e de processamento de consulta.

¹⁴ P2P é o acrônimo de *Peer to peer*, uma arquitetura de redes de comunicação independente de um ponto central para distribuição de informação. A comunicação de dados ocorre entre nós diretamente ligados.

Sistemas	<ol style="list-style-type: none"> 1. Análise de folksonomias; 2. Análise de <i>tags</i> (inclusive por semântica, semiótica e ontologias para linguagem natural); 3. Análise semântica e mapeamento de conhecimento a partir de domínio definido; 4. Auto-organização de conteúdo criado por usuários; 5. Cálculo de relevância baseado no compartilhamento de experiências passadas de buscas de usuários; 6. <i>Cloud Computing</i>¹⁵; 7. Combinação de indexação automática e indexação colaborativa humana; 8. Estratégias de atualização de documentos compartilhados em repositórios para colaboração de usuários; 9. Estruturas de políticas de serviços, de conteúdo, de dados, de serviços e de regras flexíveis; 10. Ferramentas de descoberta de informação (<i>Discovery tools</i>); 11. Histórico de buscas do próprio usuário; 12. Índices e informações (inclusive históricos de buscas) compartilhados em máquinas em rede (P2P); 13. Inteligência artificial para contextualização a partir de descrições em redes semânticas; 14. Interação entre ontologias para obter nova informação; 15. Ontologias (Web semântica); 16. Recomendações de <i>tags</i> baseadas no vocabulário utilizado em sistemas diferentes por um mesmo usuário; 17. Recomendações para colaboração em outras atividades ou com outros usuários; 18. Redes sociais (Relacionamento e recomendações); 19. Sincronização de buscas com mudanças na base de dados; 20. Sistema distintos de consulta e de processamento de consulta; e 21. Sistema de Gerenciamento de Conteúdo (CMS – <i>Content Management System</i>).
----------	---

Quadro 11 - Pontos de interesse para a RI encontrados nos artigos analisados

O primeiro conjunto de pontos de interesse para a RI encontrados nos artigos analisados se constitui na abordagem voltada para os usuários, envolvendo aspectos que de forma subjetiva intensificam a experiência na interação deles com sistemas de informação e com outros usuários, além de estudar problemas como confiabilidade, controle de alteração de informação e autoria.

Nas iniciativas estudadas são descritas características que indicam os benefícios do emprego da colaboração na Web para estímulo ao acesso dos sistemas pelos usuários. O próprio fato de serem considerados sistemas de colaboração os torna mais interessantes para que as pessoas os utilizem, o que se reflete em possível percepção de melhor efetividade. Existem ainda diversos modelos para emprego da colaboração, mesmo de forma implícita (por exemplo, o histórico de navegação ou busca de um usuário pode ser utilizado para auxiliar a experiência de outro), aplicáveis a sistemas mais abrangentes ou para um uso

¹⁵ Execução de tarefas de processamento em serviços disponíveis remotamente em servidores interligados.

específico. Isso permite melhor definição dos objetivos para os usuários e até mesmo o direcionamento para públicos-alvo específicos, inclusive para a promoção de inclusão social.

Os benefícios para a interação de usuários se fundamentam, entre outros fatores, na transparência do funcionamento dos sistemas colaborativos. Os usuários não devem precisar se preocupar com as estruturas internas dos sistemas, dotados de interface intuitiva para estimular o uso. A interação social é facilitada através de relacionamentos em redes sociais (inclusive com recomendações dos próprios sistemas), resultando em melhorias para o compartilhamento, atribuição de significados para a informação e para a aprendizagem.

A criação de vínculos entre os usuários em redes sociais permite ainda que questões como controle de conteúdo, confiabilidade e autoria sejam de domínio do próprio grupo de usuários. Isso não representa necessariamente a falta de níveis de autoridade sobre a informação ou sobre recursos, ou a inexistência de regras para filtragem de informação e permissões de confiabilidade entre usuários. Os sistemas colaborativos devem prever a flexibilidade da definição destes níveis e regras pelos próprios usuários.

O segundo conjunto de tópicos de interesse para a RI engloba características das iniciativas de colaboração na Web, presentes nos artigos analisados, que atuam sobre a definição de contexto da atuação dos sistemas, influenciando seus objetivos, sua disponibilidade e a utilidade que eles terão para seus usuários ou mesmo para outros sistemas.

O fato de os sistemas colaborativos lidarem com conteúdo criado pelos próprios usuários, relacionados entre si em redes sociais, determina a necessidade de melhorias em técnicas de disponibilidade da informação. Por isso, é maior o interesse em estratégias para controle que permitam a visualização e a edição concorrente de informação, além da obtenção de informação em outras fontes disponíveis através dos próprios relacionamentos dos usuários, como em dispositivos interligados através de redes P2P.

Uma necessidade cada vez mais presente para os sistemas é o acesso através de dispositivos móveis. Isso deve influenciar não apenas aspectos de disponibilidade (infra-estrutura de rede, segurança, etc.) e interfaces de sistemas, mas também permitir que eles possam adquirir informações contextuais dos usuários que possibilite maior relevância para os processos de RI. Um exemplo já

implementado em alguns serviços de busca é a procura por estabelecimentos comerciais em um celular. Caso o mecanismo reconheça o posicionamento do usuário, ele pode sugerir resultados que estejam em uma área próxima a que se encontra, o que em tese seria de maior utilidade para o usuário.

Outro fenômeno que influencia o uso dos sistemas é a recombinação da informação. A existência de diversas fontes e bases de informação e a livre permissão de publicação na Web possibilita que tanto usuários como serviços informatizados não apenas reutilizem informação, mas também a recombinaem para gerar novo conteúdo. Isso pode ocasionar problemas de autoria, como a indefinição da real fonte de informação; redundância, replicação de informação em diferentes locais; e, portanto, de qualidade. Uma saída apontada nos artigos é a previsão para que as próprias bases de informação ou sistemas de resposta disponham de interfaces externas, acessadas por interfaces de outros sistemas ou serviços especializados de consulta, que então fariam a personalização da informação (simplificando a visualização) de acordo com os perfis dos usuários que atendem.

O último conjunto de pontos de interesse de RI encontrados nos artigos analisados envolve as características que atuam sobre o funcionamento interno dos sistemas: algoritmos para funcionalidades de análise, auto-organização de conteúdo, cálculo de relevância, indexação e recomendações; estruturação da informação; e infra-estrutura de funcionamento, compartilhamento e sincronização.

A colaboração de usuários na Web definiu novas possibilidades e também dificuldades para os sistemas. Por um lado, podem ser desenvolvidas funcionalidades para análise semântica de *tags*, folksonomias, históricos de informação do próprio usuário ou compartilhados em sua rede social, gerando indexação automatizada, admitindo indexação colaborativa e permitindo melhorias nos cálculos de relevância e de recomendações. Por outro, se torna necessário desenvolver algoritmos de organização do conteúdo que é gerado livremente pelos usuários e de gerenciamento de atualização de informação no ambiente colaborativo.

A flexibilidade das estruturas de padronização de dados deve ser prevista nos sistemas de informação de forma a permitir que os usuários se tornem mais livres para descrever conteúdo. Com isso, os sistemas não apenas podem comportar diversos formatos, mas também se torna possível o intercâmbio de informação entre

sistemas diferentes, desde que sejam compartilhadas também as estruturas de dados. Isso ocorre com iniciativas que utilizam ontologias, como a Web semântica.

As infra-estruturas dos sistemas de informação são influenciadas pela colaboração de usuários na Web não apenas com o compartilhamento em rede. O maior volume de informação resulta na grande quantidade de tarefas de RI executadas, em serviços estruturados em *Cloud Computing* (disponibilizados remotamente por servidores interligados), ou em sistemas especializados de consulta interligados a outros serviços especializados em respostas de consultas. A necessidade de encontrar informação com maior valor estimula a busca em estruturas interligadas, como ontologias, inclusive em computadores em redes P2P, através de ferramentas de descoberta (*discovery tools*). Já a constante mudança nas informações disponíveis estimula a sincronização ou pelo menos o alerta aos usuários das modificações nas bases de dados que impactem os resultados de buscas por eles realizadas.

Uma vez descritas características que indicam os artigos analisados de iniciativas de colaboração de usuários na Web como pertinentes ao estudo da RI, bem como de tópicos abordados com maior ou menor interesse pelos autores e previstos como objetivos ou preocupações da RI, além de pontos de interesse para o desenvolvimento da RI presentes nos artigos, o capítulo seguinte contém a confrontação entre os resultados encontrados e os objetivos propostos pela pesquisa, de acordo com a sua temática.

7 Conclusão

A partir da importância e da presença constante da RI na sociedade da informação, este estudo teve como intuito pesquisar se iniciativas de colaboração de usuários na Web abrangem conceitos de RI. A investigação dessa relação pôde abordar também de que forma a RI se desenvolve ou pode se desenvolver baseada nessas novas ferramentas.

Para fundamentar a análise pretendida pela pesquisa, foi necessário o estudo de conceitos da RI, da sua relação com a CI e de características do seu desenvolvimento histórico, incluindo os sistemas de RI, métodos de avaliação, avanços e dificuldades. Associando os objetivos e subjetividades que a área apresenta com o crescente valor da informação para o desenvolvimento da sociedade atual, foi possível descrever a relevância que a RI assume nas mais variadas tarefas do cotidiano, mesmo sem que as pessoas percebam isso.

A Web foi descrita como um novo meio de comunicação de informação e como ambiente de especial interesse para a RI. A análise não apenas de seu alcance social, mas das transformações que promove ou ajuda a promover sobre os seus usuários, ou grupo de usuários, se tornou de fundamental importância para o escopo e o desenvolvimento da RI.

Entre as novas ferramentas criadas para a Web destacam-se as aplicações colaborativas, baseadas na participação de usuários para compartilhamento de informação. Uma vez descritas as relações entre a RI e a Web, observou-se a falta de aprofundamentos em pesquisas que buscassem estendê-las às novas iniciativas de colaboração de usuários na Web.

Com o estudo de definições de RI e de RI na Web foi possível delimitar 20 características de RI que, agrupadas em cinco sub-áreas de interesse, permitiram analisar:

- 1) Se existe relação entre as iniciativas que envolvem a colaboração de usuários na Web e a RI, configurando as primeiras como de interesse para o estudo da segunda;
- 2) Qual a frequência com que essas características foram empregadas nas iniciativas; e
- 3) Qual a relevância dessas características nas iniciativas.

A forma encontrada para aplicar essas características na análise das iniciativas de colaboração de usuários na Web foi através da busca de artigos que abordassem o assunto. Foi escolhida uma das maiores bases de dados de informação acadêmica do mundo, a *Web of Science*, onde foi realizada busca por artigos cuja indexação contivesse três palavras-chave (“Web”, “User” e “Collaboration”), que fossem classificados como sendo de áreas de pertinência aos objetivos da pesquisa e publicados nos cinco anos anteriores à pesquisa. Após a aplicação destes parâmetros, 58 artigos foram obtidos. Destes, doze não se enquadravam na pesquisa por não abordarem efetivamente a colaboração de usuários na Web e outros sete não puderam ser analisados pois apresentavam restrição ao acesso de seus conteúdos.

O primeiro grupo de características estabelecidas previamente continha definições conceituais da RI. Com isso, a localização desses itens nos artigos analisados os enquadrava como de interesse para o estudo da RI, tornando-os aptos para as demais etapas da avaliação. Dos 39 artigos efetivamente analisados, todos apresentaram ao menos uma das características do grupo. Além disso, esse foi o grupo com maior representatividade em citações nos artigos. Conclui-se que todas as iniciativas de colaboração de usuários na Web abordadas pelos autores continham objetivos ou preocupações comuns à RI (possivelmente até mesmo de forma inconsciente), demarcando a sua relação com a colaboração de usuários na Web, o que sugere o desenvolvimento de novas pesquisas em RI que efetivamente empreguem técnicas de colaboração de usuários na Web.

Outros três grupos abordaram os sistemas informatizados, elementos mais comumente associados à RI na operação cotidiana pelos usuários. Verificou-se que os autores se apresentavam menos preocupados com questões quantitativas associadas à performance do sistema ou mesmo de armazenamento de dados, do que em relação a questões como estruturação de dados, compreensão de formatos diversos ao texto, interfaces de usuários e personalização na apresentação da informação. Essa diferenciação pode ser considerada como um sinal dos avanços no desenvolvimento de infra-estruturas de *hardware* e até mesmo de comunicação de dados, não necessariamente acompanhados por evoluções na efetividade com que as ferramentas informatizadas (*software*) atendem a seus usuários.

Apesar do foco dos autores em questões qualitativas dos sistemas de RI, tópicos anteriormente estudados e considerados como de importância para a RI

ainda se apresentaram com destaque mediano ou abaixo da média nos artigos analisados. Enquadram-se nessa situação a efetiva descrição de métodos de avaliação de sistemas de RI e a preocupação com a escalabilidade do número de usuários, imprescindíveis para garantir a eficácia e a eficiência de sistemas colaborativos. Além dessas características, observou-se a possibilidade de maior aplicação de técnicas para a indexação automatizada e para a ordenação de resultados, funcionalidades abordadas pela RI na Web, mas com novas possibilidades de emprego com as aplicações colaborativas.

O último grupo de características para análise abordou benefícios e desafios para a RI observados com o desenvolvimento da sociedade da informação. Nos artigos analisados, apenas o uso das recomendações/classificações de usuários foi observado acima da média como uma forma de melhorar a efetividade da RI. A análise de *links* para auxílio à definição de contexto da informação, outro ponto estudado como portador de possível melhoria para a área, foi pouco abordada. Com frequência ainda menor, apresentaram-se as dificuldades em relação à concentração de acessos em poucos registros de informação, o que pode fazer com que sistemas de RI desconsiderem os demais registros – mesmo que apresentem informações relevantes às necessidades informacionais dos usuários; e iniciativas que considerem o uso da colaboração de usuários na Web como forma de desenvolver a inclusão social de públicos-alvo específicos, preocupação legítima para a RI, mas que necessita de maior quantidade de soluções efetivas.

Uma vez que todos os artigos analisados apresentaram características do grupo A e foram considerados como pertinentes ao estudo da RI, se tornou possível indicar, além das 20 características determinadas quando do estudo do referencial teórico, outros pontos de interesse que podem representar caminhos para o desenvolvimento de novas pesquisas em RI. Estes foram classificados em três áreas (Usuários, Contextualização e Sistemas), de acordo com a correlação entre eles e com o viés em que podem representar avanços para a ciência.

Na abordagem de usuários, devem ser observadas as oportunidades da colaboração de usuários na Web em relação a estímulos para interação e aprendizagem dos usuários entre si e com os sistemas de informação; considerando a transparência destes últimos, a preocupação com o desenvolvimento das interfaces e o fortalecimento de vínculos em redes sociais.

Eventuais melhorias em contextualização influenciam a necessidade e a utilidade dos sistemas de RI. Um fenômeno típico da Web e que apresenta grande potencial para maiores estudos é a recombinação da informação, baseada na existência e no acesso (inclusive em redes P2P) de diversas fontes e na livre permissão para publicação, gerando novo conteúdo e também problemas em relação à autoria e à qualidade da informação. Outra área de interesse nessa classificação diz respeito ao emprego cada vez maior de dispositivos móveis para acesso à Web e reconhecimento do contexto do usuário, que podem representar avanços em efetividade de RI.

Foram destacados pontos de interesse para o desenvolvimento de sistemas de RI baseados em colaboração de usuários na Web. Basicamente foram obtidos benefícios e dificuldades criados com as novas formas de participação de usuários e compartilhamento de informação. Assim como deve ser prevista a flexibilidade das estruturas de dados e dos formatos, devem ser previstas funcionalidades de compartilhamento de serviços (para execução em paralelo de tarefas de processamento de informação) e de descoberta de informação em redes de dados criadas a partir do relacionamento dos usuários.

O desenvolvimento desta pesquisa permitiu, portanto, identificar características básicas da RI (incluindo a sua relação com a sociedade da informação e com Web) que puderam ser analisadas em artigos acadêmicos de iniciativas de colaboração de usuários na Web de forma a caracterizá-los como de interesse da RI. Foram ainda identificados outros tópicos conceituais da RI, cuja presença nos artigos permite não apenas qualificar o efetivo emprego da RI nessas iniciativas, mas também indicar caminhos potenciais para novos estudos. Finalmente, considerando todos os artigos analisados como pertinentes ao estudo da RI, foram destacados pontos que podem representar novos desenvolvimentos para o cumprimento dos objetivos da RI em relação ao seu papel perante a sociedade da informação.

7.1 *Sugestões de trabalhos futuros*

Como o intuito principal da pesquisa era relacionar a RI e a colaboração de usuários na Web, as discussões e as conclusões apresentadas se tornam passos iniciais para outros estudos que, de forma geral, estendam os vínculos entre elas.

Em relação à própria metodologia aplicada, podem ser realizadas variações, como o uso de bases de dados distintas, indexando artigos de eventos específicos, nacionais ou internacionais; e o emprego de palavras-chave diferentes, mas relacionadas à colaboração de usuários na Web, como *groupware*, participação, cooperação, etc. Os resultados obtidos a partir de tais variações serviriam para enriquecer a discussão a respeito do relacionamento entre RI e colaboração de usuários na Web.

Na análise das características de RI em artigos de colaboração de usuários na Web, observou-se a possibilidade de desenvolvimento de pesquisas com maior destaque à escalabilidade de usuários, avaliação, indexação automatizada e *ranking* de resultados em sistemas de RI colaborativos. Além disso, há a necessidade de explorar não apenas benefícios para a RI da sociedade da informação, como as recomendações/classificações de usuários e a análise de *links* como agentes de contexto, mas principalmente problemas como a concentração de acessos em poucos registros informacionais e a exclusão social.

Da mesma forma como características de RI podem apresentar sugestões para desenvolvimento de pesquisas relacionando a área e a colaboração de usuários na Web, foram destacados nos artigos pontos de interesse classificados como pertinentes para novas abordagens relacionadas a usuários, contextualização e sistemas de RI em trabalhos futuros.

Por exemplo, pesquisas em RI voltadas para o estudo de usuários podem ser direcionadas para compreender e empregar o estímulo do uso de sistemas de RI quando em formato colaborativo, ou mecanismos de confiança e recomendações em redes sociais. Estudos em contextualização de sistemas podem destacar o uso de dispositivos móveis ou diferentes fontes e usos (recombinação) da informação. Por fim, as próprias estruturas de sistemas de RI, como modelagem de dados, políticas de uso, entre outras, deverão ser projetadas de forma a se tornarem flexíveis para melhor representação por seus usuários.

Referências

ADEVERIA. **Timeline of web browsers**. 2010. Altura: 1.780 pixels. Largura: 2.400 pixels. 52 Kb. Formato SVG. Disponível em: <http://commons.wikimedia.org/wiki/File:Timeline_of_web_browsers.svg>. Acesso em 2 maio 2010.

AMAZON.COM, INC. **amazon.com**. [S.l.]: Amazon.com, Inc, 2010. Disponível em: <<http://www.amazon.com>>. Acesso em: 18 maio 2010.

ANDERSON, C. The Long Tail. **Wired**, San Francisco, n. 12.10, Oct. 2004. Disponível em: <<http://www.wired.com/wired/archive/12.10/tail.html>>. Acesso em 26 fev. 2010.

ARAYA, E.; VIDOTTI, S. Direito autoral e tecnologias de informação e comunicação no contexto da produção, uso e disseminação de informação: um olhar para as Licenças Creative Commons. **Informação & Sociedade: Estudos**, João Pessoa, v. 19, n. 3, p. 39-51, set./dez. 2009. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/3900>>. Acesso em 10 maio 2010.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval**. New York: ACM Press, 1999. 499 p.

BARBOSA, R. R.; SEPÚLVEDA, M. I. M.; COSTA, M. U. P. Gestão da informação e do conhecimento na era do compartilhamento e da colaboração. **Informação & Sociedade: Estudos**, João Pessoa, v. 19, n. 2, p.13-24, maio/ago. 2009. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/2378>>. Acesso em 27 out. 2009.

BATES, M. J. Conceptualizing Users and Uses. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY ANNUAL MEETING, 2002, Philadelphia. **Oral presentation...** Philadelphia: American Society for Information Science and Technology, 2002.

BBC BRASIL. **Adolescentes estão perdendo interesse nos blogs, indica pesquisa**. [S.l.]: BBC Brasil, 2010. Disponível em: <http://www.bbc.co.uk/portuguese/noticias/2010/02/100204_blogs_pesquisa_rw.shtml>. Acesso em 10 maio 2010.

BELKIN, N. J. Progress in Documentation: information concepts for Information Science. **Journal of Documentation**. Londres, v. 34, n. 1, p. 55-85, mar. 1978.

BERNERS-LEE, T. **Weaving the Web**. New York: HarperCollins Publishers Inc., 2000. 256 p.

BLUNDELL, J. **ISI Web of Knowledge**. [S.l.]: Thomson Scientific, oct. 2005. Disponível em: <<http://www.biblioteche.unibo.it/portale/materiali-didattici/la-piattaforma-wok/WoKItaly.pdf>>. Acesso em 13 maio 2010.

BORKO, H. Information science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, 1968.

BRASIL. Decreto nº 5.296, de 2 de dezembro de 2004. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 2 dez. 2004. Disponível em: <http://www.planalto.gov.br/ccivil/_ato2004-2006/2004/Decreto/D5296.htm>. Acesso em: 08 fev. 2010.

BUSH, V. As we may think. **The Atlantic Monthly**, [S.l.], July, 1945. Disponível em: <<http://www.theatlantic.com/doc/194507/bush>>. Acesso em 06 nov. 2009

CALADO, P. Dados – Informação – Dados. Uma Versão Circular da História da RI. In: Workshop Internacional em Ciência da Informação, 2009, Brasília. **Apresentação oral...** Brasília: Programa de Pós-Graduação em Ciência da Informação da Universidade de Brasília, 2009.

CAPURRO, R.; HJORLAND, B. O Conceito de informação. **Perspectivas em Ciência da Informação**, tradução de Ana Maria Pereira Cardoso, Maria da Glória Achtschin Ferreira e Marco Antônio de Azevedo, Belo Horizonte, v. 12, n. 1, p. 148-297, jan./abr. 2007. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/view/54>>. Acesso em 17 nov. 2009.

CASTELLS, M. **A sociedade em rede**. Tradução: Roneide Venâncio Majer. São Paulo: Paz e Terra, 1999. 698 p. (A era da informação: economia, sociedade e cultura, v. 1)

CERVO, A. L.; BERVIAN, P. A.; DA SILVA, R. **Metodologia científica**. 6. ed. São Paulo: Pearson Prentice Hall, 2007. 162 p.

CLEVERDON, C.; MILLS, J.; KEEN, M. **ASLIB Cranfield Research Project**. Factors determining the performance of indexing systems. Volume 1. Design. Part 2. Appendices. Cranfield: College of Aeronautics, 1966. 177 p.

CREATIVE COMMONS. **Creative Commons**. [S.l.]: Creative Commons, 2010. Disponível em: <<http://creativecommons.org>>. Acesso em 18 maio 2010.

CUNHA, A. G. **Dicionário Etimológico Nova Fronteira da Língua Portuguesa**. 2. ed. Rio de Janeiro: Nova Fronteira, 1986. 839 p.

DAVENPORT, T. **Ecologia da informação**. Tradução: Bernadette Siqueira Abrão. São Paulo: Futura, 1998. 316 p.

DAVENPORT, T.; PRUSAK, L. **Conhecimento empresarial**. Tradução: Lenke Peres. Rio de Janeiro: Campus, 1999. 237 p.

DERTOZOS, M. **O que será**. Como o novo mundo da informação transformará nossas vidas. Tradução: Celso Nogueira. São Paulo: Companhia das Letras, 1997. 413 p.

DUQUE, C. G. **SiRILiCO**: Uma Proposta para um Sistema de Recuperação de Informação baseado em Teorias da Linguística Computacional e Ontologia, 2005. 118 p. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

EBAY INC. **eBay**. [S.l.]: eBay Inc., 2010. Disponível em: <<http://www.ebay.com>>. Acesso em 18 maio 2010.

ELLIS, D. **New horizons in Information Retrieval**. London: Library Association Publishing Ltd, 1990. 138 p.

FACEBOOK. **Facebook**. [S.l.]: Facebook, 2010. Disponível em: <<http://www.facebook.com>>. Acesso em 18 maio 2010.

FAVARO, T. O rival da Britannica. **Veja**, São Paulo, n. 2143, 16 dez. 2009. Disponível em: <<http://veja.abril.com.br/161209/rival-britannica-p-019.shtml>>. Acesso em 10 maio 2010.

FREIRE, G. Ciência da informação: temática, histórias e fundamentos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 1, p. 6-19, jan./abr. 2006. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/view/442>>. Acesso em 06 nov. 2009.

FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Normas de Apresentação Tabular**. 3. ed. Rio de Janeiro, 1993.

GILES, J. Internet encyclopaedias go head to head. **Nature**. London, n. 438, p. 900-901. dec. 2005.

GOOGLE. **Google**. [S.l.] Google, 2010. Disponível em: <<http://www.google.com>>. Acesso em: 18 maio 2010.

_____. **Orkut**. [S.l.]: Google, 2010. Disponível em: <<http://www.orkut.com>>. Acesso em: 18 maio 2010.

GNU PROJECT. **GNU Free Documentation License**. [S.l.] : GNU Project, 2010. Disponível em : <<http://www.gnu.org/licenses/gfdl.html>>. Acesso em: 18 maio 2010.

LAUDON, K. C.; LAUDON, J. P. **Sistemas de informação gerenciais**. 7. ed. São Paulo: Pearson Prentice Hall, 2007. 452 p.

LESK, M. **The Seven Ages of Information Retrieval**. Conference for the 50th anniversary of “As We May Think”, 12-14 October, 1995. MIT Press, 1995.

LÉVY, P. **A inteligência coletiva**. Por uma antropologia do ciberespaço. 5. ed. Tradução: Luiz Paulo Rouanet. São Paulo: Edições Loyola, 2007. 212 p.

_____. **Cibercultura**. Tradução: Carlos Irineu da Costa. São Paulo: Editora 34, 1999. 264 p.

LINKEDIN CORPORATION. **Linkedin**. [S.l.]: Linkedin Corporation. Disponível em: <<http://www.linkedin.com>>. Acesso em 18 maio 2010.

KLEINBERG, J. M. Authoritative Sources in a Hyperlinked Environment. **Journal of the ACM**. v. 46, n. 5, p. 604-632, Sept. 1999.

MANNING, C. D; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. Cambridge: Cambridge University Press, 2009. 544 p. Disponível em: <<http://nlp.stanford.edu/IR-book/information-retrieval-book.html>>. Acesso em 08 fev. 2010.

MARCONI, M. A; LAKATOS, E. M. **Fundamentos de metodologia científica**. 6. ed. São Paulo: Editora Atlas, 2007. 315 p.

MCAFEE, A. P. Enterprise 2.0: The Dawn of Emergent Collaboration. **MIT Sloan Management Review**, Cambridge, v. 47, n. 3, p. 21-28, Spr. 2006.

MORVILLE, P.; ROSENFELD, L. **Information Architecture for the World Wide Web**. 3 ed. Sebastopol: O'Reilly Media, Inc, 2006. 456 p.

MYSAPCE.COM. **MySpace**. [S.I.]: MySpace.com, 2010. Disponível em: <<http://www.myspace.com>>. Acesso em 18 maio 2010.

NEGROPONTE, N. **A vida digital**. Tradução: Sérgio Tellaroli. São Paulo: Companhia das Letras, 1995. 210 p.

NIELSEN, J. **Participation Inequality: Encouraging More Users to Contribute**. [S.I.]: useit.com, 2006. Disponível em: <http://www.useit.com/alertbox/participation_inequality.html>. Acesso em 3 ago. 2010.

O'REILLY, T. **What is Web 2.0**. Design Patterns and Business Models for the Next Generation of Software. [S.I.]: O'Reilly Media, Inc, 2005. Disponível em: <<http://oreilly.com/web2/archive/what-is-web-20.html>>. Acesso em 26 out. 2009

PROGRAMA SOCIEDADE DA INFORMAÇÃO. **sociedade da informação no Brasil**: livro verde. Brasília: Ministério da Ciência e Tecnologia, 2000. 195 p.

ROBREDO, J. **Documentação de Hoje e de Amanhã**. Uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas. 4. ed. Brasília: Edição de autor, 2005. 410 p.

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

<<http://www.eci.ufmg.br/pcionline/index.php/pci/article/view/235/0>>. Acesso em 06 nov. 2009.

SARACEVIC, T. Evaluation of evaluation in Information Retrieval. In: ANNUAL ACM CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 1995, Seattle. **Proceedings...** New York: ACM, 1995. p. 138-146.

SCHONS, C. H. A contribuição dos wikis como ferramentas de colaboração no suporte à gestão do conhecimento organizacional. **Informação & Sociedade: Estudos**, João Pessoa, v. 18, n. 2, p.79-91, maio/ago. 2008. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/1706>>. Acesso em 27 out. 2009.

TAPSCOTT, D. **Growing Up Digital**. New York: McGraw-Hill, 1998. 336 p.

_____. **Net Gen Education Challenge underway**. [S.l.]: DonTapscott.com, 2010a. Disponível em: <<http://dontapscott.com/2010/03/13/net-gen-education-challenge-underway/>>. Acesso em 28 abr. 2010.

_____. **Teaching kids democratic values**. [S.l.]: DonTapscott.com, 2010b. Disponível em: <<http://dontapscott.com/2010/03/04/teaching-kids-democratic-values/>>. Acesso em 28 abr. 2010.

_____. **Using social media to keep students engaged**. [S.l.]: DonTapscott.com, 2010c. Disponível em: <<http://dontapscott.com/2010/02/25/using-social-media-to-keep-students-engaged/>>. Acesso em 28 abr. 2010.

TAPSCOTT, D.; WILLIAMS, A. D. **Wikinomics**. How Mass Collaboration Changes Everything. New York: Portfolio, 2006. 324 p.

THE NIELSEN COMPANY. **Nielsen Insights**. Ratings and Rankings. Internet. New York: The Nielsen Company, 2010. Disponível em: <<http://en-us.nielsen.com/rankings/insights/rankings/internet>>. Acesso em 28 abr. 2010.

_____. **The Global Online Media Landscape**. Identifying Opportunities in a Challenging Market. New York: The Nielsen Company, 2009. Disponível em: <<http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/04/nielsen-online-global-lanscapefinal1.pdf>>. Acesso em 08 fev. 2010.

THOMSON REUTERS. **ISI Web of Knowledge Help**. [S.l.]: Thomson Reuters, 2009.

THOMSON REUTERS. **Web of Knowledge**. What it is. [S.I.]: Thomson Reuters, 2010. Disponível em: <<http://wokinfo.com/about/whatitis/>>. Acesso em 29 abr. 2010

TWITTER. **Twitter**. [S.I.]: Twitter, 2010. Disponível em: <<http://twitter.com>>. Acesso em 18 maio 2010.

VANDER WAL, T. **Folksonomy**. Folksonomy Coinage and Definition. [S.I.]: vanderwal.net, 2007. Disponível em: <<http://vanderwal.net/folksonomy.html>>. Acesso em 11 maio 2010.

VAN RIJSBERGEN, C. J. **Information Retrieval**. 2. ed. London: Butterworths, 1979. 208 p.

VIEGAS, W. **Fundamentos lógicos da metodologia científica**. 3. ed. Brasília: Universidade de Brasília, 2007. 241 p.

WIKIMEDIA FOUNDATION. **Wikipedia**. [S.I.]: Wikimedia Foundation, 2010. Disponível em: <<http://www.wikipedia.org>>. Acesso em 18 maio 2010.

WORLD WIDE WEB CONSORTIUM. **Architecture of the World Wide Web**, volume one. [S.I.], 2004a. Disponível em: <<http://www.w3.org/TR/2004/REC-webarch-20041215/>>. Acesso em 08 nov. 2009.

_____. **OWL Web Ontology Language**. W3C Recommendation 10 February 2004 [S.I.], 2004b. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em 26 jul. 2010.

_____. **OWL 2 Web Ontology Language**. Document Overview. [S.I.], 2009. Disponível em: <<http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>>. Acesso em 10 mar. 2010.

_____. **Pre-W3C Web and Internet Background**. [S.I.], 2004c. Disponível em: <<http://www.w3.org/2005/01/timelines/timeline-2500x998.png>>. Acesso em 28 abr. 2010.

_____. **RDF**. Resource Description Framework (RDF). [S.I.], 2004d. Disponível em: <<http://www.w3.org/RDF/>>. Acesso em 26 jul. 2010.

WORLD WIDE WEB CONSORTIUM. **Resource Description Framework (RDF):** Concepts and Abstract Syntax. [S.l.], 2004e. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>>. Acesso em 10 mar. 2010.

WURMAN, R. S. **Ansiedade de informação**. Tradução: Virgílio Freire. São Paulo: Cultura Editores Associados, 1991. 380 p.

WYLLYS, R. E. **Information Architecture**. Austin: The University of Texas at Austin. Graduate School of Library & Information Science, 2000. Disponível em: <<http://www.ischool.utexas.edu/~l38613dw/readings/InfoArchitecture.html>>. Acesso em 18 fev. 2010.

YAHOO! **Delicious**. Social bookmarking. [S.l.]: Yahoo!, 2010. Disponível em: <<http://delicious.com>>. Acesso em 18 maio 2010.

_____ **flickr**. [S.l.]: Yahoo!, 2010. Disponível em: <<http://www.flickr.com>>. Acesso em 18 maio 2010.

YOUTUBE, LLC. **YouTube**. [S.l.]: YouTube, LLC, 2010. Disponível em: <<http://www.youtube.com>>. Acesso em 18 maio 2010.

APÊNDICE A - Características do conjunto de artigos recuperados na base *Web of Science*

Autores

Com 2 publicações (3,45%¹⁶) apenas IGNAT, C., NORRIE, M., VIEGAS, F. e WATTENBERG, M. Com 1 publicação (1,72%), todos os demais.

Países de origem

Tabela 3 - Quantidade de artigos recuperados na base *Web of Science* de acordo com o país de publicação

País	Quantidade de registros	Percentual
Estados Unidos	24	41,38%
Canadá	4	6,90%
Inglaterra	4	6,90%
Espanha	4	6,90%
Taiwan	4	6,90%
Irlanda	3	5,17%
Finlândia	2	3,45%
Índia	2	3,45%
Israel	2	3,45%
Itália	2	3,45%
China	2	3,45%
Coréia do Sul	2	3,45%
Suíça	2	3,45%
Argentina	1	1,72%
Austrália	1	1,72%
Áustria	1	1,72%
Brasil	1	1,72%
Alemanha	1	1,72%
Japão	1	1,72%
Holanda	1	1,72%
Noruega	1	1,72%
Cingapura	1	1,72%
Sri Lanka	1	1,72%
Sem identificação	1	1,72%

Fonte: THOMSON REUTERS, 2009, tradução nossa

Notas: O somatório de registros por países (68) supera o total de registros recuperados (58) pois um mesmo registro pode ter sido publicado por instituições de mais de um país.

Instituições

Com 2 publicações (3,45%) apenas ETH, *Indiana University*, *National Cheng Kung University* e *Universidad Politécnica de Madrid*. Com 1 publicação (1,7241%) todos os demais.

¹⁶ Todos os dados percentuais são arredondados.

Tipos de documento

Tabela 4 - Quantidade de artigos recuperados na base *Web of Science* de acordo com o tipo de publicação

Tipo de documento	Quantidade de registros	Percentual
Total	58	-
Artigo	41	70,69%
Trabalho publicado em anais	17	29,31%

Fonte: THOMSON REUTERS, 2009, tradução nossa

Idioma

Tabela 5 - Quantidade de artigos recuperados na base *Web of Science* de acordo com o idioma

Idioma	Quantidade de registros	Percentual
Total	58	-
Inglês	57	98,28%
Espanhol	1	1,72%

Fonte: THOMSON REUTERS, 2009, tradução nossa

Ano de publicação

Tabela 6 - Quantidade de artigos recuperados na base *Web of Science* de acordo com o ano de publicação

Ano de publicação	Quantidade de registros	Percentual
Total	58	-
2010	1	1,72%
2009	26	44,83%
2008	9	15,52%
2007	7	12,07%
2006	13	22,41%
2005	2	3,45%

Fonte: THOMSON REUTERS, 2009, tradução nossa

Fonte de publicação

Tabela 7 - Quantidade de artigos recuperados na base *Web of Science* de acordo com as principais fontes de publicação

Fonte de publicação	Quantidade de registros	Percentual
Total	58	-
<i>Lecture Notes in Computer Science</i>	8	13,79%
<i>Computers & Education</i>	3	5,17%
<i>IEEE internet computing</i>	3	5,17%
<i>Concurrency and Computation - Practice & Experience</i>	2	3,45%
<i>Decision Support Systems</i>	2	3,45%
<i>Electronic Library</i>	2	3,45%
<i>Expert Systems with Applications</i>	2	3,45%
<i>IEEE Communications Magazine</i>	2	3,45%
<i>IEEE Transactions on Visualization and Computer Graphics</i>	2	3,45%
<i>International Journal of Cooperative Information Systems</i>	2	3,45%
<i>Library Hi Tech</i>	2	3,45%

Fonte: THOMSON REUTERS, 2009, tradução nossa

Nota: As demais fontes apresentaram apenas 1 registro cada (1,72%)

Áreas de assunto

Tabela 8 - Quantidade de artigos recuperados na base *Web of Science* de acordo com as áreas de assunto

Área de assunto	Quantidade de registros	Percentual
Ciência da Computação, Sistemas de Informação (<i>Computer Science, Information Systems</i>)	17	29,31%
Ciência da Computação, Teoria e Métodos (<i>Computer Science, Theory & Methods</i>)	15	25,86%
Ciência da Computação, Engenharia de Software (<i>Computer Science, Software Engineering</i>)	12	20,69%
Ciência da Informação e Biblioteconomia (<i>Information Science & Library Science</i>)	12	20,69%
Engenharia, Elétrica e Eletrônica (<i>Engineering, Electrical & Electronic</i>)	8	13,79%
Ciência da Computação, Inteligência Artificial (<i>Computer Science, Artificial Intelligence</i>)	6	10,34%
Telecomunicações (<i>Telecommunications</i>)	6	10,34%
Ciência da Computação, Aplicações Interdisciplinares (<i>Computer Science, Interdisciplinary Applications</i>)	5	8,62%
Pesquisa Operacional e Ciência de Gerenciamento (<i>Operations Research & Management Science</i>)	4	6,90%
Pesquisa em Educação e Educacional (<i>Education & Educational Research</i>)	3	5,17%
Ciência da Computação, Hardware e Arquitetura (<i>Computer Science, Hardware & Architecture</i>)	2	3,45%
Gerenciamento (<i>Management</i>)	2	3,45%
Ciências Sociais, Interdisciplinar (<i>Social Sciences, Interdisciplinary</i>)	2	3,45%
Agricultura, Multidisciplinar (<i>Agriculture, Multidisciplinary</i>)	1	1,72%
Sistemas de Automação e Controle (<i>Automation & Control Systems</i>)	1	1,72%
Negócios (<i>Business</i>)	1	1,72%
Comunicação (<i>Communication</i>)	1	1,72%
Engenharia, Multidisciplinar (<i>Engineering, Multidisciplinary</i>)	1	1,72%
Instrumentos e Instrumentação (<i>Instruments & Instrumentation</i>)	1	1,72%
Ciências Multidisciplinares (<i>Multidisciplinary Sciences</i>)	1	1,72%

Fonte: THOMSON REUTERS, 2009, tradução nossa

Notas: Um mesmo registro pode estar classificado em mais de uma área de assunto, por isso são apresentadas áreas de assunto que não foram selecionadas ou foram descartadas no refinamento inicial da pesquisa.

Percentual calculado a partir do número de registros por área de assunto dividido pelo total de registros recuperados (58) e não pelo somatório de registros por área de assunto (101). Desta forma, pelo mesmo motivo da nota acima, o somatório dos percentuais supera os 100%.

Citações

- Citação média por trabalho: 1,60
- Citações médias por ano: 18,60
- h-index¹⁷: 4

¹⁷ O h-index é um índice relativo da ferramenta *Web of Science* de publicações que apresentam mais citações e que possui um valor único para todo o conjunto de resultados. Com isso é possível analisar quais artigos possuem relativamente maior número de citações. No exemplo da pesquisa, o h-index indica que existem 4 trabalhos que possuem acima de 4 citações, sendo portanto de maior destaque (Ver tabela 2).

Tabela 9 - Quantidade de citações por ano dos artigos recuperados na base *Web of Science*

	Ano	Quantidade de citações	Percentual
	Total	93	-
	2010	11	11,83%
	2009	56	60,22%
	2008	19	20,43%
	2007	3	3,23%
	2006	4	4,30%

Fonte: THOMSON REUTERS, 2009, tradução nossa

APÊNDICE B - Análise individual dos artigos recuperados na base *Web of Science*

Observação: A correlação entre as características encontradas em cada artigo e o resumo apresentado no quadro 11 apresenta-se após cada característica com a legenda U, C ou S (respectivamente, para Usuário, Contextualização ou Sistema) acompanhada do número de ordem na respectiva categoria. Determinada característica pode apresentar relação com mais de uma categoria.

- 1) VIEGAS, F. et al. Many eyes: A site for visualization at internet scale. **IEEE Transactions on Visualization and Computer Graphics**, Los Alamitos, v. 13, n. 6, p. 1121-1128. Nov./Dec. 2007.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Formatos flexíveis (S9); e
- Tipos de visualizações pré-determinados, mas de livre contribuição pelos usuários (S9).

- 2) MONAHAN, T.; MCARDLE, G.; BERTOLOTTO, M. Virtual reality for collaborative e-learning. **Computers & Education**, Oxford, v. 50, n. 4, p. 1339-1353, May 2008.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Realidade virtual (U10);
- Aprendizagem de usuários por interação social (U2); e
- Acesso por dispositivos móveis (C1).

- 3) ZHA, X. F.; DU, H. Knowledge-intensive collaborative design modeling and support. Part II: System implementation and application. **Computers in Industry**, Amsterdam, v. 57, n. 1, p. 56-71, Jan. 2006.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Ontologias (Web semântica) (S15).

4) SCHALL, D.; TRUONG, H.; DUSTDAR, S. Unifying Human and Software Services in Web-Scale Collaborations. **IEEE Internet Computing**, Los Alamitos, v. 12, n. 3, p. 62-68, May-June 2008.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Acesso por dispositivos móveis (C1); e
- Estruturas de serviços e regras flexíveis (S9).

5) BERNOFF, J.; LI, C. Marketing - Harnessing the power of the oh-so-social web. **MIT Sloan Management Review**, Cambridge, v. 49, n. 3, Spr. 2008.

Tipo: Revisão de literatura

Características específicas do artigo de interesse para a RI:

- Conteúdo criado por usuários (C4);
- Redes sociais para divulgação de informação (U12, C12 e S18); e
- Estruturas de serviços flexíveis (S9).

6) KAJEWSKI, M. Emerging technologies changing our service delivery models. **Electronic Library**, West Yorkshire, v. 25, n. 4, p. 420-429, 2007.

Tipo: Revisão de literatura

Características específicas do artigo de interesse para a RI:

- Ferramentas colaborativas para divulgação de informação e *feedback* de usuários (C7).

7) HELD, M.; BLOCHINGER, W. Structured collaborative workflow design. **Future Generation Computer Systems**, Amsterdam, v. 25, n. 6, p. 638- 653, June 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Controle sobre visualização de versões intermediárias de produtos de informação (C3); e

- Personalização de informação para divulgação da informação aos usuários (C7).

8) LAI, L.; TURBAN, E. Groups Formation and Operations in the Web 2.0 Environment and Social Networks. **Group Decision and Negotiation**, Dordrecht, v. 17, n. 5, p. 387-402, Sept. 2008.

Tipo: Revisão de literatura

Características específicas do artigo de interesse para a RI:

- Recombinação de informação (C11);
- Análise de folksonomias (S1);
- Redes sociais (U12, C12 e S18);
- Confiança em usuários de um mesmo grupo (U4); e
- Controle de conteúdo pelos próprios usuários (U5).

9) NACAR, M. et al. VLab: collaborative Grid services and portals to support computational material science. **Concurrency and Computation: Practice and Experience**, Chichester, v. 19, n. 12, p. 1717-1728, Aug. 2007.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Estruturas flexíveis (S9); e
- Sistemas distintos de consulta e de processamento de consulta (C13 e S20).

10) LEHIKOINEN, J. et al. Meta-searches in peer-to-peer networks. **Personal and Ubiquitous Computing**, Guilford, v. 10, n. 6, p. 357-367, 2006.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Histórico de buscas do próprio usuário (S11);
- Índices compartilhados em máquinas em rede (P2P) (C9 e S12); e
- Cálculo de relevância baseado no compartilhamento automático de experiências passadas de buscas de usuários (S5).

11) COZZI, A. et al. Activity management as a Web service. **IBM Systems Journal**, Armonk, v. 45, n. 4, p. 695-712, Oct./Dec. 2006.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Ontologias (Web semântica) (S15);
- Informações das ontologias acessadas em ferramentas tradicionais, como *e-mail* (C10);
- Sincronização de buscas com mudanças na base de dados (S19); e
- Recomendações para colaboração em outras atividades ou com outros usuários (U11 e S17).

12) COOK, N.; ROBINSON, P.; SHRIVASTAVA, S. Design and implementation of Web services middleware to support fair non-repudiable interactions.

International Journal of Cooperative Information Systems, Singapore, v. 15, n. 4, p. 565-597, Dec. 2006.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de descrição de protocolo de comunicação de redes de dados.

13) IGNAT, C.; NORRIE, M. Flexible collaboration over XML documents. In: COOPERATIVE DESIGN, VISUALIZATION, AND ENGINEERING, 2006.

Proceedings... Berlin: Springer-Verlag, 2006, p. 267-274

Tipo: Revisão de literatura

Características específicas do artigo de interesse para a RI: Nenhuma

14) HWANG, Y.; KIM, D. Understanding Affective Commitment, Collectivist Culture, and Social Influence in Relation to Knowledge Sharing in Technology Mediated Learning. **IEEE Transactions on Professional Communication**, Piscataway, v. 50, n. 3, p. 232-248, Sept. 2007.

Obs.: Artigo com abordagem pertinente à colaboração na Web na análise de envolvimento social dos usuários, mas sem descrição de iniciativas de colaboração.

15) ABELS, E.; WHITE, M.; KIM, S. Developing Subject-related Web Sites

Collaboratively: The Virtual Business Information Center. **Journal of Academic Librarianship**, New York, v. 33, n. 1, p. 27-40, Jan. 2007.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de estudo de caso para criação de uma ferramenta através da colaboração de entidades envolvidas.

16) ANACLETO, J. et al. Using common sense to recognize cultural differences. In: **ADVANCES IN ARTIFICIAL INTELLIGENCE - IBERAMIA-SBIA**, 2006. **Proceedings...** Berlin: Springer-Verlag, 2006, p. 370-379.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Inteligência artificial para contextualização a partir de descrições em redes semânticas (S13).

17) CHIANG, T. et al. Friends night out - A working prototype of a blended lifestyle service enabled through IMS. **Bell Labs Technical Journal**, Hoboken, v. 10, n. 4, p. 17-23, Win. 2006.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Fontes de informação distintas utilizadas por uma aplicação/interface única (C8); e
- Contexto obtido a partir de dispositivos móveis (C5).

18) NISHANTHA, D. et al. A system for international telemedicine through integrated synchronous/asynchronous collaboration. **IEICE Transactions on Information and Systems**, Tokyo, v. E89D, n. 1, p. 271-280, Jan. 2006.

Obs.: Acesso restrito.

19) PUZIS, R. et al. Collaborative attack on Internet users' anonymity. **Internet Research**, West Yorkshire, v. 19, n. 1, p. 60-77, 2009.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de pesquisa em segurança da informação.

20) CABEZAS-CLAVIJO, A.; TORRES-SALINAS, D.; DELGADO-LOPEZ-COZAR, E. **Ciencia 2.0: catálogo de herramientas e implicaciones para la actividad**

investigadora. **Profesional de La Información**, Barcelona, v. 18, n. 1, p. 72-79, enero/feb. 2009.

Tipo: Revisão de literatura

Obs.: Indexado como “Science 2.0: Tools catalogue and consequences for scientific activity”

Características específicas do artigo de interesse para a RI:

- Recombinação de informação (C11); e
- Conteúdo criado por usuários (C4).

21) KUMAR, C. Performance evaluation for implementations of a network of proxy caches. **Decision Support Systems**, Amsterdam, v. 46, n. 2, p. 492-500, Jan. 2009.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de descrição de método/ferramenta para melhoria de performance em infra-estrutura de redes.

22) RINGBERG, H.; SOULE, A.; REXFORD, J. WebClass: Adding Rigor To Manual Labeling of Traffic Anomalies. **Computer Communication Review**, New York, v. 38, n. 1, p. 35-38, Jan. 2008.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de descrição de método/ferramenta para melhoria de performance em infra-estrutura de redes.

23) BERGAMASCHI, S.; SALA, A. Virtual Integration of Existing Web Databases for the Genotypic Selection of Cereal Cultivars. In: ON THE MOVE TO MEANINGFUL INTERNET SYSTEMS: COOPIS, DOA, GADA, AND ODBAS, PT 1, 2006. **Proceedings...** Berlin: Springer-Verlag, 2006, p. 909-926.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de descrição de método/ferramenta para fins agrários.

24) EVERITT, S. Opportunistic collaboration - Unlocking the archives of the Birmingham Institute of Art and Design. **Program-Electronic Library and Information Systems**, West Yorkshire, v. 39, n. 4, p. 312-323, 2005.

Tipo: Estudo de caso

Características específicas do artigo de interesse para a RI:

- Acesso estimulado pela colaboração (U1);
- Significância da informação reconhecida por colaboradores externos (U13); e
- Indexação colaborativa humana (S7).

25)BIER, E.; CARD, S.; BODNAR, J. Principles and Tools for Collaborative Entity-Based Intelligence Analysis. **IEEE Transactions on Visualization and Computer Graphics**, Los Alamitos, v. 16, n. 2, p. 178-191, Mar./Apr. 2010.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Sincronização de buscas com mudanças na base de dados (S19); e
- Compartilhamento e permissão de acesso a versões intermediárias de produtos de análise (C3).

26)LOPEZ, D. et al. Ciclope Robot: Web-Based System to Remote Program an Embedded Real-Time System. **IEEE Transactions on Industrial Electronics**, Piscataway, v. 56, n. 12, p. 4791-4797, Dec. 2009.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de descrição de método/ferramenta para interação entre usuários para a realização de atividades sem fins de RI.

27)FEENEY, K.; LEWIS, D.; O'SULLIVAN, D. Service-Oriented Policy Management for Web-Application Frameworks. **IEEE Internet Computing**, Los Alamitos, v. 13, n. 6, p. 39-47, Nov./Dec. 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Estruturas de políticas de serviços flexíveis (S9); e
- Autoridade específica para grupos/usuários sobre recursos (U3).

28)GREGG, D. Developing a collective intelligence application for special education. **Decision Support Systems**, Amsterdam, v. 47, n. 4, p. 455-465, Nov. 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Usuários com necessidades especiais (Público-alvo) (U14); e
- Estruturação e definição de objetivos para os usuários (U7).

29) ERICKSON, J. et al. Content-Centered Collaboration Spaces in the Cloud. **IEEE Internet Computing**, Los Alamitos, v. 13, n. 5, p. 34-42, Sept./Oct. 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- *Cloud Computing* (S6); e
- Estruturas de conteúdo flexíveis (S9).

30) Fu, F.; Wu, Y.; Ho, H. An investigation of cooperative pedagogic design for knowledge creation in Web-based learning. **Computers & Education**, Oxford, v. 53, n. 3, p. 550-562, Nov. 2009.

Tipo: Relatório de pesquisa

Características específicas do artigo de interesse para a RI:

- Interação e diversão para melhor desempenho dos usuários (U8).

31) HERTHER, N. Research evaluation and citation analysis: key issues and implications. **Electronic Library**, West Yorkshire, v. 27, n. 3, p. 361-375, 2009.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de análise sobre citações em artigos científicos, mas sem viés colaborativo.

32) JIN, Y.; PENG, J. Information portal development and practice at Shanghai Jiao Tong University Library. **Online Information Review**, West Yorkshire, v. 33, n. 3, p. 537-547, 2009.

Tipo: Estudo de caso

Características específicas do artigo de interesse para a RI:

- Fontes de informação distintas utilizadas por uma aplicação/interface única (C8);
- Recombinação de informação (C11); e
- Base compartilhada para usuários externos (C2).

33) SINGH, V. et al. Towards Environment-to-Environment (E2E) multimedia communication systems. **Multimedia Tools and Applications**, Dordrecht, v. 44, n. 3, p. 361-388, Sept. 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Contexto obtido a partir de dispositivos móveis (C5);
- Análise semântica a partir de domínio definido (S3); e
- Ontologias (Web semântica) (S15).

34) JAMES, D.; GARRETT, M.; KREVIT, L. Discovering discovery tools. Evaluating vendors and implementing Web 2.0 environments. **Library Hi Tech**, West Yorkshire, v. 27, n. 2, p. 268-276, 2009.

Tipo: Estudo de caso

Características específicas do artigo de interesse para a RI:

- Fontes de informação distintas utilizadas por uma aplicação/interface única (C8); e
- Ferramenta de descoberta de informação (*Discovery tool*) (S10).

35) WANG, Y. et al. Cultivating Personalized Museum Tours Online and On-Site. **Interdisciplinary Science Reviews**, Leeds, v. 34, n. 2-3, p. 139-153, Sept. 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Acesso por dispositivos móveis (C1).

36) BRAGGE, J.; MERISALO-RANTANEN, H. Engineering E-Collaboration Processes to Obtain Innovative End-User Feedback on Advanced Web-Based Information Systems. **Journal of the Association for Information Systems**, Atlanta, v. 10, n. 3, 196-220, 2009.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de descrição de método/ferramenta para colaboração na obtenção de produto em processo não contínuo.

37) DANG, Y. et al. Arizona Literature Mapper: An Integrated Approach to Monitor and Analyze Global Bioterrorism Research Literature. **Journal of the American Society for Information Science and Technology**, Hoboken, v. 60, n. 7, p. 1466-1485, July 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Redes sociais (Relacionamento) (U12, C12 e S18); e
- Mapeamento de conhecimento a partir de domínio específico (S3).

38) LU, C.; TSENG, V. A novel method for personalized music recommendation. **Expert Systems with Applications**, Oxford, v. 36, n. 6, p. 10035-10044, Aug. 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Histórico de buscas do próprio usuário (S11); e
- Redes sociais (Relacionamento e recomendações) (U12, C12 e S18).

39) HSWE, P. et al. The Web Archives Workbench (WAW) Tool Suite: Taking an Archival Approach to the Preservation of Web Content. **Library Trends**, Baltimore, v. 57, n. 3, p. 442-460, Win. 2009.

Tipo: Estudo de caso

Características específicas do artigo de interesse para a RI:

- Combinação de indexação automática e indexação colaborativa humana (S7); e
- Ferramenta de descoberta de informação (*Discovery tool*) (S10).

40) HUANG, A.; CHUANG, T. Social tagging, online communication, and Peircean semiotics: a conceptual framework. **Journal of Information Science**, London, v. 35, n. 3, p. 340-357, June 2009.

Tipo: Revisão de literatura

Características específicas do artigo de interesse para a RI:

- Análise de *tags* (inclusive por semântica, semiótica e ontologias para linguagem natural) (S2); e

- Recomendação de *tags* baseadas no vocabulário utilizado em sistemas diferentes por um mesmo usuário (S16).

41) LORETO, S. et al. Service Broker Architecture: Location Business Case and Mashups. **IEEE Communications Magazine**, Piscataway, v. 47, n. 4, p. 97-103, Apr. 2009.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de descrição de método/ferramenta de arquitetura de serviços de telecomunicações para a Web.

42) HOUSER, J. The VuFind implementation at Villanova University. **Library Hi Tech**, West Yorkshire, v. 27, n. 1, p. 93-105, 2009.

Tipo: Estudo de caso

Características específicas do artigo de interesse para a RI:

- Fontes de informação distintas utilizadas por uma aplicação/interface única (C8).

43) PIERCE, M. et al. Using Web 2.0 for Scientific Applications and Scientific Communities. **Concurrency and Computation – Practice & Experience**, West Chichester, v. 21, n. 5, p. 583-603, Apr. 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Fontes de informação distintas utilizadas por uma aplicação/interface única (C8);
- Acesso por dispositivos móveis (C1); e
- Análise de folksonomias (S1).

44) HUANG, Y.; LIU, C. Applying adaptive swarm intelligence technology with structuration in web-based collaborative learning. **Computers & Education**, Oxford, v. 52, n. 4, p. 789-799, May 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Estruturas de dados flexíveis (S9); e

- Auto-organização de conteúdo criado pelos usuários (S4).

45) YELMO, J.; TRAPERO, R.; DEL ALAMO, J. Identity Management and Web Services as Service Ecosystem Drivers in Converged Networks. **IEEE**

Communications Magazine, Piscataway, v. 47, n. 3, p. 174-180, Mar. 2009.

Obs.: Artigo com abordagem não pertinente à colaboração na Web. Trata-se de descrição de método/ferramenta de arquitetura de serviços de telecomunicações para a Web.

46) TUMMARELLO, G.; MORBIDONI, C. The DBin platform: A complete environment for SemanticWeb Communities. **Journal of Web Semantics**, Amsterdam, v. 6, n. 4, p. 257-265, Nov. 2008.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Modelos de confiabilidade e regras de filtragem de informação (U9);
- Ontologias acessam classificações similares (Web semântica) (S14); e
- Informações compartilhadas em máquinas em rede (P2P) (C9 e S12).

47) HEER, J.; VIEGAS, F.; WATTENBERG, M. Voyagers and Voyeurs: Supporting Asynchronous Collaborative Visualization. **Communications of the ACM**, New York, v. 52, n. 1, p. 87-97, Jan. 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Visualizações interativas de dados e de colaboração assíncrona (C6 e S8); e
- Diversas formas/modelos de colaboração (U6).

48) GAUDIOSO, E. et al. Supporting teachers in collaborative student modeling: A framework and an implementation. **Expert Systems with Applications**, Oxford, v. 36, n. 2, p. 2260-2265, Mar. 2009.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Estruturas de dados flexíveis, modeladas a partir da colaboração de usuários (S9).

49)GODOY, D.; AMANDI, A. Collaborative Web Search Based on User Interest Similarity. **International Journal of Cooperative Information Systems**, v. 17, n. 4, p. 495-521, Dec. 2008.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Ontologias (Web semântica) definidas a partir do perfil do usuário (S15).

50)KIM, B. et al. Collaborative Web Agent Based on Friend Network. **Applied Artificial Intelligence**, Philadelphia, v. 22, n. 4, p. 331-351, Apr. 2008

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Índices e informações de históricos de buscas compartilhados em máquinas em rede (P2P) (C9 e S12).

51)ADINARAYANA, J. et al. GramyaVikas: A distributed collaboration model for rural development planning. **Computers and Electronics in Agriculture**, Oxford, v. 62, n. 2, p. 128-140, July 2008.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Sistema de Gerenciamento de Conteúdo (CMS – *Content Management System*) (S21); e
- Fontes de informação distintas (inclusive geográficas) utilizadas por uma aplicação/interface única (C8).

52)CAMP, G.; ULIERU, M. InOrder: enhancing Google via stigmergic query refinement. **Computer Systems Science and Engineering**, Leicester, v. 22, n. 4, p. 217-237, July 2007.

Obs.: Acesso restrito.

53)RAZZAK, M.; HAMAM, H. A web-based user friendly simulator of optical fiber communication systems. **Annales Des Telecommunications**, Lausanne, v. 62, n. 5-6, p. 638-652, May-June 2007.

Obs.: Acesso restrito.

54)IGNAT, C.; NORRIE, M. Supporting customised collaboration over shared document repositories. In: ADVANCED INFORMATION SYSTEMS ENGINEERING, 2006. **Proceedings...** Berlin: Springer-Verlag, 2006, p. 190-204.

Tipo: Descrição de método/ferramenta

Características específicas do artigo de interesse para a RI:

- Estratégias de atualização de documentos compartilhados em repositórios para colaboração de usuários (C6 e S8).

55)HONG, Y.; JUN, W.; KWAK, B. Design and implementation of a mobile class Web site using intelligent user interface. In: TECHNOLOGIES FOR E-LEARNING AND DIGITAL ENTERTAINMENT, 2006. **Proceedings...** Berlin: Springer-Verlag, 2006. p. 716-725

Obs.: Acesso restrito.

56)FIAIDHI, J.; MOHAMMED, S. Collaborative virtual learning model for Web intelligence. In: COMPUTER SUPPORTED COOPERATIVE WORK IN DESIGN II, 2006. **Proceedings...** Berlin: Springer-Verlag, 2006. p. 563-572.

Obs.: Acesso restrito.

57)ENCHEVA, S.; TUMIN, S. Sharing protected web resources using distributed role-based modeling. In: FRONTIERS OF WWW RESEARCH AND DEVELOPMENT – APWEB, 2006. **Proceedings...** Berlin: Springer-Verlag, 2006. p. 532-543.

Obs.: Acesso restrito.

58)RASHKOVITS, R.; GAL, A. A cooperative model for wide area content delivery applications. In: ON THE MOVE TO MEANINGFUL INTERNET SYSTEMS 2005: COOPIS, DOA, AND ODBASE, PT 1, 2005. **Proceedings...** Berlin: Springer-Verlag, 2005. p. 402-419.

Obs.: Acesso restrito.