

XIV Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB 2013)
GT 1: Estudos Históricos e Epistemológicos da Ciência da Informação

Comunicação Oral

**A RECUPERAÇÃO DA INFORMAÇÃO COMO BASE FUNDAMENTAL DA
CIÊNCIA DA INFORMAÇÃO E SUA RELAÇÃO COM A CIÊNCIA DA
COMPUTAÇÃO**

Daniel Mendes Barbosa – UFMG
Marcello Peixoto Bax – UFMG

Resumo

Apresenta-se, a título de ilustração dos benefícios de uma epistemologia ampliada da CI, a pesquisa e desenvolvimento de um sistema de recuperação de informação. Partindo do exemplo do sistema desenvolvido, argumenta-se pelos benefícios de uma abordagem integral da ciência da informação que abarque tanto os usuários quanto os sistemas. Discute o estado atual em Minas Gerais das possibilidades de integração da ciência da informação e da computação, demonstrando ser esta integração necessária para a evolução da Ciência da Informação. Finalmente conclui-se a respeito da análise desenvolvida e sobre sua relevância dentro do atual cenário da Ciência da Informação.

Palavras-Chave: Sistema de Recuperação de Informação. Epistemologia da CI.

Abstract

This paper presents as an illustration of the benefits of an expanded epistemology of CI, the research and development of a system for information retrieval. Following the example of the developed system, it is argued for the benefits of a holistic approach to information science that encompasses both users and systems. Discusses the current state of Minas Gerais in the possibilities of integration of information science and computing, is demonstrating this integration required for the evolution of information science. Finally we conclude about the analysis undertaken and about its relevance in the current scenario of Information Science.

Keywords: Information Retrieval System. Epistemology of CI.

1 INTRODUÇÃO

Em todo campo científico e especialmente na Ciência da Informação (CI), por sua natureza interdisciplinar, é crucial discutir as fundamentações teóricas e disciplinares de referência. A pesquisa interdisciplinar é um modo de investigação científica que integra duas ou mais disciplinas para avançar na compreensão e resolver problemas cujas soluções estão fora do escopo de uma única disciplina. Entretanto, como adverte Dias (2002), “*o caráter interdisciplinar que se procura caracterizar na ciência da informação tem dado margem a certos exageros na definição das fronteiras do campo*”. Nesse sentido é importante ter-se claro que conteúdos formam as disciplinas básicas e que deveriam constituir o núcleo

principal da pesquisa realizada por um programa de pós-graduação em Ciência da Informação. Outros conteúdos podem ser relevantes, mas apenas em sentido complementar.

Ultimamente, inúmeras e variadas têm sido as reflexões epistemológicas publicadas na área a fim de se colocar em debate os fundamentos filosóficos que possam colaborar para uma maior consolidação desta ciência. Para uma brevíssima introdução, obviamente sem nenhuma pretensão de exaustividade, pode-se consultar ALVARENGA (2003); DIAS (2002); FLORIDI (2002); GOMES (2001); SARACEVIC (1999).

O presente trabalho tem por objetivo realizar uma análise epistêmica não exaustiva de um importante campo dentro da Ciência da Informação: a subárea de Recuperação da Informação (RI) com seus Sistemas de Informação e Recuperação da Informação (SIRIs). Sistemas de RI fornecem acesso à enorme gama de informações textuais disponíveis; suas técnicas são também aplicadas na recuperação multimídia e em tarefas como extração de informações, sumarização, disseminação seletiva, dentre outras.

Partindo de uma revisão bibliográfica recente, a análise irá focar-se em desvelar a importância da RI como um dos conhecimentos fundantes da área de CI. Em seguida propõe-se uma síntese das possíveis consequências dessa análise para a evolução da ciência da informação. Numa conclusão inicial limitada à subárea dos SIRIs, pode-se afirmar que não há como fazer avançar as pesquisas sem uma forte colaboração interdisciplinar com a grande área das ciências exatas e da terra, mais especificamente com as áreas da Informática ou Computação, como se queira chamar. Esta última caracterizada pelas disciplinas Ciência da Computação (CC), Sistemas de Informação, Engenharia e Matemática Computacional.

Sendo assim, dada a centralidade econômica e técnica dos SIRIs na Ciência da Informação (Dias 2002), não se pode exagerar os benefícios, para ambas, mas também para toda a Ciência de forma geral, de uma ainda maior aproximação epistêmica.

Segundo Floridi (2002), a Ciência da Informação como um todo não conseguiria se fundamentar apenas na epistemologia social, como se tem muitas vezes tentado fazer. Como alternativa, ele propõe que se considere a Filosofia da Informação como fundamento da CI, e esta última seria então entendida como uma filosofia da informação aplicada. Pelo menos em alguns campos mais quantitativos da CI, a construção desta filosofia da informação poderia servir justamente como fundamentação mais consistente para as pesquisas. Ela constituiria uma base para a integração, mais premente em algumas subáreas da Ciência da Informação e da Informática.

Desta forma, nas seções 2 a 5 apresenta-se o desenvolvimento do trabalho, trazendo as análises que foram realizadas. A seção 2 apresenta os conceitos de sistemas de informação e sistemas de recuperação de informação relevantes para o presente trabalho. A seção 3 apresenta, a título de ilustração dos benefícios de uma epistemologia ampliada da CI, a pesquisa e desenvolvimento de um sistema de recuperação de informação. Partindo do exemplo do sistema desenvolvido, na seção 4 argumenta-se pelos benefícios de uma abordagem integral da ciência da informação que abarque tanto os usuários quanto os sistemas. A seção 5 discute o estado atual em Minas Gerais das possibilidades de integração da ciência da informação e da computação, demonstrando ser esta integração profícua para a evolução da Ciência da Informação. Finalmente, a seção 6 traz uma conclusão a respeito de toda a análise desenvolvida e sobre sua relevância no atual cenário da Ciência da Informação.

2 SISTEMAS DE INFORMAÇÃO E DE RECUPERAÇÃO DE INFORMAÇÃO

Segundo Turban (2003), um sistema de informação é um conjunto de componentes relacionados entre si, que coleta, processa, armazena, analisa e dissemina informação, com um propósito específico para usuários interessados. O sistema de informação abrange entrada, processamento e saída, com o objetivo de realizar processos de comunicação. Já os sistemas de recuperação da informação são exemplos de mecanismos “especificamente planejados para possibilitar a recuperação da informação”, de acordo com Araújo (1995). Segundo Lancaster (1993), recuperação de informação é o processo de recuperar os documentos que tratam de determinado assunto em uma coleção. De acordo com Baeza e Ribeiro-Neto (2011), o objetivo chave de um sistema de recuperação da informação (SRI) é recuperar todos os itens que sejam relevantes a uma consulta do usuário (revocação), além de recuperar o mínimo possível de itens que não sejam relevantes (precisão). Segundo Souza (2006), um SRI deve buscar uma boa relação entre suas taxas de revocação e precisão, onde a revocação mede o sucesso do SRI em recuperar documentos pertinentes, e a precisão mede o sucesso do SRI em não recuperar documentos que não sejam relevantes. A Figura 1 resume o processo de recuperação da informação (RI): tem-se os termos dos documentos (*docs terms*) e os termos usados pelo usuário em sua consulta (*query terms*). O SRI irá então verificar quais termos da consulta são iguais aos termos dos documentos (*match*), recuperando os documentos que possuam estes termos numa ordem definida por algum esquema de classificação (*ranking*).

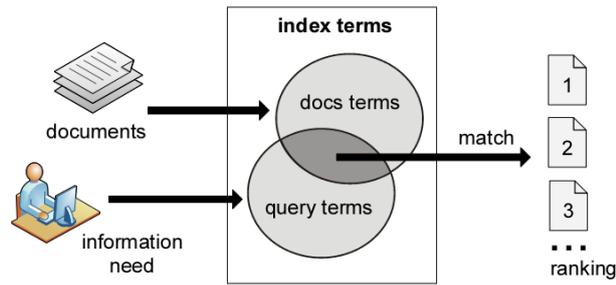


Figura 1: processo de recuperação da informação

Fonte: Baeza-Yates, Ribeiro-Neto (2011)

Os termos dos documentos devem ser previamente extraídos dos documentos da coleção que se deseja pesquisar. Os termos gerados neste processo constituirão um índice para a recuperação dos documentos em consultas posteriores. Segundo Lancaster (1986) um SRI precisa utilizar um vocabulário controlado, que seria uma delimitação dos termos presentes no índice, que representem os assuntos presentes nos documentos. No contexto da ciência da informação, Naves (2001) destaca a análise de assunto como uma das etapas mais importantes para a indexação dos documentos, onde o indexador dá início à indexação com subjetividade, conhecimento prévio, extraíndo conceitos expressos em linguagem natural, que representam o conteúdo de um documento. Em seguida é feita a tradução para termos utilizados como instrumentos de indexação, que são as chamadas linguagens de indexação. Esses termos são denominados frases de indexação por Frohmann (1990), os quais passam a ser chamados de descritores de assunto depois de traduzidos para uma linguagem de indexação. Segundo Souza (2006), o problema central na recuperação é justamente esta representação dos documentos através de operações no texto, indexação e criação do índice. A qualidade desta representação é que garante boas taxas de precisão e revocação. Quando a indexação é realizada manualmente (intelectualmente) por seres humanos, cabe a estes descobrir conceitos que sirvam de termo-índices para serem vasculhados durante as consultas, mas na indexação automática, existem dezenas de modelos para a correta ponderação do valor dos documentos, de acordo com uma explicitação de necessidade de informação, como booleano, vetorial, probabilístico etc. (SOUZA, 2006, p. 161). Uma estatística comum para suportar tal classificação é a frequência do termo, ou seja, o número de vezes que uma palavra aparece em um documento. Quanto maior a frequência de um termo em um documento, mais relevante será aquele documento para uma consulta que contenha aquele termo. Segundo Lopes (2002), os usuários de um SRI devem ser informados detalhadamente sobre o processo de busca.

Assim eles participarão mais ativamente, compreendendo sua complexidade, as limitações das bases e as interações efetuadas para o alcance dos resultados da busca.

No entanto, ao invés de continuar seguindo este caminho de perseguir a evolução de sistemas com base puramente na imitação das heurísticas intelectuais humanas, buscam-se estratégias diferenciadas, possíveis hoje graças aos ambientes com grande poder computacional (SOUZA, 2006, p. 168). Algumas destas estratégias são apresentadas na Figura 2, sendo as principais baseadas na exploração das informações semânticas intrínsecas aos documentos, nas novas possibilidades de marcação semântica, desenvolvimento de estratégias de apresentação da informação, permitindo aos usuários um refinamento dos resultados, perfis personalizados de utilização e nos mapas demográficos de relacionamentos

das redes sociais.

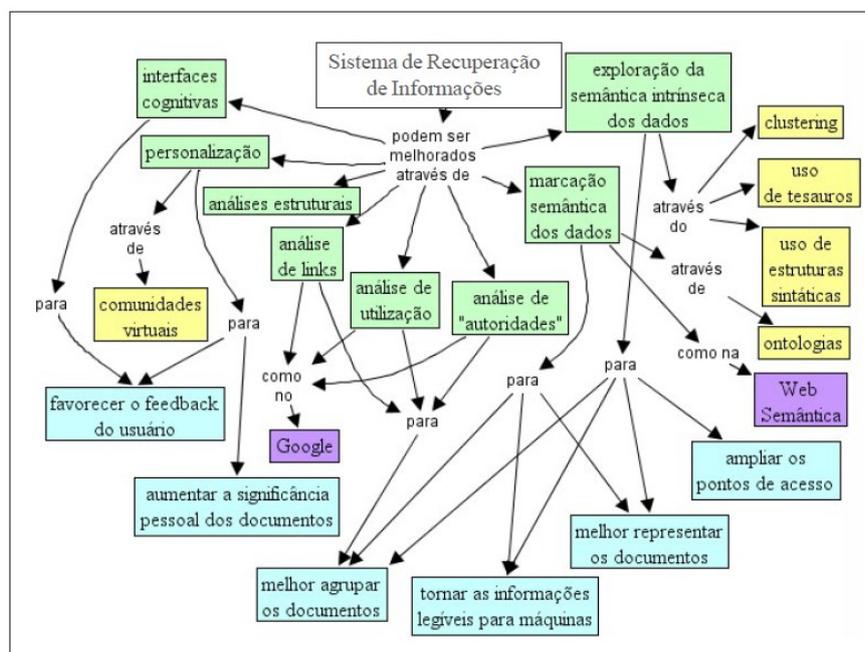


Figura 2: Estratégias alternativas para melhoria dos SRI.

Fonte: SOUZA (2006)

Estas estratégias mostram a importância da computação e da criação de novos algoritmos que alcancem bons resultados em cada uma delas. Esta junção dos conceitos da Ciência da Informação, com os recursos e técnicas provenientes da Ciência da Computação é inevitável. Com a disseminação de uso da *web*, os SRI ganharam uma atenção especial. Agora a busca deve ser feita em servidores conectados à Internet, que armazenam arquivos nos mais

diferentes formatos, mais ou menos estruturados. Isso fez e faz com que novas pesquisas tenham que ser desenvolvidas para se combinar os antigos conceitos com os novos.

Souza (2006) destaca que o mecanismo de busca do Google foi o primeiro a desenvolver um algoritmo, chamado *Pagerank*, baseado na análise de *links*. Este algoritmo calcula o “valor” de um *site* levando-se em conta as páginas que apontam para ele. Esta ideia inicial tem sido constantemente ampliada, incluindo ponderações para estes *links*, de acordo com as temáticas das páginas. Ou seja, uma página que referencia outra e as duas tratam do mesmo assunto terá um peso maior no cálculo do que uma outra página referenciadora que trata de outros assuntos.

Ao se fazer buscas nos atuais buscadores da Internet pode-se observar também a influência de outros aspectos nos resultados apresentados, bem como na ordem em que eles são apresentados: se o usuário está autenticado, a partir de onde ele está fazendo as buscas, navegações anteriores armazenadas pelo seu navegador de Internet, etc. Todas estas técnicas são feitas para atender de forma cada vez mais eficaz as necessidades informacionais do usuário, e são possíveis graças aos recursos computacionais envolvidos. A facilidade de acesso a recursos computacionais é tamanha hoje em dia que um estudante de CI que também tenha conhecimentos em Computação pode implementar um SRI, trabalhando na prática com os conceitos estudados. Um exemplo interessante é quando se deseja construir um SRI para um contexto específico. Segundo ELMASRI e NAVATHE (2011), sistemas de recuperação da informação podem ser ajustados a tipos de dados específicos. Ou seja, uma coleção específica de documentos de um determinado domínio pode ser explorada muito mais eficientemente por um sistema especializado do que por um sistema de recuperação da informação genérico.

3 PESQUISA E DESENVOLVIMENTO DE UM SRI

Como exemplo de desenvolvimento de um SRI, foi especificado e concebido um sistema de recuperação de planos de ensino de disciplinas (Barbosa et al. 2013). O objetivo foi dar suporte à função de coordenação de cursos de graduação, que necessita de uma boa gestão da informação como auxílio à tomada de decisão (como qualquer outra função de gerência em uma organização). O sistema foi desenvolvido a partir dos conceitos básicos de RI discutidos anteriormente, recuperando documentos no formato PDF contendo planos de ensino de disciplinas de cursos de graduação em um formato pré-definido. Sendo

especializado na recuperação de planos de ensino, o sistema possui uma técnica para se calcular a estatística de frequência do termo de forma específica: o peso da frequência. É atribuído um peso diferente a uma ocorrência de determinado termo de acordo com a seção em que o termo aparece. O sistema atribui o peso 1 para uma ocorrência no conteúdo programático, peso 2 para uma ocorrência nos objetivos e peso 3 para uma ocorrência na ementa, devido às respectivas relevâncias de tais seções para um plano de ensino. P.ex., se em determinado documento a palavra “informação” aparece 1 vez na ementa, 1 vez nos objetivos e 2 vezes no conteúdo programático, sua frequência neste documento será igual a 7 ($1 \times 3 + 1 \times 2 + 2 \times 1$).

A interface de busca é bem simples: apenas um campo texto, no qual o usuário digita as palavras-chaves que deseja pesquisar, separadas pelos operadores lógicos AND e OR. Os documentos são exibidos, ordenados pela soma das frequências de cada termo da consulta em cada documento. Esta soma representa o grau de relevância do documento para a consulta realizada.

É fácil perceber que a implementação deste sistema foi feita de maneira bem simples, sem sofisticação. No entanto, foi verificado que seu desempenho atende perfeitamente ao seu contexto específico.

4 USUÁRIO E SISTEMA: INTEGRAÇÃO INEVITÁVEL

A solução criada e apresentada na seção anterior tem o mérito de ser simples, e mesmo assim resolver o problema levantado. Mas isso só foi possível devido a se tratar de um contexto específico, que limitou os tipos de documentos a serem pesquisados, e conseqüentemente facilitou a indexação e a criação das estatísticas. Porém, mesmo sendo simples, um estudante com formação apenas na CI atual dificilmente seria capaz de implementar tal sistema. Soluções como a apresentada testemunham da necessária integração entre Informação e Computação. Uma realidade que precisa se concretizar na CI. No âmbito de Minas Gerais uma tal junção apenas ainda não se materializou (na UFMG) porque estas áreas permanecem formalmente separadas pela rígida estrutura da universidade, separada em diferentes departamentos, com corpos docentes atuando física e culturalmente isolados. Porém, a natureza fortemente interdisciplinar da CI pressiona e demanda uma maior integração disciplinar na academia. Integração que ocorrerá mais cedo ou mais tarde, pelo menos na UFMG. Barbosa (2008), p. ex. ressalta que há uma tendência de escolas e

departamentos de ciência da informação e/ou biblioteconomia se associarem a departamentos de administração, computação, comunicação ou educação. Destaca o autor que “o perfil do corpo docente desses programas também tem se tornado multidisciplinar, o que contribui para aumentar as interfaces entre as disciplinas”.

Outra questão muito estudada na CI é o comportamento de usuários de SRI; o seu comportamento informacional. Na Computação há também estudos relacionados, porém em relação ao uso de uma interface de um sistema de informação, inclusive em uma subárea denominada IHC (interação humano-computador). No entanto estas duas abordagens de análise são claramente complementares. Suas conclusões são mutuamente relevantes, apesar de trabalharem separadamente.

Chua e Yang (2008) fizeram um estudo bibliométrico de artigos coletados no JASIST entre os anos de 1988 e 2007, analisando-os ao longo de duas décadas. Todos os artigos coletados foram classificados de acordo com a formação de seu autor principal: núcleo da CI (acesso e recuperação da informação, organização do conhecimento, comunicação científica, economia da informação e busca *online*), gestão de sistemas de informação e tecnologia da informação e outras. Os resultados estão na Tabela 1.

Formação dos autores principais	1988-1997	1998-2007
Núcleo da Ciência da Informação	38 (61,3%)	73 (47,7%)
Gestão de Sistemas de Informação	4 (6,5%)	13 (8,5%)
Tecnologia da Informação	10 (16,1%)	38 (24,8%)
Outras	10 (16,1%)	29 (19%)
Total	62 (100%)	153 (100%)

Tabela 1: distribuição das áreas de formação dos autores principais nas duas décadas pesquisadas (adaptado de CHUA e YANG, 2008).

Como se vê na Tabela 1, na segunda década pesquisada houve um declínio de 13,6% dos pesquisadores originários do núcleo tradicional da CI (de 61,3% para 47,7%). Enquanto isso houve aumento considerável da presença dos autores pesquisadores ligados à gestão e principalmente daqueles ligados à tecnologia da Informação. Esses últimos aumentando em 8,7% (de 16,1% para 24,8%).

Outro estudo bibliométrico importante é o de Guo e Sheffield (2008), onde foram analisados os artigos de 5 revistas de gestão no período entre 2000 e 2004. Nos artigos sobre gestão do conhecimento (outro núcleo importante da CI), há uma forte tendência de

crescimento de trabalhos empíricos após 2002, sempre dependendo também da Computação, com o uso de intenso de *softwares* de pacotes estatísticos para análises complexas de dados, como análise fatoriais multivariadas, p.ex.

5 INFORMAÇÃO E COMPUTAÇÃO: INTEGRAÇÃO NECESSÁRIA

Saracevic (1999) reconhece a importância da Computação e da Tecnologia para a CI, afirmando que “a Ciência da Informação está inexoravelmente ligada à Tecnologia da Informação”. Ele ainda complementa, ressaltando que a CI é interdisciplinar por natureza, no entanto, as suas relações com diversas disciplinas estão mudando, e que esta evolução interdisciplinar está longe de terminar. Afirma ainda que o conceito original e importante para o surgimento da CI apareceu na década de 1950 e foi o conceito de RI. Ele permitiu a automação do processamento de textos por meio da lógica formal. Tratando da RI como base da CI, e suas relações com a Computação, Saracevic (1999) traz mais algumas constatações interessantes, aqui resumidas:

- Calvin Mooers cunhou o termo “recuperação da informação” (*information retrieval*) na década de 1950 e seguiu uma abordagem com ênfase na construção de sistemas; abordagem esta que foi a única durante décadas (centrada no sistema).
- No final da década de 1970 uma nova linha de pensamento apareceu, centrada na cognição, usuários, uso, interatividade com os sistemas, situações, contexto, e não somente nos sistemas de RI isolados (centrada no usuário).
- Em algum momento as interações entre as duas linhas de pesquisa interromperam-se e apareceu a questão de como fazer as duas abordagens trabalharem juntas.
- A Computação foca na manipulação de símbolos, enquanto a CI foca em conteúdos e para isso a manipulação de símbolos é instrumento indispensável.

Como se pode perceber, as duas ciências (informação e computação) são complementares, mas suas pesquisas ainda caminham separadas, com abordagens distintas. Isso se reflete não só na pesquisa, como também na educação. Saracevic (1999) também defende que a CI não recebe a merecida atenção na academia, sendo muitas vezes uma especialização da biblioteconomia, baseada no modelo de Shera, que falha ao não ensinar os formalismos lógico-matemáticos fundantes. O resultado é a formação incompleta dos profissionais, exclusivamente centrados no usuário final, que aliás os procuram cada vez menos numa sociedade de informação ubíqua, de acesso à informação cada vez mais

desintermediado. Já o modelo de Salton foi usado como especialização na Ciência da Computação, resultando em profissionais com a abordagem exclusivamente centrada no sistema. As duas abordagens são hoje completamente independentes e não são conectadas. Embora em cada modelo haja um aumento do conhecimento do outro, não há uma integração educacional das duas abordagens. Com efeito, “o alto nível de especialização das disciplinas produz uma disjunção que torna o conhecimento especializado incapaz de proporcionar uma compreensão multidimensional do fenômeno estudado” (GOMES, 2001).

No Brasil a situação não é diferente. Numa rápida pesquisa pelos *sites* dos programas de pós-graduação em CI, observa-se que em apenas três deles existem linhas de pesquisa cujos nomes possuem a palavra “tecnologia”. São os programas das seguintes universidades: UFBA (Universidade Federal da Bahia), UFRGS (Universidade Federal do Rio Grande do Sul) e UNESP (Universidade Estadual Paulista). Já os termos “lógica” e “computação” são inexistentes. No entanto, a partir das descrições dos assuntos pesquisados pelas linhas, tem-se o claro entendimento de que todas são baseadas exclusivamente na abordagem centrada no usuário, com descrições como “compreensão de fenômenos sociais mediados pelas tecnologias da informação e comunicação (TIC)” ou ainda “investiga os processos informacionais e comunicacionais, considerando as relações entre as tecnologias da informação e da comunicação e os diferentes campos do conhecimento científico e técnico, seus padrões, demandas e uso de informação”. Além disso, tais programas são oferecidos por unidades acadêmicas totalmente desvinculadas da ciência da computação e da área de exatas, sendo normalmente oferecidos por uma escola ou departamento próprios, ou ainda dentro de uma escola de comunicação.

Saracevic (1999) conclui afirmando que “a maior recompensa para a CI virá se e quando ela integrar com sucesso as pesquisas e aplicações centradas no sistema e no usuário.” De fato, principalmente com a grande evolução tecnológica recente, fazendo com que o ambiente digital se torne o mais utilizado para armazenar, representar e recuperar informações, esta junção não só é necessária como é urgente. Vários autores da CI compartilham desta opinião, dentre os quais Alvarenga (2003):

A ciência da informação não está iniciando hoje a sua jornada. Ela já caminha a passos largos, há muito tempo. Urge que os conhecimentos já sistematizados nesse campo de pesquisa, ao longo de seu processo evolutivo que remonta aos arquivos e à biblioteconomia tradicionais, sejam utilizados na resolução de problemas que se colocam no âmbito da **presente revolução dos meios de representação e comunicação de informações**. (ALVARENGA, 2003, p. 39, grifo nosso.)

Mesmo diante de tantas evidências da necessária integração das duas áreas, ainda há trabalhos que insistem por reduzir a atuação da CI à atividade de "acesso à informação", como em Dias (2002), p.ex.. Essa visão compreende o profissional da informação como um (mero) intermediário entre o usuário e a informação. Segundo o autor, o acesso a informação publicada seria próprio da biblioteconomia, o acesso à Informação especializada seria próprio da ciência da informação e o acesso à informação não publicada seria próprio da arquivologia. Esse entendimento posiciona a CI como área irmã às áreas da biblioteconomia e da arquivologia. Contudo, como atualmente quase todo o acesso à informação é automatizado por sistemas de informação, tal visão implica em reduzir o profissional da informação a um mero alimentador de sistemas (catálogos) e simples usuário de SRIs. A situação se agrava estando este profissional desprovido de formação mínima nos campos da matemática e da lógica. Ora, sabe-se que a crescente humanização da tecnologia implica na desintermediação da relação usuário e informação, um fenômeno cada vez mais expressivo na sociedade da informação ubíqua.

Finalmente, deve-se atentar ainda para o fato de que esta visão restrita da CI mina a necessária integração entre a formação de graduação e pós-graduação. Não é possível superestimar a importância dessa integração para a consolidação de qualquer ciência ou campo profissional. Em nível de pós-graduação a CI acaba por atrair discentes melhor preparados de cursos de graduação de áreas afins, dificultando assim a continuidade pedagógica tão necessária para o fortalecimento do conhecimento na área.

6 CONCLUSÃO

A partir das análises realizadas pode-se concluir que a subárea de RI é um dos pilares da Ciência da Informação (possivelmente o maior e mais importante), que sustentam a razão de ser desta área na academia. De fato, nota-se que todas as principais subáreas da CI estão de alguma forma relacionadas com a RI: indexação, representação, classificação, gestão da informação e do conhecimento, comunicação científica, controle bibliográfico, memória, acesso dos usuários, necessidades dos usuários e inclusão digital etc. No fim, todas estas atividades são realizadas com o objetivo de se recuperar a informação, previamente representada, indexada e classificada. Esta centralidade da RI na CI pode ser observada também nos GT's (grupos de trabalho) da ANCIB. Com uma breve análise, o fato pode ser percebido, uma vez que a RI é a razão pela qual se organiza e representa o conhecimento (GT

2 – Organização e representação do conhecimento). Os SI são considerados como mediadores (GT 3 – Mediação, circulação e apropriação da informação) entre a informação e o usuário. Também não se faz a Gestão da Informação e do Conhecimento (GT 4 – Gestão da informação e do conhecimento nas organizações) sem os SIRIs. Toda política e economia da informação (GT 5 – Política e economia da informação) tem por finalidade facilitar e promover a recuperação da informação, sendo um exemplo disso a lei de acesso a informação para os órgãos públicos. Os SIRIs estão também na base da produção e comunicação da informação em todos os contextos, devido ao amplo de tais sistemas nos mais diversos contextos (GT 7 – Produção e comunicação da informação em ciência, tecnologia e inovação).

Araújo (2010) analisa estes GT's com base nos três paradigmas de Capurro, e chega à conclusão de que os modelos físico, cognitivo e social da informação são complementares em todos os GT's. Segundo Araújo (2010):

Em lugar de se vislumbrar formas de pensamento que se opõem (o modelo cognitivo “contra” o físico, o social “contra” o cognitivo), o que a análise da presença dos três nas várias subáreas permite evidenciar é, antes de tudo, sua complementaridade. Um modelo não surge para substituir o outro, mas sim para completá-lo, buscando analisar justamente aquilo que o modelo precedente não dá conta, ou deixa de fora. (ARAÚJO, 2010, p.102)

Consequentemente pode-se concluir que eles são complementares entre si e também à RI, que pode ser pensada como a base, pelo menos instrumental, de todos os GT's. Ao se trabalhar com os três modelos para se resolver os atuais problemas de pesquisa, necessariamente deve-se recorrer também à Ciência da Computação. O modelo físico, principalmente, requer a análise e o uso de recursos computacionais e seus SIRIs, uma vez que não faz mais sentido existir um SRI manual, sem o uso da tecnologia.

No entanto, ao longo dos anos, a epistemologia da CI relativa à subárea de RI foi caminhando para abordagens científicas centradas apenas no usuário, deixando os próprios SIRIs, seus métodos e técnicas, conceitos, especificações e implementações em segundo plano, o que também pode ser entendido como uma prevalência do modelo cognitivo. Ao final da década de 1980, dada a relevância que foi tomando a informação distribuída em redes, a Ciência da Computação logo tratou de se apropriar de tal abordagem, centrada epistemologicamente na engenharia do sistema. Abordagens estas originalmente abarcadas pela Ciência da Informação, segundo nos lembra Saracevic (1999).

Com vistas a ilustrar quão profícua pode ser uma abordagem integrada, um exemplo utilizado neste trabalho foi a implementação de um SRI específico, com a criação de uma métrica de relevância, para uma busca em uma coleção específica: planos de ensino de disciplinas de cursos de graduação. O sistema criado obteve resultados interessantes, na eficácia e eficiência, sendo uma forma concreta de se consolidar conceitos através da prática. Mas sua implementação só foi possível devido à formação multidisciplinar do autor, que possui graduação em Ciência da Computação e mestrado em Ciência da Informação.

Com isso chega-se a um impasse: se a RI é um pilar para a CI e com a evolução das TICs ela tem sido mais bem explorada pela Computação, pergunta-se como a CI poderá sustentar seu futuro dentro da academia, se justificando? Num cenário futuro pessimista a CI poderá se encontrar como que cada vez mais esvaziada e pouco relevante. Cercada que está por todos os lados de concorrentes mais bem consolidados. Para abordar os aspectos sociais da informação a CI já sofre a concorrência dos cursos de Sistemas de Informação (cursos mais estabelecidos em departamentos de Computação). À medida em que a área da Administração percebe cada vez mais as oportunidades oferecidas pela área da gestão da informação e do conhecimento, esta vai se apropriando deste filão, também original da CI.

As contribuições da CI são inegáveis, mas precisarão ser melhor aproveitadas e valorizadas numa abordagem conjunta, complementada pelo ensino de aspectos lógico-matemáticos fundantes. Assim ela poderia estar centrada ao mesmo tempo no sistema e no usuário. Somente a interdisciplinaridade não será suficiente para resolver este impasse. É necessário se caminhar em direção a uma mais expressiva integração ou mesmo junção entre estas áreas na academia, aproximando mais os pesquisadores de ambas as áreas. Nota-se que naturalmente, e apesar das resistências que insistem e reduzir as possibilidades dos profissionais, esta aproximação já vem acontecendo aos poucos nos últimos anos, por ser inevitável. No contexto acadêmico no qual se produz uma ciência cada vez mais super-especializada, vale lembrar que muitas vezes os próprios cursos de graduação de ciência da computação, e talvez principalmente de sistemas de informação reconhecem suas bases epistêmicas nos institutos de ciências exatas não sem dificuldades. Com efeito, a penetração social direta dos artefatos de tecnologia da informação na sociedade é tal, que hoje em dia essas ciências poderiam muito bem ser percebidas com fortes interseções com as ciências sociais aplicadas.

No Brasil este cenário está aparentemente longe de se concretizar. Os programas de pós-graduação em CI encontram-se em sua maioria totalmente separados da Computação, e todos com a abordagem centrada no usuário, de Shera. Há inclusive classificações das áreas de conhecimento muito atrasadas neste sentido, como a do próprio CNPq, discutida por Dias (2002). Nesta classificação, a RI aparece apenas como “técnicas de recuperação da informação” e ainda pertencendo à biblioteconomia, e esta sim pertencendo à CI. Classificações como esta dificultam uma visão mais ampla e integrada da CI, dentro de seu próprio campo e suas relações com os demais.

Finalmente, como relacionar o debate epistêmico atual com o que defendemos a respeito do importante alinhamento da CI e da Computação? Nesse sentido concordamos com L. Floridi que neste novo contexto a nascente filosofia da informação fornecerá a base sob a qual se fundamentaria a CI. Por ser a ciência aplicada que lida diretamente com a informação, a CI necessita da Computação, que também se fundamenta na mesma filosofia. Para melhor pontuar a questão, vejamos um exemplo de visão com a qual não nos alinhamos, citando o filósofo Rendón Rojas:

Devido ao surgimento de **novas disciplinas** científicas, tais como Ciência da Informação e Documentação, que estudam os fenômenos documentários informativos, enquanto a permanência das disciplinas de Biblioteconomia e Arquivologia têm o mesmo objeto de estudo, há um crescente interesse nas investigações epistemológicas que tentam explicar a relação de subordinação, ou harmonização entre essas disciplinas. (Rojas, 2011. Trad. livre e grifo nosso.)

Tal pensamento revela a opção por se referir a "novas disciplinas" e não a "novos contextos", como faz Floridi. Pode-se presumir que Rojas não concorda com o fato de que o surgimento da Internet tenha alterado as coisas de maneira substancial, originando um novo contexto sócio-técnico bastante diferente do anterior. Cabe questionar, entretanto, se são mesmo "novas disciplinas" (como quer Rojas) ou se, de fato, são "novos contextos". Floridi (2004) defende posição inversa: o surgimento da sociedade da informação e da infosfera (ambiente semântico que engloba milhões de pessoas) constituem um contexto radicalmente novo (e revolucionário) e pede um pensamento novo. Nesse contexto novo a filosofia da informação seria o campo filosófico preocupado com a investigação crítica da natureza conceitual e os princípios básicos da informação, incluindo a sua dinâmica, utilização e ciências. Segundo Floridi, na filosofia da informação, o foco sai da memória e das linguagens de organização do conhecimento (meros instrumentos de gestão da infosfera) e passa para concentrar-se na natureza da essência e produção, a própria informação. Para o mesmo autor,

a Informação surge como um conceito tão fundamental e filosoficamente importante quanto 'ser', 'conhecimento', 'vida', 'inteligência', 'significado' ou 'bem e mal'. A ontologia do ser humano, que para Aristóteles é um ser político e para Marx é um ser econômico/social, se volta, para a filosofia da informação, a considerá-lo um ser informacional. Desse modo, trata-se de um novo contexto sócio-técnico que origina profundas implicações antropológicas.

Esse pensamento contrasta frontalmente com a visão de Rojas, que considera que trata-se apenas de novas disciplinas tendo que ser harmonizadas com novas disciplinas. Rojas chega a chamar a Documentação de nova disciplina, o que é duvidável, pois a Documentação é uma disciplina muito mais antiga, mesmo já estabelecida no Séc. XIX. Se o termo "ciência da informação" é de origem recente, reflexões sobre a circulação de documentos são muito mais antigas. Segundo WIKIPEDIA (2013), Gabriel Naudé cria a biblioteconomia no século XVII ao escrever "*l'Advis pour dresser une bibliothèque*". Paul Otlet separa a ciência da informação da biblioteconomia e publica seu Tratado da Documentação em 1934. Em 1917, ele define a Documentação como se constituindo de todos os meios próprios para transmitir, comunicar e transmitir informações. Nota-se então claramente que não se trata de um problema de "novas disciplinas (CI e Documentação) *versus* disciplinas mais antigas (biblioteconomia)". As questões epistemológicas da área não estão sendo bem colocadas por esse pensamento, pois ele revela um desconhecimento e uma conseqüente deflação da importância do fenômeno Internet, que de forma alguma representaria uma "revolução", que altera mais profundamente a sociedade.

A área de CI/Biblioteconomia tem debatido sua fundamentação teórica e status acadêmico desde os anos 1930 (com o curso de pós-graduação Chicago *School Library*). O problema, segundo o Floridi é que o fenômeno da informação hoje teria algo de fundamentalmente novo e muito mais impactante. Assim, a área teria adotado cedo demais fundamentos filosóficos da Epistemologia Social que não dão conta de fundamentar esse "novo fenômeno" informacional. Nas palavras de Floridi:

Os pesquisadores de CI/B foram atraídos por uma variedade de filosofias pré-estabelecidas "amigas", em vez de lutar por seu lugar próprio no campo filosófico. Formando 'guetos intelectuais' (Floridi, 2004).

Acreditamos nós que a web (rede + protocolo + linguagens de marcação) estabeleceu um novo contexto no qual disciplinas nem-tão-novas (como a CI e Documentação) ganham relevância. E outras, para as quais a adaptação ao novo contexto é mais penosa (a biblioteca precisa antes de tudo tratar livros físicos) sofrem relativa falta de atratividade, numa difícil

fase de transição, como tantas outras anteriores ocorreram na história das técnicas. Dessa vez entretanto, segundo Stiegler (*apud* IVESON 2012), tem-se o fenômeno da “gramatização” radical da noética, possibilitada pelo “digital”.

Nesse sentido, afirma Floridi que a crise na CI/B seria muito mais devida à influência da abordagem pragmática que levou à falta de fundamentos teóricos. Quando se buscou os fundamentos, esses vieram exclusivamente da epistemologia social. Segundo Zwadlo (*apud* Floridi 2002), para os bibliotecários:

obter uma filosofia é algo parecido com o empréstimo de um livro de nossas bibliotecas. Mas, como os livros emprestados, as filosofias emprestadas não pertencem a nós, parecem precisar sempre ser renovado o empréstimo, e acabamos por devolvê-las e pegar outras emprestadas.

A nós parece preferível à CI/B aproveitar o momento (revolucionário) para construir sua própria fundamentação teórico/filosófica, do que continuar a pedi-la emprestado de outras áreas (comunicação, educação, linguística, gestão, sociologia – matemática, física etc.), de forma mediada pela SE e eternizar os “guetos intelectuais” existentes na área. É importante desenvolver a epistemologia e a ontologia da própria área e esta é uma oportunidade histórica. É necessário fortalecer a área com fundamentos teóricos sólidos, preparando-a melhor para a construção interdisciplinar. Nas palavras de Pierce (1992):

Nosso campo (CI/B) importa teorias da comunicações, educação, linguística, gestão, psicologia, sociologia, e uma série de outras disciplinas. Que estranho. Não são muitas áreas que aceitam dissertações baseadas em tradições intelectuais de outras áreas. Este tipo de atitude em pesquisa é menos interdisciplinar do que "não-disciplinar". Em geral busca-se fundamentação teórica em outras disciplinas [...]. Vivemos em espécies de "guetos intelectuais"; nossos pesquisadores mais talentosos procuram favor imitando práticas de disciplinas consideradas superiores à nossa. Prestamos tão pouca atenção à nossa própria história intelectual que podemos ter de reconstruí-la, quase do zero. (p. 643).

Este artigo é apenas mais um dentre muitos que chamam a atenção para tal situação, com o intuito de ampliar a reflexão sobre o tema. Somente assim será possível caminhar em direção a este novo cenário de ensino e pesquisa em CI, mais abrangente (pluralismo metodológico) e ao mesmo tempo coeso. Um cenário que integra a formação em graduação àquela da pós-graduação. Um cenário mais produtivo, e mais adequado ao atendimento das fortes demandas pelo profissional da informação, cada vez mais relevante na atual sociedade da informação.

Pôde-se perceber que a epistemologia da CI inspira-se fortemente nos três paradigmas de Capurro e das abordagens de Shera e Salton, talvez com alguma prevalência das ciências sociais, porque os sistemas de informação são fundamentalmente sociais, mas são também

sistemas técnicos (i.e., são sistemas sócio-técnicos). Ficou claro que o pluralismo metodológico é um tema que pode e deve ser apoiado por todos os pesquisadores, independentemente de suas tendências epistemológicas. Este trabalho também mostrou que o pluralismo metodológico é irresistível e inevitável em Ciência da Informação.

REFERÊNCIAS

ALVARENGA, L. **Representação do conhecimento em tempo e espaço digitais**. Encontros bibli, Florianópolis, n. 15, p. 18-40, 2003.

ARAÚJO, Carlos Alberto Ávila. **O conceito de informação na Ciência da Informação**. Inf. & Soc.: Est., João Pessoa, v. 20, n. 3, p. 95-105, set./dez. 2010. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/6951>>.

ARAÚJO, V.M.R.H. **Sistemas de recuperação da informação: nova abordagem teóricoconceitual**. Ciência da Informação, Brasília, v. 24, n. 1, p. 54-76, jan./abr. 1995.

BAEZA Y. R.; RIBEIRO-NETO, B. **Modern Information Retrieval - The Concepts and Technology behind Search**. 2nd Edition, Pearson, 2011.

BARBOSA, D. M. ; BAX, M. P. ; BARACHO, R. M. A. **Estratégia de Recuperação da Informação - SIRIPE: Sistema de Recuperação da Informação em Planos de Ensino**. III Seminário de Estudos da Informação, 2013, Niterói/RJ.

BARBOSA, R. R. **Gestão da informação e do conhecimento: origens, polêmicas e perspectivas**. Informação & Informação, v. 13, p. 1-25, 2008.

CHUA, A. Y. K; YANG C. C. **The Shift Towards Multi-Disciplinarity in Information Science**. Journal of the American Society for Information Science and Technology, 59(13):2156-2170, 2008.

DIAS, E. W. **Ensino e pesquisa em ciência da informação**. Datagramazero, Rio de Janeiro, v. 3, n.5, p. 1-10, 2002.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de Banco de Dados**. 6. ed. São Paulo: Pearson Education do Brasil, 2011.

FLORIDI, L. **Afterword LIS as Applied Philosophy of Information: A Reappraisal**. Library Trends, v. 52, n. 3, p. 658-665, 2004.

_____. **On defining library and information science as applied philosophy of information**. Social Epistemology, v. 16, n. 1, p. 37-49, 2002.

FROHMAN, B. **Rules of indexing: a critique of mentalism in information retrieval theory**. Journal of Documentation, v. 46, n. 2, p. 81-101, 1990.

GOMES, Henriette Ferreira . **Interdisciplinaridade e Ciência da Informação: de característica a critério delineador de seu núcleo principal**. Datagramazero, Rio de Janeiro, v. 2, n.4, p. 1-15, 2001.

GUO, Z.; SHEFFIELD, J. **A paradigmatic and methodological examination of knowledge management research: 2000 to 2004**. Decision Support Systems, v. 44, n. 3, p. 673 – 688, Feb. 2008.

IVESON R. **Rewiring the Brain Or, Why Our Children Are Not Human: On Bernard Stiegler's *Taking Care of Youth and the Generations*** in *Parallax* 18:4 (2012), 121-125. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/13534645.2012.715462>

LANCASTER, F.W. **Indexação e resumos: teoria e prática**. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 1993. 347 p.

_____. **Vocabulary control for information retrieval**. 2th ed. Arlington: Information Resources Press, 1986. 270 p.

LOPES, I. L. **Estratégia de busca na recuperacao da informação: revisão de literatura**. Ciência da Informação, Brasília, v. 31 , n. 2, p. 60-71, maio/ago. 2002.

NAVES, M.M.L. **Estudo de fatores interferentes no processo de análise de assunto**. Perspectivas em Ciência da Informação, Belo Horizonte, v. 6, n. 2, p. 189-203, jul./dez. 2001.

PIERCE, S. **Dead Germans and the theory of librarianship**. *American Libraries*, 23, 641–643, 1992.

ROJAS, Miguel Ángel Rendón. **Bibliotecología, archivística, documentación: intradisciplina, interdisciplina o transdisciplinariedad**. 2011

SARACEVIC, T. **Information Science**. Journal of the American Society for Information Science, 50(12):1051–1063, 1999.

SOUZA, R. R. **Sistemas de recuperação de informações e mecanismos de busca na web : panorama atual e tendências**. Perspectivas em Ciência da Informação (Impresso), Belo Horizonte, v. 11, n.2, p. 161-173, 2006.

TURBAN, E. **Administração de Tecnologia da Informação: teoria e prática**. Rio de Janeiro: Campus. 2003. 598 p.

WIKIPÉDIA, *l'encyclopédie libre*. **Science de l'information**. Consultado em 8 de agosto 2013. Disponível em: <http://fr.wikipedia.org/w/index.php?title=Science_de_l%27information&oldid=95663077>.