



XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)
ISSN 2177-3688

GT 1 – Estudos Históricos e Epistemológicos da Ciência da Informação
Comunicação Oral

CIÊNCIA DA INFORMAÇÃO – SINALIZAÇÕES PARA O PRESENTE E O FUTURO¹

INFORMATION SCIENCE – TRENDS FOR THE PRESENT AND FOR THE FUTURE

Renata Maria Abrantes Baracho, UFMG
renatambaracho@gmail.com

Maurício Barcellos Almeida, UFMG
priv.mba@gmail.com

Renato Souza Rocha, FGV – UFMG
rsouzaufmg@gmail.com

Jacqueline Pawlowski Oliveira, UFMG
jackiepaw@gmail.com

Resumo: O presente trabalho discute como o realinhamento da Ciência da Informação à suas origens, em questões seminais como a representação e a organização para fins de recuperação da informação, proporciona benefícios para a ciência e para a sociedade no contexto atual. A interdisciplinaridade característica da Ciência da Informação possibilita a solução de problemas que não seriam resolvidos isoladamente, uma vez que a complexidade atual exige de abordagens integradas. O objetivo desse trabalho é contextualizar algumas das premissas fundamentais da Ciência da Informação, para então conectá-las às modernas tendências como: *web semântica*, *redes sociais*, *linked data*, *big data*, *cloud computing*, dentre outros. Ainda que “*buzz-words*”, esses termos refletem a explosão da informação como característica marcante e intrínseca da contemporaneidade. Para tal, conduziu-se breve revisão de literatura como suporte à percepção da alta relevância das especialidades históricas da Ciência da Informação no contexto moderno. Advoga-se que apenas a reconexão da Ciência da Informação a essas especialidades históricas levará o campo científico a realizar toda sua potencialidade. Os processos de criação, representação, armazenamento, organização, disseminação e consumo de informação são objetos

¹ O conteúdo textual deste artigo, os nomes e e-mails foram extraídos dos metadados informados e são de total responsabilidade dos autores do trabalho.

a privilegiar na pesquisa, assim como o contexto, desde que os fins não se confundam com os meios. De fato, a abordagem aqui apresentada corrobora com a percepção de que problemas de décadas atrás, que justificaram o aparecimento de uma assim chamada "Ciência da Informação" são hoje em dia ainda mais importantes. Espera-se que esse primeiro estudo, de uma série de muitos por vir, reacenda questões sobre a importância dos principais paradigmas históricos da Ciência da Informação para a pesquisa atual.²

Palavras-chave: Ciência da informação; *big data*; computação nas nuvens; recuperação da informação; representação do conhecimento; redes sociais.

Abstract: The present work discusses how the re-alignment between Information Science to its own origins, in issues such as representation and organization for purposes of information retrieval, provides benefits for the science “*per se*” and for the society in the current context. The inter-disciplinarily character of Information Science enables one to solve problems which could not be solved in an isolated context, insofar as the current complex problems requires integrated approaches. The goal of this work is to contextualize some fundamental premises of Information Science, and therefore to connect them to modern trends as: semantic web, social networks, linked data, big data, cloud computing, to mention but a few. Even though these terms are buzz-words, they represent the information explosion as the main and intrinsic characteristic of the contemporaneous environment. In reach your goals, we conducted a brief literature review as a support to our perception about the nowadays high relevance of the historical specificities of Information Science. We advocate that, only through the re-connection of the Information Science with such historical specificities will make possible to the field to reach its full potential. The processes of creation, representation, storage, organization, dissemination and consume of information are object to be emphasized in Information Science research, as well as the social context, since the ends is not confused with the means. Indeed, the approach presented here corroborates with the perception that the problems we had already seen decades ago, which then justified the arising of a so-called Information Science, are nowadays still more important. We hope that this essay, the first of a planned series, could shed some light to issues about the main historical paradigms of Information Science and their importance to the current research.

Keywords: *Information Science; big data; cloud computing; information retrieval; linked data.*

1 INTRODUÇÃO

Desde os anos de 1990, o aumento de informação textual não estruturada tem impactado fortemente sobre a pesquisa de comunidade acadêmicas diversas, e sobremaneira em campos especializados na organização e recuperação da informação, como a Ciência da Informação (CI). Não que a informação não tenha sempre mantido um status de alta relevância para a sociedade, e um papel chave na investigação da CI. Entretanto, a evolução da tecnologia e sua popularização via recursos como celulares, *tablets*, vídeos, redes sociais, comércio eletrônico, sistemas de geo-referenciamento, etc., tem tornado a investigação em CI cada vez mais complexa, e seu papel cada vez mais relevante.

² Trabalho parcialmente financiado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Governo do Estado de Minas Gerais, Brasil.

De fato, debates sobre o crescimento exponencial da informação e explosão da disseminação há muito se fazem presentes nas discussões e pesquisas da CI. Isso tem ocorrido, mesmo que tenha se enfatizado o estudo do contexto social em detrimento de especialidades conhecidas da CI, principalmente, a organização e a representação da informação para fins de recuperação (ALMEIDA, SOUZA e BARACHO 2015).

A reflexão que se busca no presente trabalho surgiu da constatação da importância em discutir e em analisar limites da CI e interseções da CI com áreas de cunho tecnológico que tradicionalmente lidam com informação. Acredita-se que a ênfase deva recair na tríade composta pela aplicação da tecnologia (e não em sua produção), pelas pessoas e pela sociedade, em particular, as instituições sejam públicas ou privadas. Considerando-se a quantidade de informação gerada, armazenada e distribuída continuamente em diferentes dispositivos tecnológicos, destaca-se a necessidade da atuação dos pesquisadores da CI de forma efetiva, uma vez que profissionais de outras áreas que lidam com informação não recebem formação sobre representação e recuperação da informação.

Além da preocupação com a integração do componente tecnológico em representação e organização para fins de recuperação reconhece-se, em contexto mais amplo, a importância de entender como as pessoas buscam e adquirem informação. Unindo o uso de soluções tecnológicas para resolver problemas sociais via pessoas, desenha-se um campo de pesquisa promissor. Além da preocupação com os meios tecnológicos, muitos dos quais as pessoas usam já intuitivamente (celulares, para um exemplo simples), está a preocupação com as necessidades e as motivações humanas, o que leva uma pessoa ou um grupo de pessoas a buscar e utilizar a informação.

Nesse contexto, o presente ensaio tem como objetivo ensinar uma reflexão sobre os caminhos da CI. Como princípio norteador, se vale de duas citações seminais: uma relativa às fronteiras e outra relativa às origens da CI. Com relação às fronteiras, Dias (2002, p. 2) explica que “*o caráter interdisciplinar que se procura caracterizar na ciência da informação tem dado margem a certos exageros na definição das fronteiras do campo*”. Abordando as origens e a definição da CI como campo de pesquisa, Saracevic (1996) define:

A Ciência da Informação é um campo dedicado às questões científicas e à prática profissional voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação. No tratamento destas questões são consideradas de particular interesse as vantagens das modernas tecnologias informacionais.

Dias (2002) explica assim que, mesmo interdisciplinar, uma ciência tem seu núcleo principal. Dessa forma, mesmo no âmbito interdisciplinar, há pesquisas fronteiriças que são válidas e pesquisas além da fronteira do razoável que deveriam ser objeto de estudo em outras áreas. Mesmo que se diga que essa fronteira não é clara nem facilmente identificável, o que parece ser razoável, também parece razoável afirmar que é possível distinguir uma publicação de Engenharia de uma de Medicina. Assim, da mesma forma, parece razoável que alguém consiga distinguir uma publicação da CI de uma publicação de História, ou Sociologia, ou Antropologia.

Saracevic (1996), ao mencionar o caráter interdisciplinar da CI, reconhece também a conexão entre Tecnologia da Informação (TI) e CI, considerando mesmo que a segunda está inexoravelmente ligada à evolução da primeira. Explica por exemplo, a relação da CI com a Inteligência Artificial (IA), a qual, pode causar alguma surpresa dentre os que não acreditam na pervasividade atual da tecnologia. Na verdade, a IA lida com a criação de esquemas de classificação na tentativa de proporcionar contexto às máquinas, de forma muito similar ao que a CI faz na sua busca por recuperação da informação via vocabulários controlados. Na verdade, o termo “vocabulário controlado”, cunhado no âmbito da Biblioteconomia e CI, é bastante adequado para explicar a forte tendência atual de conectar dispositivos, conhecido popularmente como “Internet das Coisas”. Essa assim chamada internet das coisas não se realiza sem vocabulários controlados nas máquinas, os quais são especialidade histórica da CI.

De forma a levar adiante essa primeira reflexão sobre os rumos e papéis da CI na sociedade moderna, o restante desse artigo está organizado da seguinte forma. A seção 2 apresenta um breve contexto histórico, bem como questões contemporâneas relacionadas. A seção 3 apresenta algumas das tendências tecnológicas que carecem da intervenção de profissionais e pesquisadores da CI. A seção 4 apresenta uma breve revisão de literatura que busca suportar, no contexto desse ensaio, os argumentos para a conexão entre CI e TI. A seção 5 descreve alguns sinais evidentes da conexão supracitada a partir de iniciativas internacionais. Finalmente, a seção 6 apresenta as considerações finais e perspectivas de trabalhos futuros.

2 CONTEXTO HISTÓRICO E CONTEMPORANEIDADE

Na tentativa de traçar um histórico do campo científico da CI, parte-se do seu surgimento na década de 50, então relacionada ao contexto de Recuperação da Informação (RI). Em 1950, *Calvin Mooers* cunhou o termo “recuperação da informação” (*information retrieval*) na década e

seguiu com ênfase na implementação do sistema (centrada no sistema). A automação do processamento fazia uso de lógica formal sem muita preocupação com usuários da informação.

No final da década de 1970, surgiu o desenvolvimento de sistemas de informação centrado no usuário. O usuário deixa de ser passivo, aquele que recebe o sistema pronto, para ser ativo. O foco passa a ser a cognição, a interatividade, as situações e contextos de uso dos sistemas. Dois estudos bibliométricos dão suporte a tal questão analisada (BARBOSA e BAX 2013).

Chua e Yang (2008) realizaram estudo bibliométrico de artigos coletados no *Journal of the American Society of Information Science and Technology* (JASIST) entre os anos de 1988 e 2007. 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. A pesquisa mostra como resultado um declínio de 13,6% dos pesquisadores originários do núcleo tradicional da CI, de 61,3% para 47,7%, ao mesmo tempo em que registra aumento considerável da presença dos pesquisadores de gestão, e principalmente daqueles ligados à tecnologia da Informação (os últimos aumentaram em 8,7%, de 16,1% para 24,8%). Guo e Sheffield (2008) analisaram artigos de cinco revistas de gestão, no período entre 2000 e 2004, encontrando artigos sobre gestão do conhecimento, com forte tendência de crescimento de estudos relacionados a TI.

Essa tendência da área já reconhecida pela área de CI, como mostra 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 da informação. (ALVARENGA, 2003, p. 39).

A declaração corrobora com a posição de Wersig (1993), retratada em Almeida, Souza e Baracho (2015), que enfatiza a necessidade de uma perspectiva pragmático-evolutiva, junto às tentativas de construção teórica da CI. Ambos os objetos de investigação empírica e interfaces com outros campos funcionam como um sistema de apoio para a permanente construção e reconstrução do campo de investigação conjunto.

Em uma publicação também seminal, Borko (1968, p.3) afirma que a CI investiga as propriedades e o comportamento da informação, as forças que regem o fluxo de informações, e os meios de processamento de informações para a acessibilidade e usabilidade. Está preocupada com que o corpo de conhecimento sobre origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação. Isto inclui a investigação de representação de informação em sistemas naturais e artificiais, o uso de códigos de mensagens para transmissão, e o estudo de dispositivos de processamento e técnicas, tais como computadores e seus sistemas. Trata-se assim, de uma ciência interdisciplinar derivada de e relacionada a áreas como matemática, lógica, linguística, psicologia, informática, artes gráficas, comunicações, biblioteconomia, gestão e outros campos semelhantes.

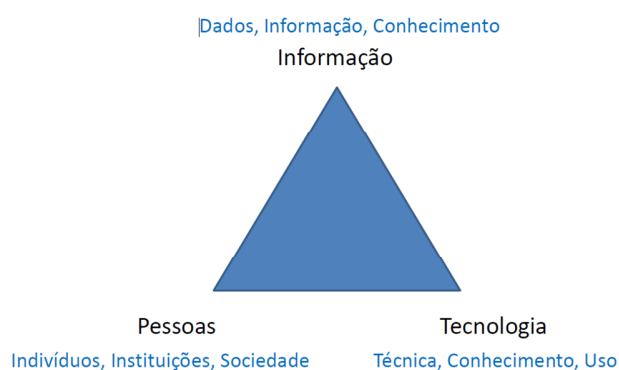
Os processos que se realizam sobre a informação incluem a criação, representação, armazenamento, organização, disseminação e consumo. Isto significa que os problemas de décadas atrás que justificam o aparecimento de uma assim chamado "Ciência da Informação" são, hoje em dia, objetos de pesquisa em crescimento exponencial de alta relevância para a sociedade. Dessa forma, são necessários e mesmo essenciais para o crescimento da área.

As questões de compreensão dos fenômenos da informação, em suas mais variadas manifestações, mais a compreensão do comportamento humano e dos problemas aplicados para tornar acessível um acervo crescente, incluindo os respectivos ajustes tecnológicos, não podem ser resolvidos no âmbito de uma única disciplina. Para exemplificar a abrangência dos temas envolvidos se faz necessário, por um lado, o entendimento do indivíduo, suas relações com a sociedade, com o contexto, a definição de suas necessidades; e, por outro lado, a elaboração de uma solução conceitual sobre a representação de um domínio, a qual precisa ser explicitada em linguagens para a criação de modelos, ao mesmo tempo em que se dá o diálogo outros campos no que se refere aos aspectos da informação, seu processamento e fluxos nos sistemas.

Em termos de contemporaneidade, acredita-se assim na necessidade de a CI alinhar-se com as necessidades de representação, organização e recuperação implícitas nos anseios da sociedade moderna, de forma para proporcionar recuperação da informação condizente com as necessidades e usos da informação pelas pessoas. As aplicações computacionais e as pesquisas de informação são campo fértil de desenvolvimento e crescimento da sociedade da informação (FLORIDI 2004).

A proposta de interseção entre áreas reflete o conceito interdisciplinar da CI, mas traz em seu bojo, ao mesmo tempo, a necessidade de delimitar uma área de conhecimento para está se consolide (DIAS 2002). Sinaliza-se assim para a necessidade da pesquisa em CI ocupar-se da informação e do conhecimento, das necessidades humanas e da compreensão sobre as formas sistemáticas de organização do conhecimento, se valendo dos recursos da tecnologia e gestão. Esse pesquisador poderá atuar em diversos contextos, o que exige formação para entendimento de questões de cunho social. Mas o contexto social deve ser o meio para a recuperação da informação e não um fim em si. Isso difere a pesquisa da CI das pesquisas em outras áreas limítrofes. O mesmo vale, e já foi explicado, com relação à relação com a TI. Essa proposta por ser representada na tríade Informação / Tecnologia / Pessoas, onde a informação e pessoa abrangem sociedade que faz uso da tecnologia em seu dia a dia. De fato, os problemas reais de organização da informação que têm sido enfrentados pela CI há anos, apenas recentemente foram sentidos pela TI (VICKERY 1997).

FIGURA 1 – Os pilares Conhecimento / Tecnologia / Pessoa



Fonte: dos autores

Três questões norteiam a busca por formas de atender tal demanda (BATES 1999). Tem-se a questão física, que abrange o estudo das características e das leis do universo da informação e do conhecimento registrado; a questão social, que realiza o estudo sobre como os usuários especializados buscam, usam e se relacionam com a informação e o conhecimento; a questão da organização, que envolve a busca por formas mais efetivas para acesso a informação e ao conhecimento registrado, do ponto de vista humano e tecnológico.

Assim, busca-se promover a pesquisa no âmbito de três grandes tendências atuais, de alta relevância para a sociedade, a pesquisa e a educação: a tendência de conexão e interação entre equipamentos via web, através de vocabulários controlados, popularmente denominados internet das coisas; a explosão de fontes de informação e conhecimento heterogêneas, que tem sido denominada de *big data* e a necessidade de torná-las úteis; a necessidade de se aprimorar os sistemas de organização do conhecimento, melhor identificar as necessidades dos usuários em contextos especializados.

Para tal, é preciso desenvolver capacidades para pesquisa e inovação com perfil e habilidades específicas para organização e representação da informação, agregando ainda o uso da tecnologia e da gestão. A proposta é a união de competências básicas para planejamento, armazenamento, organização, recuperação, disseminação e acesso a informação, alinhado com expertise sobre o estado da arte de ferramentas computacionais; competências complementares envolvem habilidades analíticas e representacionais, habilidade de expressão e comunicação em linguagens formais e naturais, metodologias de pesquisa, produção de conhecimento científico, bem como sensibilidade para as questões sociais, históricas e culturais necessárias ao entendimento do impacto das soluções encontradas num contexto global e social.

Nesse contexto, a importância subáreas como arquitetura e organização da informação além de gestão e tecnologia da informação, parece evidente. A primeira, considerando aspectos classificatórios da organização de domínios de conhecimento, taxonomias facetadas, sistemas de conceitos, análise de domínio, organização de recursos imagéticos e diferentes mídias, proteção do conhecimento, estudos métricos da informação e do conhecimento, avanços em análise de citação, além de acesso e recuperação do conhecimento em contextos digitais. A segunda, abrangendo visualização de dados, processamento de linguagem natural, computação científica, mineração de dados, modelagem de tópicos, modelagem de informação espacial, aprendizagem por máquinas, tecnologias da informação na web, modelos organizacionais, ontologia do social, recuperação em bases de dados científicas, *linked-data*, *big-data*, *cloud computing* além de engenharia organizacional.

3 LINKED DATA – BIG DATA – CLOUD COMPUTING – SOCIAL NETWORKS

Apresentam-se e analisam-se nessa seção, termos que tem sido utilizados para referências ao grande volume de informação produzido e disponibilidade na *web*, a saber: *semantic web*, *linked data*, *big-data*, *cloud computing*, *social networks*.

Em 2006, surge uma proposta inovadora de interligação de dados (*Linked Data*) liderada por Tim Bernes-Lee, fundador da W3C, para a formalização da informação disponível na Web considerando os conceitos e suas relações. Surgiu assim o termo “web semântica” (*semantic web*) (BERNERS-LEE et al, 2001). Cinco anos mais tarde, em 2006, Berners-Lee apresenta o termo *Linked Data* (BERNERS-LEE 2006). Consolidado em 2009, *Linked Data* é definido como (BIZER, HEATH & BERNERS-LEE; 2009):

“...um conjunto de melhores práticas para publicação e conexão de dados estruturados na Web. Essas melhores práticas foram adotadas por um número crescente de fornecedores de dados ao longo dos últimos três anos, levando à criação de um espaço global de dados contendo milhares de milhões de declarações - a Web de Dados”.

Linked data referência os dados da web de forma legível por máquina como objetivo de conectar toda a web a partir de diferentes fontes. Surge a necessidade de interoperabilidade com ligações entre base dados, dados externos que podem estar interligados a outros dados externos.

O termo *Big Data* representa, de uma maneira geral, o grande volume de dados produzido e que não pode ser processado sem estruturas computacionais específicas. Volume, velocidade e variedade (3 Vs) são suas características. O volume, a articulação em rede, a velocidade e diversidade possibilitam a produção de novos dados, a partir dos dados já existentes (BOYD; CRAWFORD, 2011).

O termo computação nas nuvens (*cloud computing*) também ganhou impulso nos últimos anos. Por se tratar de uma tecnologia inovadora, a computação em nuvem vem recebendo maior atenção da academia nos últimos anos. Surgiu a fim de gerir e utilizar enorme quantidade de dados que se tornaram disponíveis com o desenvolvimento da Internet. O uso da computação em nuvem tem apresentado vantagens significativas (KAMALA, PRIYA e NANDHINI 2012). As vantagens são: redução de custos, grande capacidade de armazenagem, escalabilidade, ausência de instalação e manutenção de software, acessibilidade dos serviços, dentre outros.

Este fenômeno acrescenta um nível abstrato para o processo de desmaterialização dos registros de informação. Algum tempo atrás, ainda sem confiança, as pessoas mudaram a partir

de suportes físicos para os digitais como *tablets* e dispositivos móveis. A computação em nuvem separa os usuários desses dispositivos que armazenam seus dados. Serviços de armazenamento onipresentes e muitas vezes livremente disponíveis (*DropBox*³, *Google Drive*⁴, *iCloud*⁵, para mencionar alguns) são parte da vida de milhões de pessoas, e carecem de organização e representação, como proporcionadas pela CI.

As redes sociais (*social networks*) são estruturas que reúnem atores – indivíduos e instituições – através de *links* (EASLEY et al, 2010), os quais podem ser reificados como arranjos sociais ou tecnológicos. A representação e estudo dessas redes incluem, mas não estão limitadas, as áreas de Antropologia, Biologia, Comunicação, Economia, Geografia, Ciência da Informação, Psicologia, Sociologia e sociolinguística. As redes sociais são construídas a partir de arranjos sócio-técnicos, como é o caso de dispositivos na web, por exemplo, *Twitter*⁶ e *Facebook*⁷. Em seus primeiros dias, o *Twitter* foi descartado porque não foi considerado uma fonte de informação relevante, mas em 2010 as suas mensagens de 140 caracteres começaram a ser coletadas⁸ por meio da *Library of Congress*. Em 2014, mais de 500 milhões de mensagens foram trocadas no *Twitter* por dia⁹. Da mesma forma, em 2014, o *Facebook* tinha cerca de 1,1 bilhão de usuários que trocavam um volume surpreendente de informação¹⁰.

A principal característica deste tipo de informação é a fluidez, a dinâmica e a efemeridade temática. São explorados nas atividades de monitoramento ambiental, análise de sentimento (Liu, 2012) e até mesmo a gravação da história humana. As redes sociais e seus fenômenos relacionados apontam para o caráter efêmero de ambos os registros informacionais e os instrumentos necessários para organizar esses registros. Na verdade, essas construções coletivas permitem criar produtos de extensão inimaginável, que exigirá novas abordagens e estudos (SUROWIECKI, 2004), os quais se imaginam como parte da pesquisa em CI.

Os profissionais que utilizam técnicas tradicionais de análise de assunto, classificação e categorização adaptam os seus conhecimentos para as coleções gigantescas e para a velocidade de produção de resumos e índices (MAGNUSSON; VANHARANTA, 2003; NODUS LABS,

3 <https://www.dropbox.com/>

4 <https://drive.google.com/>

5 <https://www.icloud.com/>

6 <https://twitter.com/>

7 <https://pt-br.facebook.com/>

8 <http://www.businessinsider.com/library-of-congress-is-archiving-all-of-americas-tweets-2013-1>

9 <http://www.internetlivestats.com/twitter-statistics/>

10 <http://www.statisticbrain.com/facebook-statistics/>

2012). As redes sociais têm desafiado os limites geográficos de conectividade; a nova explosão de informações eliminou a possibilidade de um tratamento individualizado de registros de informação; e a computação em nuvem tem dividido os registros.

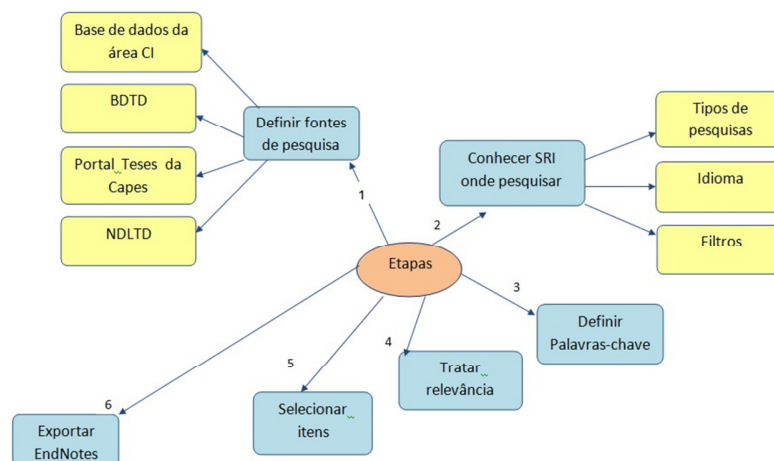
4 REVISÃO DE LITERATURA

A revisão de literatura aqui apresentada não é exaustiva, e foi realizada para dar suporte aos argumentos estabelecidos no decorrer da pesquisa. O intuito é apresentar um panorama por meio de um levantamento onde se verificou o que está sendo pesquisado sobre a CI no contexto proposto no presente artigo. Foram consultados artigos nas bases de dados da área de ciências sociais aplicadas disponíveis no portal CAPES (dentre elas LISA, SCIELO, Web of Science, ISTA, GALE, Springer Link, Wiley) além de Bibliotecas Digitais de Teses e Dissertações. Os descritores utilizados para as buscas foram: ciência da informação (*information science*), computação nas nuvens (*cloud computing*), megadados (*big data*), informação (*information*). Foram selecionados artigos, teses e dissertações que abordam pesquisas referentes a estudos realizadas na área da ciência da informação, computação nas nuvens e *big data*. Após uma leitura, um total de 41 artigos que corresponderam aos critérios de inclusão estabelecidos foram selecionados.

Este estudo verifica como estas temáticas vêm sendo abordadas na área da CI. Percebe-se que estes processos estão em fases de estudos e consolidação na área, tanto o desenvolvimento do tema *big data* como o de computação nas nuvens. É notório que para avançar em determinado campo do conhecimento é preciso primeiro conhecer o que já foi desenvolvido por outros pesquisadores. Para tanto, o objetivo dessa seção é realizar um levantamento de estudos por meio de uma revisão sistemática em relação as pesquisas desenvolvidas.

Para a preparação da revisão sistemática foi adotado o método de Sampaio e Mancini, (2007), que apresenta diretrizes essenciais para integrar as informações de um conjunto de estudos realizados separadamente sobre o tema. A revisão de literatura foi realizada a partir da formulação de uma pergunta específica que conduz a busca das publicações: “O que existe de estudo em relação ao *big data* e computação nas nuvens no âmbito da CI”?

FIGURA 2 – Procedimentos Metodológicos



Fonte: Elaborado pelos autores, 2015.

Com algumas adaptações no modelo de revisão de Sampaio e Mancini (2007), definiu-se o fluxo dos procedimentos a serem realizados no processo de revisão. Identificaram-se as bases de dados a serem pesquisadas e definiram-se os critérios para a seleção dos artigos. A figura 2 apresenta o fluxo dos procedimentos realizados nesta etapa de forma detalhada.

Na 1a. etapa avaliou-se quais fontes de pesquisas seriam selecionadas para realizar as buscas. Foram consideradas bases de dados da área de ciências sociais aplicadas disponíveis no portal da CAPES. Dentre elas estão: LISA, SCIELO, *Web of Science*, ISTA, GALE, *Springer Link*, Wiley, *Annual Reviews*. Além das bases de dados, foram consideradas como fontes de pesquisa as Bibliotecas Digitais de Teses e Dissertações (BDTD). As teses e dissertações foram consideradas por serem importantes recursos de informação científica e tecnológica que podem auxiliar no desenvolvimento de novas pesquisas. Para a seleção das BDTDs que fariam parte do conjunto de fontes a serem pesquisadas, foram analisadas quais instituições de ensino superior possuíam curso de pós-graduação na área de ciência da informação. A consulta foi realizada na relação de cursos recomendados e reconhecidos pela CAPES.

Foram selecionadas as Bibliotecas Digitais de Teses e Dissertações das instituições: Universidade Federal de Santa Catarina (UFSC); Universidade Federal de São Carlos (UFSCAR); Universidade de Brasília (UNB); Universidade Federal do Rio de Janeiro (UFRJ); Instituto Brasileiro de Ciência e Tecnologia (IBICT). Universidade Estado de São Paulo (UNESP), Universidade Federal Fluminense (UFF). Outras fontes de informações consultadas

foram Banco de Teses da CAPES, que é um sistema online oficial do governo brasileiro para depósito de teses e dissertações brasileiras, vinculado ao Ministério da Educação (MEC), e também, a *Networked Digital Library of Theses and Dissertation* – NDLTD que reúne os repositórios de teses e dissertações de diversos países do mundo.

A 2ª etapa foi direcionada ao reconhecimento dos tipos de pesquisas disponíveis em cada uma das fontes. Uma busca eficaz envolve não só uma estratégia que inclua termos adequados, mas também conhecer a base de dados para proceder com as buscas corretamente. Vale ressaltar que as buscas realizadas nas BDTDs da UNB, UFF, UFSCAR, apresentaram resultados insatisfatórios. O sistema considera palavras como “da” e “nas” como termo de busca e apresenta um quantitativo de resultado muito alto. Foi necessário ajustar a busca utilizando opções de pesquisa avançada.

No intuito de acessar os estudos que abordavam pesquisa sobre CI, *big data* e computação nas nuvens, na 3ª. etapa do processo de revisão sistemática foram selecionados como termos de busca os descritores: ciência da informação (*information science*), computação nas nuvens (*cloud computing*), megadados (*big data*) e informação (*information*). Utilizou-se combinações de termos com operadores booleanos. Cabe ressaltar que os resultados encontrados nas BDTDs foram bem menores do que o esperado. Melhores resultados quando se tratavam de artigos nas bases de dados do Portal CAPES. Na tabela 1 estão descritos alguns exemplos de resultados. Apesar dos descritores estarem em português, a pesquisa foi realizada também em inglês.

TABELA 1 – Estratégias e resultados de buscas

Estratégia de busca	IBICT	UFSCAR	NDLTD	UNB	UFF
Campo assunto: ciência da informação AND <i>big data</i>	0	142* 0 Relevância	1 1 Relevância (baixa)	17 0 Relevância	9 0 Relevância
Campo assunto: ciência da informação AND computação nas nuvens	0	0	0	39*** 0 Relevância	91*** 0 Relevância
Campo assunto: computação nas nuvens	3 1 Relevância	0	349 0 Relevância	54 0 Relevância	127*** 0 Relevância
Campo assunto: <i>big data</i>	9 0 Relevância	70 0 Relevância	148** 0 Relevância	33 0 Relevância	27 0 Relevância
Campo Título: computação nas nuvens	2 1 Relevância	9 0 Relevância	113 0 Relevância	3 0 Relevância	1 0 Relevância

Campo título: <i>big data</i>	2 0 Relevância	26 0 Relevância	12 0 Relevância	16 0 Relevância	23 0 Relevância
Campo livre “cloud computing” AND “information science”	1 0 Relevância	2072	18.246	50	105

*Resultados insatisfatórios considerando os termos pesquisados. Havia registros que não continham as palavras pesquisadas

** Apenas 76 itens eram dos idiomas selecionados para inclusão. A maioria em chinês e árabe. Apenas 1 item em português.

***Sistema de busca separa o termo “computação” e “ciência” dos descritores “ciência da informação” e “computação nas nuvens” e a maioria do resultado são pesquisas de assunto de “ciência da computação”

Fonte: Elaborado pelos autores.

Na busca, foram analisados e selecionados apenas os estudos cujo título, resumo ou corpo do artigo tivessem relação com o objeto do presente estudo e respondessem a questão: “Quais os estudos estão sendo realizados sobre big data e computação nas nuvens no âmbito da ciência da informação?” Após a leitura dos resumos dos estudos selecionados, pertinentes à questão proposta, foi realizada a recuperação dos artigos em texto completo.

É possível observar na tabela 1 que as buscas realizadas nas BDTDs apresentaram um baixo índice de resultados sobre *big data* e computação nas nuvens no âmbito da ciência da informação. Ressalta-se a baixa relevância em torno de estudo proposto. Verificou-se na análise que a maioria dos estudos recuperados são voltados para a área da ciência da computação os quais, foram descartados seguindo os critérios de exclusão adotados. Em contrapartida, foram realizadas algumas buscas pelo campo de título o que também não apresentou resultados condizentes com a proposta. A busca realizada em campo livre trouxe uma quantidade grande de resultados, fazendo com que a precisão ficasse ruim e inviável analisar a relevância.

Foram definidos critérios de inclusão e exclusão dos artigos, com base na pergunta que norteia a revisão, critérios como tipo de estudo, entre outros. Estes critérios não foram muitos rígidos devido a relação que a CI tem com outras áreas de conhecimento. Avaliou-se a pertinência do conteúdo por meio de leitura prévia dos resumos dos trabalhos encontrados.

Tabela 2 - Critérios de inclusão e exclusão de estudos para a revisão sistemática

Inclusão	Exclusão
Título que corresponde ao assunto do estudo e que responda a questão da revisão	Estudos que focam apenas a área da ciência da computação
Idioma português, inglês, italiano e espanhol	Conteúdo diferente ao tema investigado
Teses, dissertações, artigos, trabalhos de conclusão de cursos	

modalidade de estudo: teórico/revisão, empírico/pesquisa ou relato de experiência/estudo de caso.	
---	--

Fonte: Elaborado pelos autores.

A busca nas bases de dados mencionadas foi realizada em julho de 2015. O estudo foi realizado por meio de pesquisas em artigos, teses, dissertações e trabalhos de conclusão de cursos sem limitar períodos de data de publicação, com o objetivo de avaliar também, se estudos referentes a área de CI relacionados a big data e computação nas nuvens são recentes.

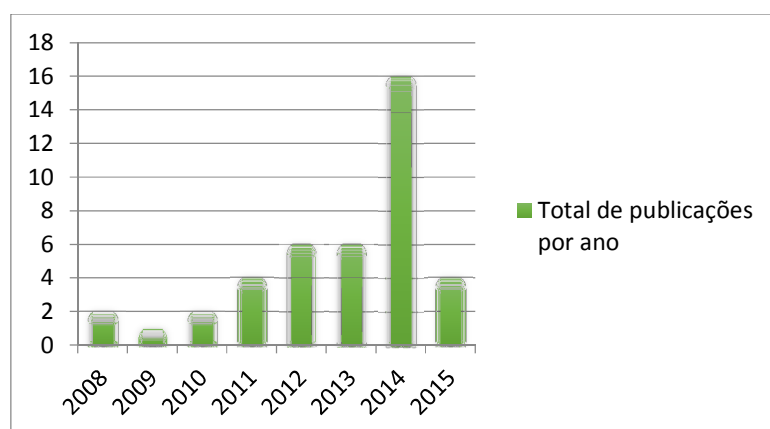
Para as etapas 5 e 6 foi realizada a análise e seleção dos estudos encontrados, utilizou-se a premissa de que quando o título e o resumo não fossem esclarecedores o estudo seria selecionado.

Descrição dos trabalhos analisados contendo autoria e ano de publicação: Aharony, Noa (2014); Aharony, Noa (2015); Bansode, S. Y. and Pujar, S. M (2012); Aharony, Noa (2014); Valdo Pasqui (2010); Yan Han (2013); B. N. Thimmaiah (2014); Goldner, Matt (2011); Bushhousen, Ellie (2011); R. M. Chang; R. Kauffman; Y. Kwon (2014); Philip Chen, C.L. ; Zhang, C.-Y. (2014); Colmenero-Ferreira, Fernando; de Oliveira, Adicinéia (2012); Janna Hastings; Gwen Alexandra Frishkoff ; Barry Smith; Russell A Poldrack; Anita Bandrowski; Jessica A Turner (2014); Doug Howe; Maria Costanzo; Petra Fey; Takashi Gojobori; Linda Hannick; Winston Hide; David P. Hill; Renate Kania; Mary Schaeffer; Susan St Pierre; Simon Twigger; Owen White; Seung Yon Rhee (2008); James, Roger (2012); Li, Yuelin; Sonnenwald, Diane H. (2014); Liao, Jianxin; Yang, Di; Li, Tonghong; Wang, Jingyu; Qi, Qi; Zhu, Xiaomin (2014); Lunardi, Marcia S. ; De Castro, Jose Muanis F. ; Monat, Andre S. (2008); May, Mike (2014); Melo, Humberto Gustavo de (2014); Mihai Doinea ; Paul Pocatilu (2014); Pang, L.Y.; Zhong, Ray Y.; Fang, J.; Huang, George Q. (2015); Osterhaus, Leah C. (2010); Waxer, Nathan ; Ninan, David ; Ma, Alfred ; Dominguez, Norma (2013); Poulymenopoulou, M.; Malamateniou, F.; Vassilacopoulos, G. (2014); Prantosh K Paul; Kalyan Kumar ; Dipak Chatterjee (2014); Bhagyashree Dhakulkar; Seetal S. Shinde; Nandkishor Gosavi (2012); Claudio José Silva Ribeiro (2014); S. P. Singh; R. S. R. Veralakshmi (2012); Xiaocan (Lucy) Wang; Jie Huang Xiaocan (Lucy) Wang; Jie Huang (2011); Corrado, Edward M.; Moulaison, Heather Lea (2012); Murley, Diane (2009); Geeta Rani, Renu Bala (2015); Erik W. Kuiler (2014); Tinati, Ramine; Halford, Susan; Carr, Leslie; Pope, Catherine (2014); Vartika Kulshrestha; Dr. Seema Verma (2013); Westphall, Carlos; Villarreal, Sergio (2013); Ricardo M. Pimenta (2013); Danah boyd;

Kate Crawford (2011); José Fernando Modesto da Silva; Marcos Luiz Mucheroni (2013); Elkin, Lauren S. (2015).

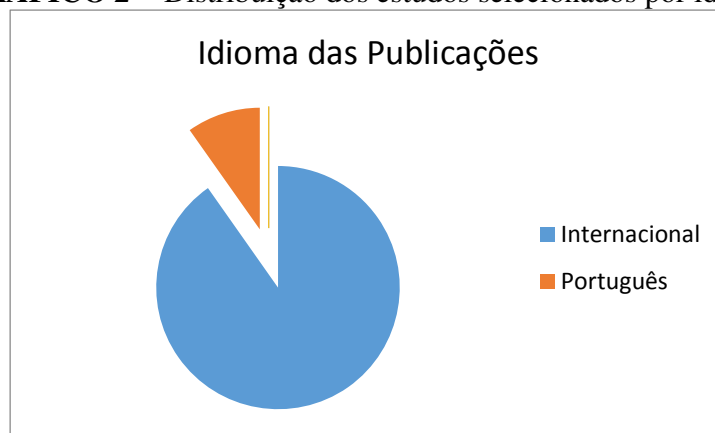
O gráfico 1 apresenta a distribuição do total de publicações por ano e o gráfico 2 a distribuição dos estudos selecionados por idioma. Foi possível observar que o índice de publicações sobre a temática nos últimos oito anos cresceu muito, o que pode indicar uma maior preocupação da ciência sobre a importância do tema (gráfico 1). Foi verificado também que o número maior de estudos selecionados é o de publicações internacionais (gráfico 2) comparando com as de produções brasileiras. Dentre os 41 estudos selecionados apenas 4 são em português o que corresponde a 9,75%. Tais dados apontam para o fato de que o estudo sobre *big data* e computação nas nuvens é ainda bastante incipiente no Brasil, o que reforça a necessidade de novas pesquisas em nosso contexto.

GRÁFICO 1 - Distribuição dos estudos selecionados por ano



Fonte: Elaborado pelos autores, 2015.

GRÁFICO 2 - Distribuição dos estudos selecionados por idioma



Fonte: Elaborado pela autora, 2015.

Uma análise criteriosa da literatura selecionada com o objetivo de avaliar sua qualidade, identificar conceitos importantes e concluir sobre o que a literatura informa em relação a intervenção proposta, apontando questões que necessitam de novos estudos, fará com que este movimento de pesquisa possibilite que outros pesquisadores interessados possam se engajar nesta discussão, levando este tema para além da tecnologia.

É necessário estimular a realização de novas pesquisas sobre a temática, possibilitando melhor entendimento dessa relevante questão no contexto brasileiro pois, a abordagem do assunto na literatura de Ciência da Informação ainda é considerada baixa.

5 SINALIZAÇÕES DO FUTURO: AS *I-SCHOOLS*

Para refletir sobre a CI nacional, é necessário verificar a experiência de países estrangeiros, em diferentes estágios de evolução tecnológica, como por exemplo Portugal e Estados Unidos. Ambos adotam a moderna visão da CI, ainda inédita no Brasil, as *i-schools*¹¹. Essas iniciativas enfatizam o entendimento da relação entre informação, tecnologia e pessoas, pressupondo que a expertise em todas as formas em que a informação se manifestam é requisito básico para o progresso na ciência, nos negócios, na educação e cultura.

O movimento é denominado *i-schools*, ou Escolas de Informação. Nestas escolas têm sido criados, já há alguns anos, os modernos sucessores dos programas de Ciência da Informação. Os

¹¹ <http://ischools.org/>

programas de pós-graduação no formato *i-school* enfatizam o entendimento da relação entre informação e conhecimento, tecnologia e pessoas. Esses programas pressupõem que a *expertise* em todas as formas em que a informação e o conhecimento se manifestam é requisito básico para o progresso na ciência, nos negócios, na educação e cultura. Essa *expertise* abrange uma ampla compreensão dos usos e dos usuários especializados, bem como de técnicas de organização do conhecimento, no contexto das tecnologias e suas aplicações.

As *i-schools* são uma resposta direta às necessidades relacionadas à organização da grande massa de dados hoje disponível, a partir de soluções novas e criativas. Implícita a essa iniciativa, está a necessidade real de uma ciência para lidar com informação em todas as suas manifestações. Em muitas universidades, em várias partes do mundo, as escolas de ciência da informação foram transformadas em "escolas de informação", onde grupos diversos reúnem os perfis necessários capazes de exercer as competências sócio-técnicas interdisciplinares demandadas pela sociedade. Além disso, as conferências da *International Society of Knowledge Organization* (ISKO 2014) indicam uma direção clara que busca uma ampla cobertura, considerando essenciais o emprego de especialidades históricas da CI.

A emergência das *i-schools* tem origens na constatação básica de que o conhecimento humano, já há algumas décadas, não está restrito àquele registrado em livros ou documentos tradicionais. Estudos do início do século XX¹², Lyman, Varian (2003) já apontavam que em apenas um ano eram gerados cinco exabytes (5×10¹⁸) de informação por ano, com crescimento (também anual) de 30%. Para se ter uma ideia, cabe mencionar que cinco exabytes equivalem a informação contida em todas as coleções de cerca de 37.000 novas bibliotecas do tamanho da *Library of Congress* nos Estados Unidos. Essa informação encontra-se “desterritorializada”, registrada em uma variedade de formatos, mídias, fontes, bancos de dados, web, para citar algumas poucas possibilidades.

Pode-se dizer que os estudos sobre o tema ainda se encontram em estágio inicial, com um número ainda reduzido de pesquisas em relação. Espera-se que este movimento de pesquisa sobre os temas na área da CI, fulgure novos projetos de pesquisas e possibilite que outros pesquisadores interessados possam se engajar nesta discussão, levando este tema para além da tecnologia.

¹² <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>

6 CONSIDERAÇÕES FINAIS

O presente trabalho discutiu a necessidade de um realinhamento da CI com suas origens, em questões seminais como a representação e a organização para fins de recuperação da informação. Acredita-se que tal realinhamento proporcionará benefícios para a ciência e para a sociedade no contexto atual.

De fato, a interdisciplinaridade característica da CI possibilita a solução de problemas que não seriam resolvidos isoladamente, uma vez que a complexidade atual exige de abordagens integradas. Assim, contextualizou-se aqui algumas das premissas fundamentais da CI, para então conectá-las às modernas tendências como: web, web semântica, redes sociais, linked data, big data, cloud computing, dentre outros. Esses termos refletem a explosão da informação como característica marcante e intrínseca da contemporaneidade.

Para tal, conduziu-se breve revisão de literatura como suporte à percepção da alta relevância das especialidades históricas da CI no contexto moderno. Argumentou-se que reconexão da CI às essas especialidades históricas levará o campo científico a realizar toda sua potencialidade. Os processos de criação, representação, armazenamento, organização, disseminação e consumo de informação são objetos a privilegiar na pesquisa, assim como o contexto, desde que os fins não se confundam com os meios.

A abordagem aqui apresentada se fundamenta em problemas conhecidos, que há décadas atrás justificaram o aparecimento de uma assim chamada CI. A partir da nova realidade tecnológica, tais problemas hoje se apresentam em dimensões extremamente maiores. Espera-se que esse primeiro estudo, de uma série de muitos por vir, tenha levantado questões sobre a importância dos principais paradigmas históricos da CI para a pesquisa atual da área.

REFERÊNCIAS

ALMEIDA, M. B.; SOUZA, R. R.; PORTO, R. B. Looking for the identity of Information Science in the age of big data, computing clouds and social networks. *In: INTERNATIONAL SYMPOSIUM OF INFORMATION SCIENCE*, 14, 2015, Zadar. **Anais...** Zadar, 2015.

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

BATES, M. **The invisible substrate of information science**. Disponível em: <<http://pages.gseis.ucla.edu/faculty/bates/substrate.html>>. Acesso em: 10 jun. 2015.

BARBOSA D. M.; BAX, M. P. A recuperação da informação como base fundamental da Ciência da Informação e sua relação com a Ciência da Computação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14, 2013. Florianópolis. **Anais...**, Florianópolis, 2013.

BERNERS-LEE, T. *et al.* The semantic web. **Scientific American**, v. 284, n. 5, p. 28-37, 2001.

BERNERS-LEE, T. **Design Issues** - Linked Data. 2006.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked data-the story so far. **Semantic Services, Interoperability and Web Applications**: emerging concepts, p. 205-227, 2009.

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

BOYD, D.; CRAWFORD, K. Six provocations for big data. In: A DECADE IN INTERNET TIME: SYMPOSIUM ON THE DYNAMICS OF THE INTERNET AND SOCIETY, 21, 2011. **Proceedings...** Oxford: Oxford Internet Institute's, 2011. Disponível em: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431>. Acesso em: 6 jul. 2015.

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**, v. 59. n. 13, p. 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.

FLORIDI, L. Afterword LIS as applied philosophy of information: a reappraisal. **Library Trends**, v. 52, n. 3, p. 658-665, 2004.

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

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.

KAMALA, B.; PRIYA, B.; NANDHINI, J. M. Platform autonomous custom scalable service using service oriented cloud computing architecture. **International Journal of Engineering Research and Applications** (IJERA), v. 2, n. 2, p.1467-1471, mar./apr. 2012.

LIU, B. **Sentiment Analysis and Opinion Mining**, Chicago: Morgan & Claypool. 2012.

LYMAN, P.; VARIAN, H. R. **How much information?** 2003 .Disponível em: <<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>>. Acesso em: 08 abr. 2015.

MAGNUSSON, C.; VANHARANTA, H. Visualizing sequences of texts using collocational networks. *In: 3rd Intern. Conf. Leipzig*. v. 2734, p. 276-283, 2003.

NODUS LABS. *Texttexture*: visualize any text as a network. Retrieved July 11, 2013. Disponível em: <http://texttexture.com/> Acesso em: 08 abr. 2015.

SAMPAIO, R. F.; MANCINI, M. C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Rev. Bras. Fisioterapia**, v.11, n.1, p.83-89, jan/fev, 2007.

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

SUROWIECKI, J. *The wisdom of crowds*: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. New York: Anchor. 2004.

VICKERY, B. C. Ontologies. *Journal of Information Science*. v. 23. n.4, p. 227–286, 1997.

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