

Qualidade da Informação e Recuperação de Informação: uma revisão da literatura

Guilherme Martins Alvarez, Alexandre L. Gonçalves

Universidade Federal de Santa Catarina Campus Universitário Reitor João David
Ferreira Lima, Trindade, Florianópolis, SC, Brasil CEP 88040-900

guilherme.alvarez@gmail.com, a.l.goncalves@ufsc.br

Abstract:

With the expansion of Information and Communication Technologies, individuals have passed from just information consumers to content producers, significantly increasing the amount of information available on the web. In addition, lack of quality standards of information published on the web makes it difficult to find relevant information for users. Thus, this paper aims to conduct a systematic review of the literature about the relationship between information quality and information retrieval, mapping the intellectual production developed up to the present day. The survey was conducted in bibliographic data bases Scopus® and Web of Knowledge®, using the following keywords: "information quality" and "information retrieval". From the identification of these studies this research contributes to the understanding of the application possibilities of information quality and information retrieval in organizations. The results found highlight an overview of the current stage of studies on the relationship between information quality and information retrieval.

Palavras-chave: Qualidade da informação, Recuperação de Informação; Extração de Informação; Motores de Busca;

1. Introdução

O rápido crescimento da web como um ambiente de troca de informação e a falta de padrões sobre essas informações levaram a inúmeros problemas de qualidade de informação (KNIGHT; BURN, 2005). É fácil adicionar informações à web através de sites, *wikis* e *blogs*, porém é difícil distinguir uma informação precisa e confiável de uma informação imprecisa e não confiável (WANG; ZENG; TANG, 2010). Pois não há um controle de qualidade e qualquer pessoa pode escrever qualquer coisa na web.

Um dos maiores desafios para os sistemas de busca é percorrer a vasta massa de conteúdos questionáveis publicados na web, a fim de retornar resultados de qualidade para os seus usuários. A qualidade dessas informações é importante tanto para usuários que realizam buscas na web, quanto para organizações que as utilizam em suas tomadas de decisão.

De acordo com Redman (1998), há um grande interesse comercial na área de Qualidade da Informação devido ao aumento de custos e impacto provocado nas organizações, em consequência da baixa qualidade das suas informações. Portanto,

diante da enorme quantidade de dados disponíveis aos usuários, faz-se necessário o uso de meios para acessar essas informações considerando medidas de qualidade.

Nesse sentido, pode-se utilizar recursos da Recuperação de Informação (RI), pois o seu objetivo é facilitar o acesso as informações de maior relevância, conforme as necessidades dos usuários (CECI, 2010). De acordo com Santarém Segundo e Vidotti (2011), a Recuperação de Informação (RI) tem sido muito discutida na área da Ciência da Informação, e a busca por informação de qualidade se tornou importante nessas pesquisas.

Para Knight e Burn (2005), o foco na Qualidade da Informação sob a perspectiva da Recuperação de Informação é uma área relativamente nova, mas é absolutamente crítica, diante da necessidade de se utilizar as ferramentas de RI para a busca e extração de informações na web.

Diversos modelos têm sido utilizados para avaliar a Qualidade da Informação em ferramentas de Recuperação de Informação, considerando medidas, dimensões da qualidade e a relevância das informações para os usuários. Porém aplicar esses conceitos em sistemas recuperação da informação é uma tarefa complexa (KNIGHT; BURN, 2005).

A partir do contexto anteriormente declarado, apresenta-se a seguinte questão: Como a Qualidade da Informação tem sido utilizada em sistemas e ferramentas de Recuperação da Informação, a fim de melhorar a sua qualidade para os usuários?

2. Aspectos conceituais

Para Wang e Strong (1996), o termo Qualidade da Informação (QI) é descrito como dados que são "prontos para o uso", mas o conceito é relativo, pois as informações consideradas adequadas para determinado uso, podem não possuir qualidade suficiente para um outro uso. (TAYI; BALLOU, 1998).

De acordo com Klein (2001), a Qualidade da Informação é considerada como um conceito multi-dimensional, com diferentes características atribuídas dependendo ponto de vista filosófico do autor, em que vários fatores determinam o seu estado, existência e aplicação (KNIGHT, 2005).

O crescimento das fontes de informação na web aumentou o interesse no estudo da qualidade da informação na perspectiva da ciência da informação. Estes trabalhos têm focado em dimensões e atributos das fontes de informação e no papel dos indivíduos como consumidores de informação (TAYLOR, 2014).

Existem diversos trabalhos relacionados a mensuração da qualidade da informação, porém existem fatores que dificultam essa tarefa, como: a origem da informação, a subjetividade implícita na necessidade do usuário, a vasta massa de dados, entre outras. (NAUMANN; ROLKER, 2000)

De acordo com Shanks e Corbitt (1999) a qualidade de uma informação deve ser avaliada no contexto da sua geração e utilização. Isto é porque os atributos e dimensões de qualidade da informação podem variar, dependendo do seu contexto e domínio. No

entanto, na visão de Katerattanakul e Siau (1999), a qualidade da informação precisa ser avaliada de acordo com o uso pretendido.

Segundo Calazans (2012), diversos autores implementaram modelos de qualidade da informação, sugerindo alguns princípios que podem ser aplicados para a melhoria da qualidade da informação, como: acurácia, completude, atualização e os seus benefícios relacionados a satisfação das necessidades e expectativas do cliente, tais como eficiência, confiabilidade, efetividade, entre outros.

De acordo com Knight e Burn (2005), definir Qualidade da Informação é uma questão complexa, que se torna ainda mais difícil se avaliada no contexto da Recuperação de Informação, a partir de fontes não validadas, como a World Wide Web. Os autores afirmam ainda que, compreender a QI a partir do ponto de vista dos usuários da informação, envolve o entendimento dos processos de Recuperação de Informação na web, que envolvem a busca de um termo, seguida de um processo de decisão do usuário baseado em conceitos como precisão, atualidade e utilidade (ROSE; LEVINSON, 2004).

De acordo com Baeza-Yates e Ribeiro-Neto (1999), a Recuperação da Informação (RI) é uma disciplina da computação que estuda a representação, armazenamento, organização e acesso à itens de informação. Os autores afirmam ainda que a RI, através de suas técnicas, deve proporcionar uma forma fácil de acesso a informação desejada pelo usuário.

Zhu e Gauch (2000) afirmam que, a grande maioria dos sistemas de Recuperação de Informação, apoiam-se em algoritmos de similaridade baseados nas estatísticas de frequência dos termos de busca, ignorando a Qualidade da Informação. Dessa maneira, os documentos são recuperados sem considerar a sua qualidade.

De acordo com Salton (1968), a Recuperação de informação, é o campo de pesquisa que estuda a análise, estrutura, armazenamento, organização, recuperação e busca de informação. Para Kowalski (1997), os sistemas de RI objetivam minimizar a dificuldade do usuário em localizar a informação desejada.

Na visão de Taylor (2014), a informação é consumida através de um processo comunicativo, que inicia acessando as informações em uma determinada fonte. Uma das fontes mais dominantes da informação hoje são os motores de busca comerciais.

Os sistemas de RI podem ser classificados em duas grandes categorias, centralizados e distribuídos. Os sistemas de recuperação de informação centralizadas exigem que todos os documentos sejam armazenados em um único local e as suas consultas são controladas através desse local. Em contraste, os distribuídos permitem que os usuários acessem simultaneamente documentos armazenados em diversos locais (ZHU; GAUCH, 2000).

Segundo Rijsbergen (1979), grande parte das pesquisas de RI objetiva a melhoria da eficiência e eficácia da tarefa de recuperação das informações. A eficiência é normalmente avaliada utilizando medidas computacionais, enquanto que a eficácia, é frequentemente calculada através de medidas como recall e precision (BAEZA-YATES; RIBEIRO-NETO, 1999).

3. Método utilizado

De acordo com Sampaio e Mancini (2007), uma revisão sistemática é um método de pesquisa que utiliza como fonte de dados a literatura a respeito um tema definido, que apresenta um resumo das evidências relacionadas a uma estratégia de intervenção específica, através da aplicação de métodos explícitos e sistematizados de busca, análise crítica e síntese da informação selecionada.

As revisões sistemáticas proporcionam a utilização de uma gama maior de resultados pertinentes, pois não restringem as conclusões das pesquisas de determinado estudo à leitura de poucos materiais.

O autor, por essa lógica, define essa forma de revisão em cinco etapas, como forma de estratégia, detalhadas a seguir:

- Primeira etapa: envolve a definição da temática e formulação da pergunta de pesquisa;
- Segunda etapa: definição de termos ou palavras-chave, estratégias de busca, bases de dados e de outras fontes de informação a serem pesquisadas;
- Terceira etapa: revisão e seleção dos artigos através da avaliação dos títulos e resumos identificados na busca inicial, estabelecendo critérios de inclusão e exclusão de artigos;
- Quarta etapa: apreciação dos registros selecionados da base de conhecimento;
- Quinta etapa: apresentam-se os resultados obtidos após a análise dos estudos.

4. Execução e resultados

A temática definida para a revisão sistemática está focada em qualidade da informação e recuperação de informação. Tem-se como objetivo a obtenção de artigos científicos relacionados ao tema e disponíveis em bases de conhecimento.

Após definir o tema e o objetivo da revisão, elaborou-se o seguinte questionamento como pergunta da revisão: Como a Qualidade da Informação tem sido utilizada em sistemas e ferramentas de Recuperação da Informação, a fim de melhorar a sua qualidade para os usuários?

O próximo passo foi a definição da fonte de dados, então, optou-se pelas bases Scopus e Web of Knowledge tendo também em vista a grande quantidade de artigos disponíveis e a abrangência de áreas suportadas.

Visto que, a proposta se direciona a identificar as pesquisas já realizadas nesse tema, definiu-se como termos para a busca a combinação das palavras “*information quality*” e “*information retrieval*”.

O conector entre os termos utilizados para a busca foi “E”. Com isso, pretendeu-se recuperar os artigos relacionados aos dois termos de busca selecionados. É importante salientar que as bases de conhecimento escolhidas têm, na sua grande

maioria, artigos em inglês. A consulta montada ficou da seguinte maneira: “*information quality*” AND “*information retrieval*”.

Na Scopus, os campos de buscas selecionados foram “Title” e “Abstract”, enquanto na Web of Knowledge, os campos de buscas selecionados foram “Nome da publicação”, “Tópico” e “Título”.

As bases de dados retornaram 60 publicações (contando os duplicados) distribuídas da seguinte forma: Web of Knowledge (31) e Scopus (29). Em seguida, os resultados de cada uma das bases foram importados no *EndNote* para eliminação dos artigos duplicados entre bases. A retirada destes trabalhos culminou com a seleção de 49 trabalhos, distribuídos conforme demonstra a Figura 1.

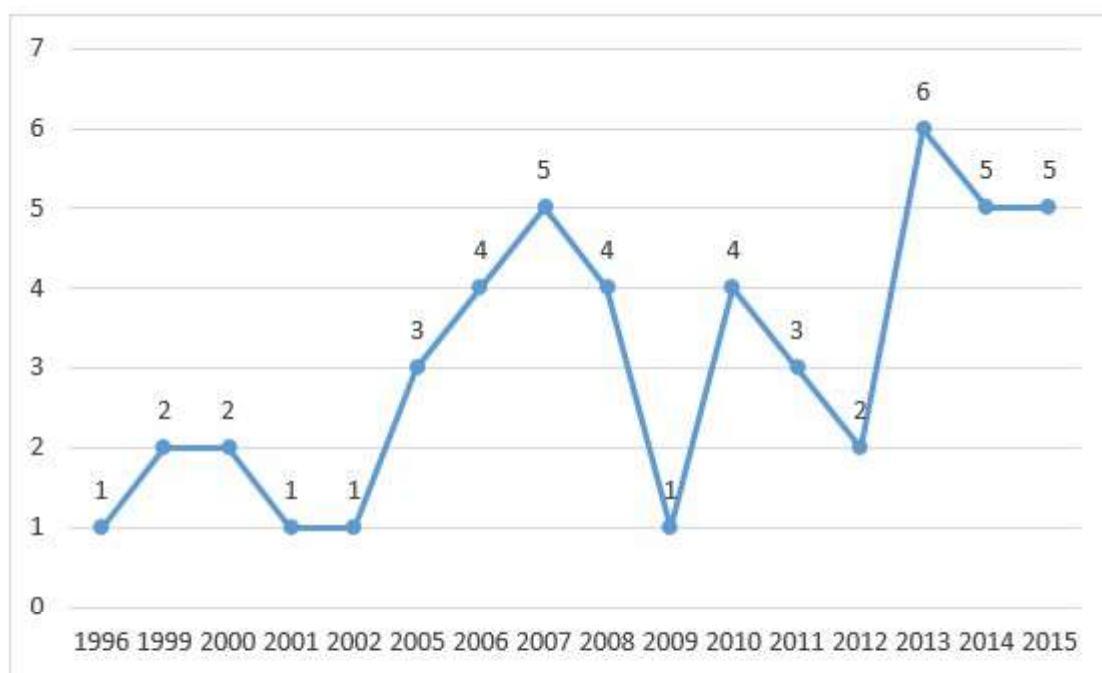


Figura 1 – Artigos retornados da busca distribuídos por ano

Dos 49 artigos resultantes da pesquisa efetuada na Scopus, apenas 35 estavam disponíveis para download. Após a efetuar a leitura dos títulos e resumos dos artigos recuperados, verificou-se que alguns não possuíam relação direta com os temas abordados nessa pesquisa. Portanto, esses trabalhos foram removidos, restando 25 trabalhos para serem analisados. A Figura 2 expõe a distribuição dos artigos coletados por ano.

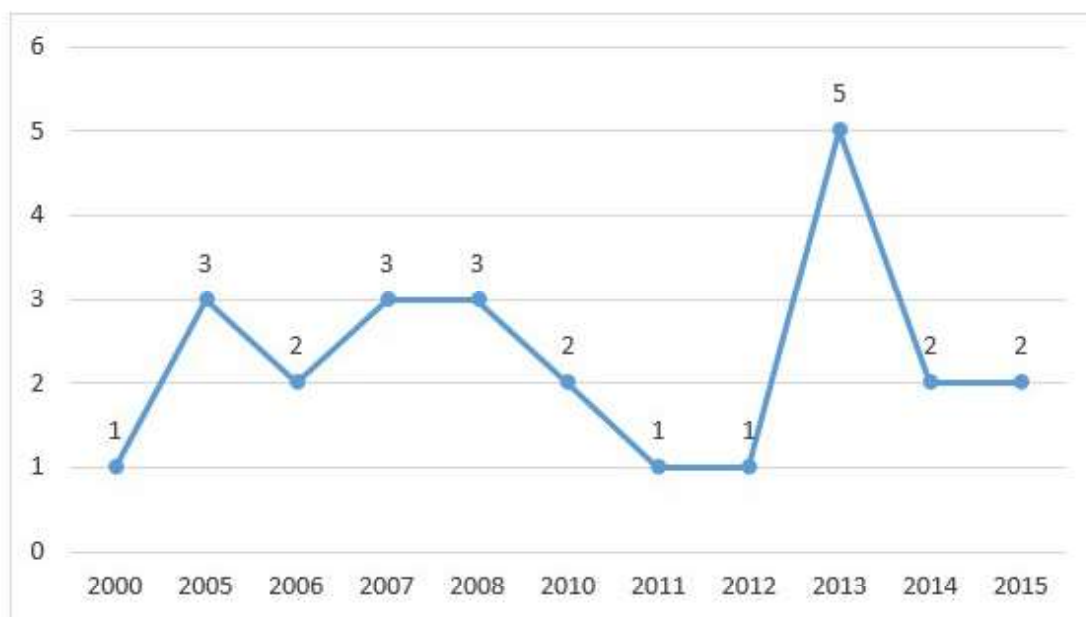


Figura 2 – Artigos coletados por ano

Dos 10 trabalhos removidos por não possuírem relação com os temas pesquisados, 3 tratavam de questões da educação, 4 abordavam temas da medicina, 1 tratava de transferência de músicas pela web através de redes P2P, 1 era da área da filosofia e 1 abordava temas da sociologia.

Ao analisar a Figura 2, percebe-se que a quantidade de publicações relacionadas à qualidade da informação e recuperação de informação teve um crescimento por volta do ano de 2013. Essa informação também pode ser constada ao analisar a Figura 1, que apresenta todos os artigos disponíveis nas bases de documentos sobre os temas.

A partir do conjunto de trabalhos recuperados, foi possível identificar 102 palavras-chave. As palavras-chaves “information quality” (qualidade da informação) e “information retrieval” (recuperação de informação) destacam-se com 11 citações cada e a palavra-chave “search engines” (motores de busca) vem logo atrás com 4 ocorrências. A análise da nuvem de tags demonstrada na Figura 3, gerada a partir das palavras-chave dos trabalhos, evidencia os termos com maior número de citações.

	retrieval and its application
Knight e Burn (2005)	Developing a Framework for Assessing Information Quality on the World Wide Web
Liu e Liang (2005)	A joint link and content approach to information retrieval and distillation
Gaudinat et al. (2006)	Health search engine with e-document analysis for reliable search results
Chung et al. (2006)	Supporting non-English Web searching: An experiment on the Spanish business and the Arabic medical intelligence portals
Herrera-Viedma et al. (2006)	A fuzzy linguistic model to evaluate the quality of Web sites that store XML documents
Eysenbach e Thomson (2007)	The FA4CT Algorithm A New Model and Tool for Consumers to Assess and Filter Health Information on the Internet
Wang et al. (2007)	Factoid mining based content trust model for information retrieval
Pereira e Da Silva (2008)	The use of cognitive authority for information retrieval in folksonomy-based systems
Scherp et al. (2008)	Interaction and user experiences with multimedia technologies: Challenges and future topics
Otterbacher (2008)	Managing information in online product review communities: A comparison of two approaches
Wang, Zeng e Tang (2010)	Using evidence based content trust model for spam detection
Bounhas et al. (2010)	Toward a Computer Study of the Reliability of Arabic Stories
Knight (2011)	The combined conceptual life-cycle model of information quality: Part 1, an investigative framework
Schulte et al. (2012)	COV4SWS.KOM: Information quality-aware matchmaking for semantic services
Puentes et al. (2013)	Information quality measurement of medical encoding support based on usability
Wang et al. (2013)	ExpertRank: A topic-aware expert finding algorithm for online knowledge communities
Figueiredo et al. (2013)	Assessing the quality of textual features in social media
Lopes e Ribeiro (2013)	Measuring the value of health query translation: An

	analysis by user language proficiency
Albaham e Salim (2013)	Quality Biased Thread Retrieval using the Voting Model
Todoran et al. (2014)	Toward the quality evaluation of complex information systems
Taylor (2014)	Postmodernist and consumerist influences on information consumption
Sağlam e Temizel (2015)	A framework for automatic information quality ranking of diabetes websites
Kim et al. (2015)	A blog ranking algorithm using analysis of both blog influence and characteristics of blog posts

Fonte: Elaborado pelo autor

O primeiro trabalho lido, desenvolvido por Zhu e Gauch (2000), apresenta um método que combina ranking de similaridade com ranking de qualidade, em ambientes de busca centralizados e distribuídos. O trabalho explora métricas para avaliar a qualidade de resultados de busca, que incluem, valor, disponibilidade, ruído de informação, autoridade, popularidade e coesão.

Ma e Tanaka (2005), propõe um modelo para extração de conteúdo de *data streams* e páginas da web, baseando-se em conceitos de estruturas de tópicos. Os autores apresentam um sistema chamado WebTelop, que aumenta o conteúdo de programas de TV em tempo real, utilizando páginas da Web complementares. O WebTeleop permite que os usuários visualizem um programa de televisão e as páginas Web com conteúdo similar mesmo tempo. Para avaliar a qualidade da informação recuperada, os autores utilizaram métricas de precisão, acurácia, completude e similaridade.

Knight e Burn (2005), apresentam o IQIP, um modelo para gerenciar a implementação e escolha de algoritmos de qualidade para um motor de busca da Internet. Inicialmente, os autores apresentam um contexto histórico dos frameworks desenvolvidos para de Qualidade da Informação e as dimensões mais utilizadas. Após abordar o contexto histórico, os autores apresentam o modelo IQIP, e definem que as dimensões de QI a serem utilizadas devem ser escolhidas com base na natureza do ambiente do sistema. Assim, os autores apresentam alguns pontos que devem ser avaliados durante a escolha das dimensões de QI, mas não definem quais dimensões de QI devem ser utilizadas no seu modelo.

No trabalho de Liu e Liang (2005), é desenvolvem um método de pesquisa baseado em conteúdo e links entre as informações em um grafo. Esse método utiliza um índice de relevância de conteúdo para medir a qualidade da informação e um modelo SWM (Sliding Window Model) para construir um grafo dinâmico e controlar o nível de busca no grafo. Além disso, é utilizada uma abordagem de amortecimento exponencial para guiar a busca e melhorar a sua eficiência, prevendo os caminhos mais valiosos de pesquisa e descartando caminhos menos relevantes. A avaliação da Qualidade de

Informação é realizada utilizando medidas de relevância e importância com indicadores propostos pelos autores.

Gaudinat et al. (2006) apresentam o WRAPIN, uma ferramenta de busca que utiliza tecnologias avançadas de Recuperação de Informação para avaliar a Qualidade da Informação sintetizando conceitos médicos, conclusões e referências contidas na literatura de saúde, para identificar fontes precisas e relevantes. A Qualidade da Informação é avaliada utilizando métodos de similaridade e frequência de termos (utilizando o algoritmo TF-IDF).

O trabalho de Chung et al. (2006), propõe uma abordagem independente de linguagem para a construção de portais de busca da Web, a fim de apoiar a busca em língua não inglesa. Para medir a Qualidade da Informação, os autores elaboraram um modelo utilizando as 16 dimensões propostas por Wang e Strong (1996), porém removendo a dimensão segurança.

[Herrera-Viedma](#) et al. (2006), desenvolvem um modelo de avaliação baseado em computação difusa para medir a qualidade da informação de sites que armazenam documentos XML. A fim de avaliar a Qualidade da Informação nos websites, os autores propõem um esquema baseado nas seguintes dimensões de qualidade: credibilidade, relevância, oportunidade, plenitude compreensibilidade de sites, originalidade, compreensibilidade de documentos, concisão e ferramentas de navegação.

No trabalho de Eysenbach e Thomson (2007), é desenvolvido um novo modelo e ferramenta para acessar e consultar informações relacionadas a saúde na web. Segundo os autores, o modelo proposto substitui a abordagem de questionário ou lista de verificação através uma abordagem dinâmica e orientada a processo. Para avaliar a Qualidade da Informação, é proposto um modelo baseado em 7 critérios de qualidade, que utiliza uma escala de pontuação para classificar cada critério.

Wang et al. (2007), apresentam um modelo que extrai dados factuais de documentos e aplica recursos de aprendizado de máquina para ranqueá-los. Os autores tratam a Qualidade da Informação como um problema de classificação, então utilizam uma abordagem de aprendizado supervisionado com SMV (Support Vector Machine) para ranquear os dados factuais.

Pereira e Da Silva (2008), propõem a atribuição de autoridade cognitiva para reconhecer a Qualidade da Informação, qualificando suas fontes, ou seja, o seu autor. Aplicou-se essa qualificação no processo de Recuperação da Informação em um sistema baseado em folksonomia e através de um processo de simulação os autores demonstraram os resultados da proposta. Os autores tratam a Qualidade da Informação como um problema de classificação e categorização da informação, De acordo com Pereira e Da Silva (2008), a qualidade da informação é subjetiva, pois depende de quem está fazendo a julgamento.

Em seu trabalho, Scherp et al. (2008), investigam temas e desafios da interação e experiência do usuário em multimídia. Os autores avaliam 3 domínios de aplicação utilizando como dimensões de Qualidade da Informação a qualidade do conteúdo, qualidade de apresentação, ambiente, interatividade e expectativa dos usuários.

Otterbacher (2008), compara duas propostas de solução para solucionar problemas de sobrecarga de informação e controle de qualidade em websites de e-

commerce. São utilizadas técnicas de mineração textual, frequência de termos, entropia e perplexidade, a fim de avaliar a qualidade das opiniões contidas em websites de e-commerce.

Em seu trabalho, Wang, Zeng e Tang (2010), desenvolvem um modelo de confiança de conteúdo para um algoritmo de detecção de spam com base em características de texto. O modelo utiliza técnicas de aprendizado de máquina para criar um algoritmo eficiente e preciso. Para avaliar a Qualidade da Informação, foram selecionadas 6 dimensões de qualidade passíveis de análise automática: coesão, valor, autoridade, disponibilidade, ruído de informação e popularidade.

Bounhas et al. (2010), apresentam metodologias e ferramentas baseadas em aspectos de confiabilidade para narrativas elaboradas em línguas árabes. QI baseada em dimensões. Para avaliar a Qualidade da Informação, os autores utilizam as dimensões autoridade, objetividade e verificabilidade, baseados no trabalho de Knight e Burn (2005).

O trabalho de Knight (2011), apresenta uma revisão da literatura do tema Qualidade da Informação e constata a falta de uma abordagem coerente e consistente para identificar e medir os sistemas de QI. Além disso, apresenta um modelo de QI baseado em ciclo de vida conceitual combinado, através do qual é possível examinar a interação do usuário/informação e percepções de QI. São relacionadas 16 dimensões de QI, que são agrupadas com base na sua semelhança em quatro classificações gerais; QI intrínseca; QI de representação; QI interacional e QI contextual. Knight (2011) desenvolve uma análise profunda comparando o seu modelo com diversos frameworks publicados.

Schulte et al. (2012), propõem uma abordagem semântica de Qualidade de Informação para matchmaking de serviços. De acordo com os autores, em matchmaking de serviços a Qualidade da Informação depende do impacto positivo que uma descrição de serviço terá sobre o resultado do processo de matchmaking. Se determinada informação tiver um impacto positivo sobre a identificação dos serviços correspondentes, a sua qualidade é assumida como sendo relativamente alta. A abordagem proposta fornece métricas de ponderação e similaridade para avaliar se o impacto gerado é positivo ou negativo.

Em seu trabalho, Puentes et al. (2013), definem e validam uma nova abordagem de Qualidade da Informação destinada a avaliar a qualidade de listas de códigos médicos. O modelo proposto é baseado em um estudo de usabilidade de como especialistas e médicos aplicam a codificação médica assistida por computador. A medida proposta, chamada ADN, avalia a precisão, dispersão e ruído das listas de códigos médicos.

Wang et al. (2013) desenvolvem um novo algoritmo de busca de especialistas, ExpertRank, que avalia conhecimentos com base na relevância de documentos e autoridade em sua comunidade de conhecimento. As informações são classificadas utilizando métricas de similaridade de conteúdo e importância baseada em links. Para avaliar a experiência dos candidatos, o modelo avalia a relevância experiência e a sua autoridade/influência na comunidade que faz parte.

Figueiredo et al. (2013), analisam a qualidade relativa das diferentes características textuais comumente encontradas em diversas aplicações populares da Revista Tecnologia da Informação e Comunicação: Teoria e Prática. Vol.1 n.1, 2017

Web 2.0. A fim de avaliar a Qualidade da Informação, são utilizadas métricas de Recuperação da Informação.

Lopes e Ribeiro (2013), avaliam o efeito de traduzir termos de busca do domínio médico escritos na língua nativa dos usuários para o idioma Inglês, utilizando usuários com diferentes níveis de proficiência em Inglês. A qualidade das informações traduzidas é avaliada através da sua legibilidade e acurácia. A legibilidade é mensurada utilizando a Medida Simples de Gobbledygook (SMOG) e a acurácia é avaliada por especialistas em medicina.

Albaham e Salim (2013) apresentam indicadores de baseados em influencia para avaliar a qualidade de conteúdo em mensagens dispostas em estruturas formadas por tópicos. A qualidade das mensagens e tópicos é avaliada utilizando métricas de baseadas no modelo de ranking de votação. De acordo com os autores, a qualidade do conteúdo das mensagens é avaliada utilizando a dimensão quantidade de dados, utilizando o número de caracteres, número de palavras, número de palavras únicas e número de sentenças em uma mensagem.

O trabalho de Todoran et al. (2014), propõe uma metodologia para estimar a Qualidade da Informação de forma dinâmica em sistemas complexos e instáveis. A Qualidade da Informação é avaliada utilizando as dimensões propostas no trabalho de Wang e Strong (1996).

Taylor (2014), examina o consumo e qualidade da informação no contexto da teoria da Ciência da Informação, consumismo, pós-modernismo e as mudanças significativas na Tecnologia da Informação, fornecendo algum contexto histórico e examinando o impacto da sua convergência. Além disso, o autor examina as dimensões da Qualidade Informações em relação ao consumismo e pós-modernismo, considerando o impacto da produção e disseminação de informação.

Belen Sağlam e Taskaya Temizel (2015), desenvolvem um framework que identifica e classifica automaticamente os sites de diabetes de acordo com a sua relevância e Qualidade da Informação baseada no conteúdo do site. O framework proposto classifica os sites de diabetes de acordo com sua qualidade de conteúdo relevância, combinando técnicas de RI com recursos léxicos.

Kim et al. (2015) apresentam um algoritmo para ranquear blogs de forma eficiente. O algoritmo executa um processo de classificação, através não só de análise de interconexão de blogs, mas também de pesos estruturais dos conteúdos contidos nos blogs.

5. Conclusões

Este trabalho mapeou e analisou o campo de pesquisas que abordam os conceitos de qualidade da informação e recuperação de informação. Foi possível delinear um panorama sistemático com indicadores bibliométricos desse campo de estudo, identificando os trabalhos publicados e apontando as tendências dessa literatura.

Ao verificar e listar os principais estudos e autores nessa área de pesquisa, este trabalho também fornece subsídios para a continuidade dos estudos em outros campos, Revista Tecnologia da Informação e Comunicação: Teoria e Prática. Vol.1 n.1, 2017

formação de redes de pesquisas, expansão de linhas de pesquisas que estabeleçam e investiguem em profundidade essas áreas.

As mídias sociais, websites, blogs, fóruns e sites de avaliação de produtos, são exemplos de meios através dos quais usuários têm publicado informações. A Recuperação de Informação e a Qualidade da Informação, são disciplinas que buscam identificar e extrair informações relevantes de determinado conteúdo, e definir a sua qualidade, grau de confiança ou relevância em relação ao objeto alvo desta busca.

Através desse trabalho, percebe-se o aumento dos estudos na área da qualidade da informação e recuperação de informação, além da constante utilização de dados e informações presentes em recursos da *Web 2.0* para a análise e posterior tomada de decisão por parte dos usuários.

Com a análise realizada, o presente trabalho avança na compreensão do relacionamento entre as áreas de qualidade da informação e recuperação de informação, demonstrando as diversas possibilidades de uso de dimensões da qualidade da informação aliadas a técnicas de recuperação da informação. Além disso, verifica-se uma carência de estudos que tratem claramente dos conceitos relativos a recuperação de informação utilizando modelos semânticos, apoiados por ontologias e dimensões de qualidade da informação.

Referências

Albaham, A.T, and Salim.N. (2013). "Quality biased thread retrieval using the voting model." Proceedings of the 18th Australasian Document Computing Symposium. ACM, 2013.

Baeza-Yates, R. and Ribeiro-Neto. B. (1999). "Modern information retrieval". Vol. 463. New York: ACM press.

Belen Sağlam, R. and Taskaya Temizel, T. (2015). "A framework for automatic information quality ranking of diabetes websites." Informatics for Health and Social Care 40.1; 45-66.

Bounhas, I. et al. (2010). "Toward a computer study of the reliability of Arabic stories." Journal of the American Society for Information Science and Technology 61.8, 1686-1705.

Calazans, A. (2012). "Qualidade da informação: conceitos e aplicações." Transinformação-ISSN 2318-0889 20.1.

Ceci, F. (2010). "Um Modelo Semi-automático Para a Construção e Manutenção de Ontologias a partir de bases de documentos não estruturados". Dissertação (Mestrado)-Universidade Federal de Santa Catarina, Florianópolis, 2010.

Chung, W, et al. (2006). "Supporting non-English Web searching: An experiment on the Spanish business and the Arabic medical intelligence portals." Decision Support Systems 42.3, 1697-1714.

Eysenbach, G. and Thomson, M. (2007). "The FA4CT Algorithm: A New Model and Tool for Consumers to Assess and Filter." Medinfo 2007 129, 142.

- Figueiredo, F. et al. (2013). "Assessing the quality of textual features in social media." *Information Processing & Management* 49.1, 222-247.
- Gaudinat, A. et al. (2006). "Health search engine with e-document analysis for reliable search results." *International Journal of Medical Informatics* 75.1, 73-85.
- Herrera-Viedma, E. et al. (2007). "A fuzzy linguistic model to evaluate the quality of Web sites that store XML documents." *International Journal of Approximate Reasoning* 46.1, 226-253.
- Kim, J. et al. (2015). "A blog ranking algorithm using analysis of both blog influence and characteristics of blog posts." *Cluster Computing* 18.1, 157-164.
- Klein, B. (2001). "User perceptions of data quality: Internet and traditional text sources." *Journal of Computer Information Systems* 41.4, 9-15.
- Knight, S. (2011). "The combined conceptual life-cycle model of information quality: part 1, an investigative framework." *International journal of information quality* 2.3, 205-230.
- Knight, S. and Burn, J. (2005). "Developing a framework for assessing information quality on the World Wide Web." *Informing Science: International Journal of an Emerging Transdiscipline* 8.5, 159-172.
- Kowalski, G. (1998). "Information retrieval systems: theory and implementation." *Computers and Mathematics with Applications* 5.35, 133.
- Katerattanakul, P. and Siau, K. (1999). "Measuring information quality of web sites: development of an instrument." *Proceedings of the 20th international conference on Information Systems*. Association for Information Systems.
- Liu, Y. and Liang, M. (2005). "A joint link and content approach to information retrieval and distillation." *E-Commerce Technology, 2005. CEC 2005. Seventh IEEE International Conference on*. IEEE.
- Lopes, C. and Ribeiro, C. (2013). "Measuring the value of health query translation: an analysis by user language proficiency." *Journal of the American Society for Information Science and Technology* 64.5, 951-963.
- Ma, Q, and Tanaka, K. (2005). "Topic-structure-based complementary information retrieval and its application." *ACM Transactions on Asian Language Information Processing (TALIP)* 4.4, 475-503.
- Otterbacher, J. (2008). "Managing information in online product review communities: a comparison of two approaches." *Proc. 16th European Conference on Information Systems (ECIS 2008)*, Association for Information Systems.
- Pereira, R. and Da Silva, S. R. P. (2008). "The use of cognitive authority for information retrieval in folksonomy-based systems." *Web Engineering, 2008. ICWE'08. Eighth International Conference on*. IEEE.
- Puentes, J. et al. (2013). "Information quality measurement of medical encoding support based on usability." *Computer methods and programs in biomedicine* 112.3, 329-342.
- Redman, Thomas C. (1998). "The impact of poor data quality on the typical enterprise." *Communications of the ACM* 41.2, 79-82.

- Van R. C. J. (1979). "Information retrieval". Dept. of Computer Science, University of Glasgow."
- Rose, D. E., and Levinson, D. (2004). "Understanding user goals in web search." Proceedings of the 13th international conference on World Wide Web. ACM.
- Salton, G. (1968). "Automatic information organization and retrieval."
- Santarem Segundo, J. E., Vidotti, S. and Borsetti G. (2013) "Representação iterativa e folksonomia assistida para repositórios digitais".
- Scherp, A. et al. (2008). "Interaction and user experiences with multimedia technologies: challenges and future topics." Proceedings of the 3rd ACM international workshop on Human-centered computing. ACM.
- Schulte, S. et al. (2012). "COV4SWS. KOM: information quality-aware matchmaking for semantic services." Extended Semantic Web Conference. Springer Berlin Heidelberg.
- Tayi, G. K. and Ballou, D. (1998). "Examining data quality." Communications of the ACM 41.2,54-57.
- Taylor, A. (2014). "Postmodernist and consumerist influences on information consumption." Kybernetes 43.6,924-934.
- Todoran, I.G. et al. (2014). "Toward the quality evaluation of complex information systems." SPIE Defense+ Security. International Society for Optics and Photonics.
- Wang, R., and Strong, D. (1996). "Beyond accuracy: What data quality means to data consumers." Journal of management information systems 12.4, 5-33.
- Wang, W. et al. (2007). "Factoid mining based content trust model for information retrieval." Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer Berlin Heidelberg.
- Wang, W. Zeng. G. and Tang, D. (2010). "Using evidence based content trust model for spam detection." Expert Systems with Applications 37.8, 5599-5606.
- Wang, G. A, et al. (2013). "ExpertRank: A topic-aware expert finding algorithm for online knowledge communities." Decision Support Systems 54.3, 1442-1451.
- Zhu, X, and Gauch. S (2000). "Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web." Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. ACM.