

Recuperação de informação em dados ligados: um modelo baseado em mapas conceituais e análise de redes complexas

Information retrieval in linked data: A model based on concept maps and complex networks analysis

Henrique Monteiro CRISTOVÃO¹  0000-0003-2011-7022

Jorge Henrique Cabral FERNANDES²  0000-0001-9954-8305

Resumo

Este artigo apresenta um modelo para recuperação de informação em dados abertos ligados, usando métodos e operações de redes complexas para classificação e seleção de informações, bem como mapas conceituais para apresentação das informações recuperadas ao usuário. O modelo evidencia os relacionamentos entre os termos de consulta que representam uma necessidade informacional e os apresenta enquanto mapas conceituais. A hipótese subjacente é que o relacionamento do usuário com a informação recuperada ocorre à luz da equação fundamental da Ciência da Informação de Brookes, em que a estrutura cognitiva do cognoscente é uma rede complexa que é modulada pela informação recuperada, a qual, por sua vez, é derivada de uma rede complexa. A rede complexa final é mapeada em um mapa conceitual resultante aprimorado com heurísticas. Com características qualitativas e o emprego de abordagem exploratória, a pesquisa realizou primeiramente um teste piloto de recuperação da informação, que permitiu aferir os algoritmos empregados no ranqueamento e seleção nas redes de informação intermediárias, servindo de base para implementação de um protótipo. O protótipo empregou uma base de conhecimento de dados abertos ligados (*linked open data*), derivada da DBpedia, sobre a qual foram realizadas operações de análise de redes complexas, apresentando revocação e precisão relevantes, perante uma validação aplicada a um grupo de 17 usuários. Os resultados são promissores quanto ao uso de operações de redes complexas e mapas conceituais na recuperação de informação, especialmente em dados ligados. Como continuidade da pesquisa, observou-se demanda por ações mais interativas e pela realização de experimentos em outras bases de conhecimento.

Palavras-chave: Mapas conceituais. Recuperação da informação. Redes de informação. *Web* semântica.

Abstract

This article presents a model for information retrieval in linked open data using methods and complex network operations for ranking and selecting information, and concept maps for presenting the retrieved information to the user. The model shows the relationships

¹ Universidade Federal do Espírito Santo, Centro Universitário Norte do Espírito Santo, Departamento de Computação e Eletrônica. Rodovia BR 101 Norte, Km 60, Bairro Litorâneo, 29932-540, São Mateus, ES, Brasil. Correspondência para/Correspondence to: H.M. CRISTOVÃO. E-mail: <henrique.cristovao@ufes.br>.

² Universidade de Brasília, Departamento de Ciência da Computação, Programa de Pós-Graduação em Ciência da Informação. Brasília, DF, Brasil.

Apoio: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior e Conselho Nacional de Desenvolvimento Científico e Tecnológico (Processo nº 205480/2014-4).

Recebido em 21 de julho de 2016, versão final reapresentada em 7 de julho de 2017 e aprovado em 21 de agosto de 2017.

Como citar este artigo/How to cite this article

Cristovão, H. M.; Fernandes, J. H. C. Recuperação de informação em dados ligados: um modelo baseado em mapas conceituais e análise de redes complexas. *Transinformação*, v. 30, n. 2, p. 193-207, 2018. <http://dx.doi.org/10.1590/2318-08892018000200005>



between query terms that represent an informational need and presents them as concept maps. The underlying hypothesis is that the user's relationship to the retrieved information occurs in the light of Brookes' fundamental equation of information science. The cognitive structure of the cognoscente is a complex network that is modulated by the retrieved information which, in turn, is derived from a complex network. The final complex network is mapped into a resulting concept map enhanced by heuristics, such as the application of controlled vocabulary. The first study conducted, with qualitative characteristics and using an exploratory approach, was an information retrieval pilot test. It allowed the assessment of the algorithms used in the ranking and selection of the intermediate information networks and provided the framework for the implementation of a prototype. The prototype used a knowledge base of linked open data, derived from DBpedia, on which complex network analysis were carried out. The validation of the model presented relevant recall and precision when applied to a group of 17 users. The results are promising for the use of complex network operations and concept maps for information retrieval, especially linked data. Further research should observe the demand for more interactive actions and conduct experiments in other knowledge bases.

Keywords: Concept maps. Information retrieval. Information networks. Semantic web.

Introdução

Muitas variáveis determinam o sucesso na Recuperação de Informação (RI), tais como o comportamento do usuário (Saracevic, 2010) e suas necessidades informacionais, bem como a organização da informação e os processos de comunicação (Wersig; Neveling, 1975). O estudo das necessidades informacionais de usuários tem tomado um novo rumo desde o surgimento da *World Wide Web* (WEB), que estabeleceu novas tecnologias de organização, busca e disseminação da informação (Saracevic, 2010). A *Web* é um fenômeno social de grande escala, que apresenta propriedades emergentes e comportamentos transformadores (Shadbolt *et al.*, 2013).

A propriedade essencial da *Web* é a sua universalidade, favorecendo a comunicação social e unindo línguas e culturas diferentes (Berners-Lee *et al.*, 2001). A *Web* é distribuída, tem grande volume de informações disponíveis em repositórios não estruturados; é ubíqua e sofre constantes mudanças (Baeza-Yates; Ribeiro-Neto, 2011). A Ciência da *Web* é uma área de pesquisa com desafios muito variados (Stuckenschmidt, 2012) e um de seus maiores problemas, apontado por autores como Berners-Lee (2010), Bauer e Kaltenböck (2012) e Auer *et al.* (2013), é a dificuldade de interoperabilidade entre sistemas, decorrente da falta de definição de formatos e da incapacidade de sistemas e organizações trabalharem de forma cooperativa. Para minimizar esse problema, o *World Wide Web Consortium* (W3C), principal organização de padronização na área de tecnologias web, emite recomendações sobre as principais linguagens e protocolos (Mika, 2007).

As tecnologias da *web* semântica, padronizadas pelo W3C, indicam soluções para o problema da interoperabilidade. Considerada como uma extensão da *Web* atual, o desenvolvimento da *web* semântica parte do pressuposto de que, no momento, a maior parte do conteúdo da *Web* é destinada à interpretação por humanos, e não é facilmente operável por sistemas computacionais. Diante dessa dificuldade, a tecnologia busca estimular a organização de dados na forma de relacionamentos conceituais, em redes, permitindo a atribuição de significados aos dados na *Web*, habilitando sistemas computacionais e pessoas a trabalharem em conjunto na RI.

Um dos desdobramentos da evolução da *Web* e da *web* semântica foi a abertura de dados por governos e instituições, viabilizada pelo uso da tecnologia de *linked data*, ou de dados ligados. (Open Definition, 2015). Uma base de *linked data* é constituída por um conjunto de triplas, representados no formato denominado de *Resource Description Framework* (RDF). Uma tripla RDF é composta por: (1) um sujeito/recurso (*subject*); (2) um objeto/valor (*object*); e (3) um relacionamento entre sujeito e objeto. O relacionamento representa um predicado/propriedade (*predicate*) e, dessa forma, uma tripla pode ser lida como uma sentença composta por sujeito, predicado e objeto. A promoção de dados ligados em um contexto de dados abertos, isto é, dados que são publicamente disponíveis na *Web*, ocorre por meio da tecnologia de *Linked Open Data* (LOD), ou dados abertos ligados, um refinamento do *linked data*. Os dados abertos ligados viabilizam a criação de serviços diferenciados, estimulando a inovação e a produção de conhecimento, como um mecanismo para gestão da integração da informação (Bauer; Kaltenböck, 2012).

Muitos são os relatos de aplicações de LOD em diversas áreas do conhecimento. Por exemplo, Santos Neto *et al.* (2013) avaliaram o impacto da LOD na integração de dados de arquivos, bibliotecas e museus, que normalmente possuem acervos isolados com seus códigos próprios e maneiras particulares de representar a informação.

Outra tecnologia baseada na utilização de redes, mas com origens na educação, é a de mapa conceitual. Um mapa conceitual é uma ferramenta que auxilia na organização e representação do conhecimento (Novak, 1977), sendo universal e ubíqua (Novak; Cañas, 2010), isto é, empregada no apoio a atividades de educação realizadas por usuários de todas as idades, em todas as partes do mundo e em qualquer idioma. Diversos autores argumentam o quanto os mapas conceituais também são bons para comunicar e disseminar informações. Para Vekiri (2002) há menor esforço mental para compreender um texto quando acompanhado de mapas conceituais. Orrantia (2012) mostra que os mapas conceituais auxiliam na disseminação da informação. Valerio, Leake e Cañas (2012) comprovaram que mapas melhoram substancialmente as habilidades de compreensão de leitura dos usuários no quesito velocidade, em comparação à leitura somente de texto. Lima (2004) argumenta que a característica gráfica do mapa conceitual auxilia na compreensão das relações entre os conceitos e do conhecimento no todo. Zhang (2008) observa que, sem o auxílio de visualização gráfica, há necessidade de maior abstração de informações e, conseqüentemente, menor percepção ou compreensão dos dados e informações.

A utilização de redes também está muito presente em estudos cognitivos ligados à informação. Nas décadas de 1970 e 1980, Brookes escreveu uma série de artigos para discutir fundamentos da Ciência da Informação (CI). Ele propôs uma equação fundamental para a informação e o conhecimento (Brookes, 1980) que ficou sendo conhecida como a “Equação Fundamental da Ciência da Informação” e teve sua importância confirmada por vários autores, como Neill (1982), Le Coadic (1996), Todd (1999), Araújo (2003), Robredo (2003), Nascimento (2006), Batista, Costa e Alvares (2007), Pereira (2008), Bawden (2011), Moraes (2013), Pontes Junior, Carvalho e Azevedo (2013), mesmo que alguns desses relatassem discordâncias quanto a algumas ideias associadas à equação. Além disso, conforme levantamento realizado por Pereira (2008), de 1980 a 2008, 106 autores apresentaram trabalhos que citam os artigos de Brookes.

Brookes (1980) formulou sua equação da seguinte forma: $K[S] + \Delta I = K[S + \Delta S]$, onde $K[S]$ denota a estrutura cognitiva do sujeito; ΔI é uma nova informação recebida pelo sujeito que, relacionando-se com a sua estrutura cognitiva atual $K[S]$, provoca alterações representadas por $+\Delta S$; $K[S + \Delta S]$ representa a nova estrutura cognitiva do sujeito após relacionamento com a nova informação ΔI e em função do seu novo estado $S + \Delta S$. A equação segue o paradigma cognitivo da Ciência da Informação, baseado na ideia de que a informação provoca transformações nas estruturas cognitivas de um indivíduo.

Brookes (1980) também observou que a parcela ΔI poderia ser definida como um pequeno pedaço de conhecimento ΔK , gerando a equação alternativa $K[S] + \Delta K = K[S + \Delta S]$. Porém, ele esclareceu que ΔI pode ter diferentes efeitos sobre diferentes estruturas de conhecimento e, portanto, poderia sem prejuízo permanecer ΔI . Além disso, o autor também deixa claro que a equação não diz que o conhecimento é simplesmente aumentado com a chegada da nova informação, mas que a absorção da nova informação em contato com a estrutura de conhecimento do sujeito causa uma modificação nos relacionamentos conceituais já existentes e, portanto, não deve ser admitida como um simples incremento de informação, sendo que a percepção desta é dependente da observação sensorial do sujeito. Esses relacionamentos conceituais também podem ser representados na forma de redes.

A teoria e a prática de análise e manipulação de redes de quaisquer tipos são investigadas na Ciência das Redes. Trata-se de um campo de pesquisa interdisciplinar que busca compreender a emergência de redes na natureza, na tecnologia e na sociedade, por meio de um conjunto de ferramentas e princípios unificadores (Barabási, 2013). De forma simplificada, redes são quaisquer agregados de elementos relacionados entre si, tais como moléculas inter-relacionadas em uma solução, pessoas inter-relacionadas em uma sociedade, palavras inter-relacionadas em um texto, conceitos inter-relacionadas em um mapa. Nesse contexto, as redes complexas são

aquelas redes não triviais, que apresentam propriedades impossíveis de serem expressas em uma forma sintética, especialmente devido à dinâmica emergente trazida pelos relacionamentos não lineares e dinâmicos entre os seus elementos constituintes. As redes complexas se manifestam na estrutura e comportamento de todas as entidades vivas do mundo, em várias áreas do conhecimento, como, por exemplo, as redes informacionais encontradas na *Web*, as redes biológicas, as redes linguísticas etc.

O estudo das redes complexas busca compreender a natureza persistente dos fenômenos emergentes de uma rede e aplica-se a uma ampla gama de áreas do conhecimento. O estudo das redes complexas apresenta sobreposição com o campo mais restrito da Análise de Redes Sociais (ARS). A ARS estuda redes, mas trata apenas das redes cujos elementos são entidades sociais, tais como pessoas e organizações (Wasserman; Faust, 1994), relacionando-se entre si. A ARS busca explicar os fenômenos sociológicos de forma estrutural e, portanto, está mais centrada no universo das relações entre agentes sociais humanos. Na sobreposição entre as redes complexas e as redes sociais encontram-se as redes de informação (Newman, 2010), cuja análise depende de muitas ideias oriundas das Ciências Sociais, e grande parte da linguagem usada para descrever essas ideias reflete sua origem sociológica. Dessa forma, os métodos descritos na ARS são amplamente utilizados no estudo de redes complexas, e vice-versa. As pesquisas em redes complexas cada vez mais revelam o quão as ideias da complexidade estão conectadas, isto é, fenômenos na Biologia, na Ciência da Computação, na Sociologia e na Física podem estar intimamente conectados a uma explicação de origem comum, na estrutura de relacionamentos entre agentes interativos quaisquer (Buchanan, 2002).

A fim de lançar mão simultânea dos conceitos presentes nos campos de estudo das redes complexas e da análise de redes sociais, utiliza-se o termo Análise de Redes Complexas, no presente trabalho, para se referir ao conjunto dos conceitos empregados em ambas as correntes de pensamento.

Procedimentos metodológicos

Com características qualitativas e o emprego de uma abordagem exploratória e de prototipagem, a pesquisa concebeu um modelo de RI em dados abertos ligados, usando operações de análise de redes complexas para auxiliar na classificação e seleção da informação recuperada, bem como desenvolveu técnicas para geração de mapas conceituais para apresentar a informação recuperada. O mapa conceitual resultante apresenta mais facilmente relacionamentos entre os termos inicialmente propostos. Assim, o foco não é descobrir atributos individuais, definições ou explicações de um ou outro termo, mas apresentar conexões e novos termos que sejam relevantes no relacionamento dos termos de consulta do usuário. Isso tudo, é claro, dentro dos limites e escopo da base de conhecimento.

Os autores usaram módulos de *software* independentes, por meio dos quais foi realizado um teste piloto com o ciclo completo de RI, desde o recebimento de um conjunto arbitrário de termos de consulta, passando pela geração de redes de informação intermediárias, até a síntese do mapa conceitual resultante. Usando o método de inspeção visual, que é normalmente empregado em uma das etapas da análise de redes complexas (Nooy; Mrvar; Batagelj, 2011), e atuando de forma experimental no refinamento dos parâmetros de cálculo e transformação executados por cada um dos módulos de *software* independentes. A realização desse teste piloto forneceu resultados importantes para aferir os algoritmos empregados nos processos de ranqueamento e seleção da informação recuperada, além de servir de base para implementação de um protótipo.

As etapas de execução do protótipo automatizaram quase por completo o modelo, que será apresentado na próxima seção. Ele permitiu a realização de maior número de testes; a descoberta de outros elementos de análise de redes complexas; a inclusão de mais iterações no algoritmo que faz a retroalimentação (resultados de uma determinada etapa do algoritmo servem de entrada para a próxima) para reiteradas expansões e reduções

da rede, até a obtenção do mapa conceitual resultante; flexibilização da quantidade de termos fornecidos pelo usuário; e aumento da modificabilidade, para permitir reconfigurações ágeis do algoritmo. O protótipo também trouxe aprimoramentos adicionais, como a inserção de heurísticas para melhoria da leitura do mapa conceitual resultante, e a possibilidade de validação com um grupo de usuários, apresentado mais adiante.

À luz da equação fundamental da CI de Brookes (1980) e dos resultados obtidos nos experimentos do teste piloto, foi concebido um modelo de RI que, em seguida, embasou a construção da primeira versão do protótipo. A execução exploratória dessa versão do protótipo sugeriu modificações no modelo que, em seguida, provocou novas alterações no protótipo. Esse processo exploratório e cíclico, de desenvolvimento do modelo de RI e protótipo, continuou por quase dois anos de pesquisa até se atingir um ponto de equilíbrio. Esse método exploratório foi determinante para a descoberta de parâmetros de cálculos e transformações de análise de redes que melhorassem empiricamente a síntese do mapa conceitual resultante.

O protótipo foi então validado por um grupo de 17 usuários. Tendo em vista a necessidade em avaliar a informação recuperada, os usuários foram escolhidos em função de sua familiarização com o uso de mapas conceituais e conhecimento do assunto no entorno dos termos da busca. De fato, Novak e Gowin (1984) sinalizam sobre a importância da preparação dos aprendizes para lidarem de forma adequada com os mapas conceituais, e Hjørland (2010) alerta que a determinação da relevância de uma informação é fortemente dependente do seu conhecimento.

A interface dos usuários com o protótipo se deu por intermédio de troca de mensagens por *e-mail* com um dos autores. Inicialmente os usuários forneciam dois conjuntos de termos semanticamente independentes, com três e seis elementos, e relacionados a uma hipotética necessidade informacional. Uma vez que os usuários eram todos falantes de língua portuguesa, mas o protótipo empregava uma base de conhecimento de dados abertos ligados em inglês, um dos autores atuou na tradução, tanto na formulação da consulta quanto na apresentação dos resultados. A base de dados abertos ligados escolhida foi a DBpedia, que representa um esforço comunitário para extrair informações estruturadas da enciclopédia Wikipedia <<https://www.wikipedia.org/>> e torná-las disponíveis na *Web*, permitindo sofisticadas consultas (Auer *et al.*, 2007). A sua base de conhecimento, além de cobrir uma grande quantidade de áreas, é amplamente usada pela comunidade de pesquisa e por diversas aplicações (Lehmann *et al.*, 2015). Apesar de disponível em vários idiomas, a escolha da sua versão em inglês para o experimento foi devida à quantidade muito maior de triplas RDF em comparação ao português.

Os usuários fizeram 47 avaliações de mapas conceituais cujos termos e relacionamentos foram gerados pelo protótipo, respondendo a um questionário enviado juntamente com a resposta à consulta por eles formulada. Inicialmente os usuários avaliaram o quanto o mapa conceitual resultante os auxiliava: (1) no entendimento das relações entre os termos de consulta, (2) como ponto de partida para uma pesquisa sobre relações com os termos de consulta, e (3) para construir um mapa conceitual mais completo. Foram também avaliadas a relevância dos novos conceitos introduzidos na informação recuperada, intermediários entre os termos enviados na consulta, bem como a relevância das proposições presentes no mapa (triplas). Finalmente, os usuários avaliaram a completude do mapa, indicando proposições fundamentais (outras triplas) que deveriam ter sido recuperadas. A partir dessas últimas avaliações foi feita uma estimativa preliminar, para o protótipo, dos seguintes indicadores de qualidade na recuperação da informação: (1) a precisão conceitual, correspondente à razão entre a quantidade dos novos conceitos relevantes e recuperados, sobre todos os novos conceitos recuperados; (2) a precisão proposicional, correspondente à razão entre a quantidade das proposições relevantes e recuperadas sobre todas as proposições recuperadas; e (3) a revocação proposicional, correspondente à razão entre a quantidade de proposições relevantes e recuperadas sobre todas as proposições relevantes indicadas por cada usuário. Os índices calculados, embora positivos, ainda não possuem significância estatística.

As principais tecnologias empregadas no teste piloto e na construção do protótipo foram: RDF (modelo de dados da base de dados ligados), SPARQL (linguagem de consulta para a base de dados ligados), SNORQL (terminal

para execução de consultas SPARQL), Gephi (usado na análise e inspeção visual das redes), *Semantic Web Import* (software usado em conjunto com o Gephi para transformar dados ligados em rede), GraphStream e GephiToolkit (bibliotecas com algoritmos de redes complexas), CmapTools (software usado para apresentação do mapa conceitual resultante), Java (linguagem de programação usada na construção do protótipo), Javacc (biblioteca de apoio para geração dos analisadores das entradas de dados), Jena (biblioteca para lidar com bases de dados ligados), JSON (formato para intercâmbio de dados entre a base de dados ligados e a biblioteca Jena), Eclipse (ambiente integrado usado no desenvolvimento do protótipo), Github (controle de versões usado no desenvolvimento do protótipo) e Egit (ferramenta para integração do Eclipse e Github).

As métricas de redes complexas empregadas foram: *Betweenness* (centralidade de intermediação dos conceitos, em relação aos demais), *Closeness* (centralidade de proximidade dos conceitos, em relação aos demais), *Eigenvector* (centralidade de autovetor dos conceitos, em relação aos demais), *K-core* (um indicador da coesão entre conceitos), *Eccentricity* (indicador do quão distante cada conceito encontra-se dos demais), *Shortestpath* (indicador de quão próximo cada conceito encontra-se dos demais), Componente Conectado (representa uma sub-rede independente do restante da rede) e *Giant Component* (um componente conectado numa rede proporcionalmente muito maior que os demais).

A apresentação detalhada dessas tecnologias e métricas de rede está fora do escopo desse artigo. Detalhes podem ser buscados em Cristovão (2016).

Resultados

A visão geral do modelo desenvolvido pela pesquisa é apresentada no diagrama da Figura 1. O diagrama descreve o fluxo informacional, que se inicia com o usuário fornecendo um conjunto de termos de busca. Em seguida, esses termos são reescritos como consultas para *Linked Open Data* (LOD) baseados numa consulta modelo. A execução dessas consultas sobre a base LOD recupera um conjunto de triplas RDF resultantes, que passam por um mapeamento, transformando-se numa rede de informação, normalmente com vários componentes conectados devido à distância semântica dos termos. É aplicada uma análise de redes complexas sobre a rede de informação para ranquear e selecionar nós em potencial, ou novos termos, para permitir a unificação dos vários componentes conectados. Se a rede ainda não possui um componente gigante que integre todos os termos do usuário, o fluxo do modelo retorna retroalimentando uma nova busca com os nós selecionados, tendo os novos RDF recuperados e mesclados na rede existente. Esse processo se repete enquanto o critério de unificação dos termos de consulta do usuário não for atendido (Figura 1).

Ao final dos ciclos de retroalimentação, seguindo o fluxo, obtém-se uma rede intermediária expandida e unificada. A partir desse ponto, inicia-se o processo de redução da rede, que é feito por uma nova análise de redes com a construção de nova rede contendo os nós selecionados nas iterações anteriores, mantendo os caminhos mínimos entre os termos de consulta em um componente gigante. Finalmente, o mapeamento da rede de informação final é feito para o mapa conceitual resultante, tendo o auxílio de um vocabulário controlado e aplicação de algumas heurísticas para aumento da legibilidade do mapa.

A Figura 2 mostra um exemplo de rede expandida após oito ciclos de retroalimentação e com alguns milhares de nós e conexões, advindos de uma consulta sobre os termos “Jean Piaget”, “Software educacional” e “Seymour Papert”. Essa rede, após redução, originou o mapa conceitual resultante mostrado na Figura 3, que foi um dos mapas avaliados na validação com usuários. As caixas de cor de fundo cinza do mapa conceitual representam os termos de consulta do usuário. Os outros conceitos do mapa, “Logo” (linguagem de programação) e “Categoria de software educacional livre”, representados em caixas de cor de fundo branca, possuem relacionamento com os termos de consulta do usuário que foram descobertos por intermédio de ranqueamentos e seleção no processamento do

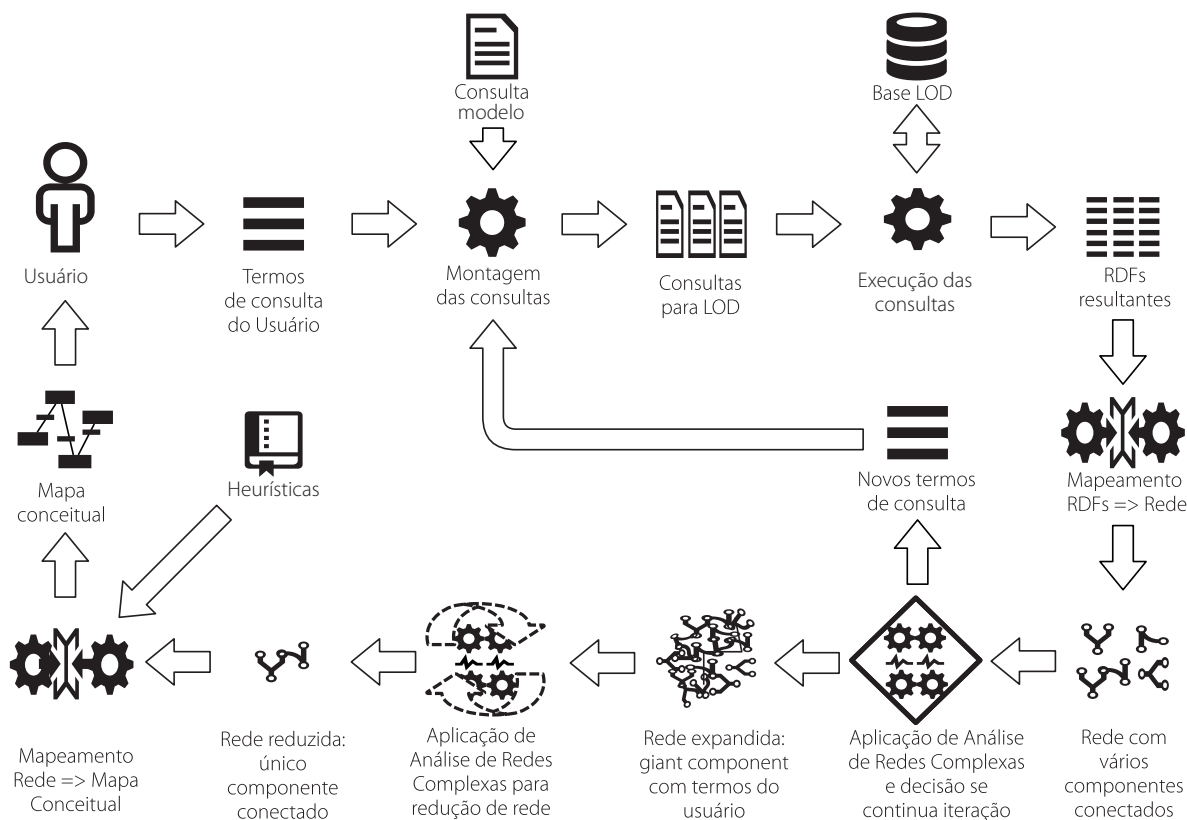


Figura 1. Diagrama geral do modelo de recuperação da informação em dados ligados.

Fonte: Elaborada pelos autores (2016).

modelo sobre a rede expandida mostrada na Figura 2. Observa-se ainda, na rede expandida, destaque para os nós “Educação” e “Tecnologia da informação”, que, apesar de possuírem as duas maiores quantidades de conexões, não foram selecionados para o mapa resultante, pois o algoritmo do modelo tem critérios que vão além de uma simples escolha como essa.

Os critérios usados para a seleção dos nós mais importantes, destacados na Figura 2 e resultantes no mapa conceitual final, representado na Figura 3, foram orientados pela criação de uma rede informacional formada por todos os nós mais bem ranqueados. Esse ranqueamento ocorreu por intermédio das métricas *Betweenness*, *Closeness* e *Eigenvector*, e também por todos os nós intermediários, que formam o *Shortestpath* e garantem a unificação da rede em um único componente conectado, *Giant Component*. Além disso, foram usados o algoritmo *K-core* e um *ranking* de nós com a métrica *Eccentricity* no auxílio da redução da rede, pela eliminação de nós menos relevantes, até a formação do mapa conceitual resultante. Detalhamento desses critérios e do algoritmo completo podem ser consultados em Cristovão (2016) (Figuras 2 e 3).

O protótipo desenvolvido, descrito na seção dos métodos, implementou quase por completo o modelo apresentado. Não foram implementados a parte inicial referente à interface com o usuário que cuida da entrada dos termos de consulta; uma das heurísticas que cuida do balanceamento dos tipos de conceitos finais do mapa, isto é, conceitos individuais e conceitos gerais; e o ajuste fino no *layout* do mapa conceitual resultante.

A avaliação dos usuários sobre o quanto o mapa conceitual resultante auxilia no entendimento das relações entre os termos de consulta, obteve 15 respostas “muito”, 16 “razoável”, 15 “pouco” e 1 “nenhum”. Sobre o quanto o

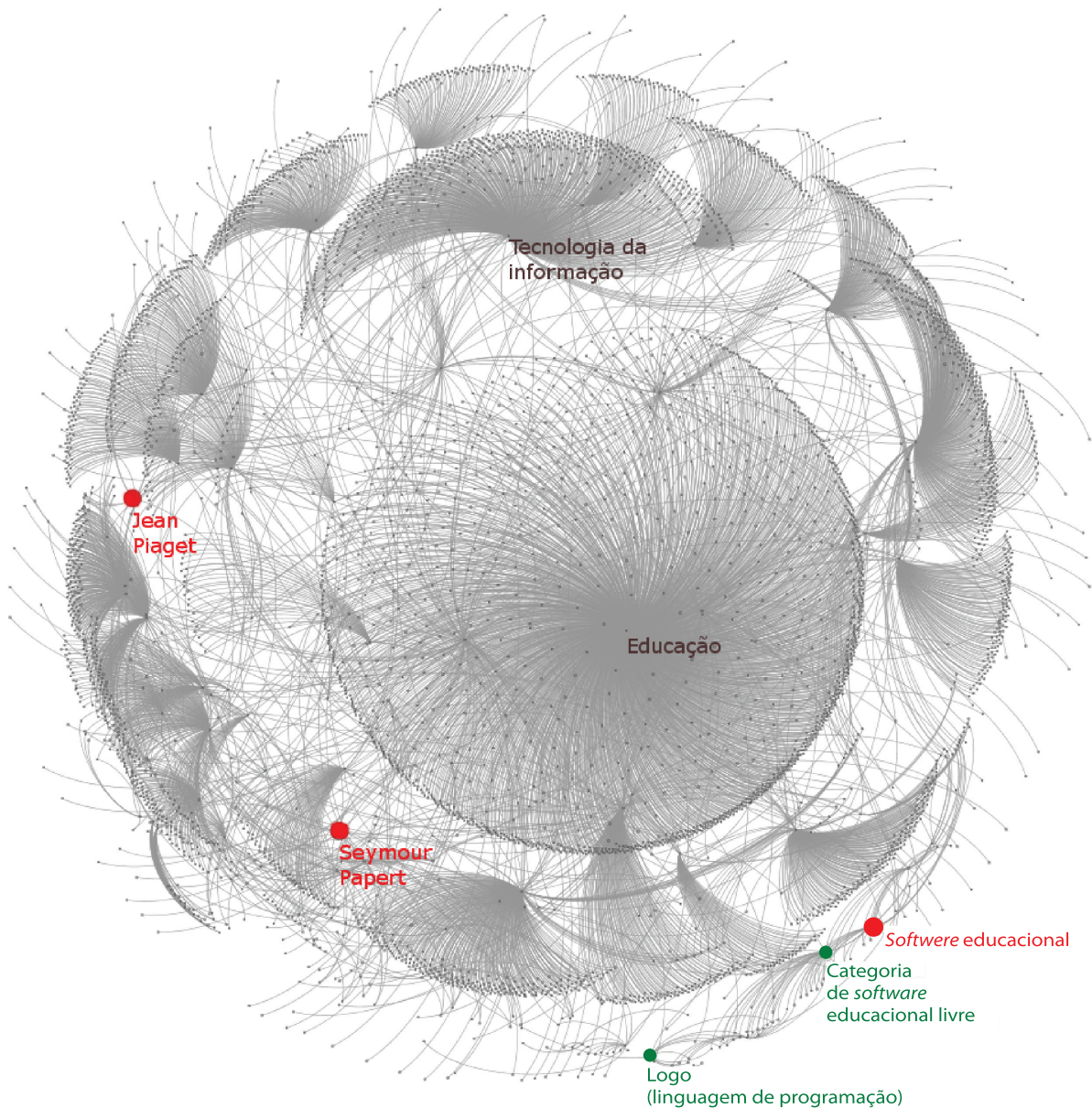


Figura 2. Rede de informação expandida com 4.285 nós e 4.909 conexões a partir dos termos de consulta “Jean Piaget”, “Software educacional” e “Seymour Papert”.

Fonte: Elaborada pelos autores (2016).

mapa auxilia como ponto de partida para uma pesquisa sobre relações com os termos de consulta, foram obtidas 17 respostas “muito”, 19 “razoável” e 11 “pouco”. Sobre o quanto o mapa auxilia para construir um mapa conceitual mais completo, foram obtidas 21 respostas “muito”, 16 “razoável” e 10 “pouco”.

A partir da indicação, pelos usuários, da quantidade de conceitos e proposições relevantes, e das proposições que faltaram, foram estimados preliminarmente os valores da precisão e revocação da informação recuperada. Ao todo foram 47 avaliações realizadas sobre 33 mapas conceituais distintos. Dentro desse universo, obteve-se a média de 65% de precisão nos novos conceitos e 64% nas proposições recuperadas. Fazendo-se uma distribuição

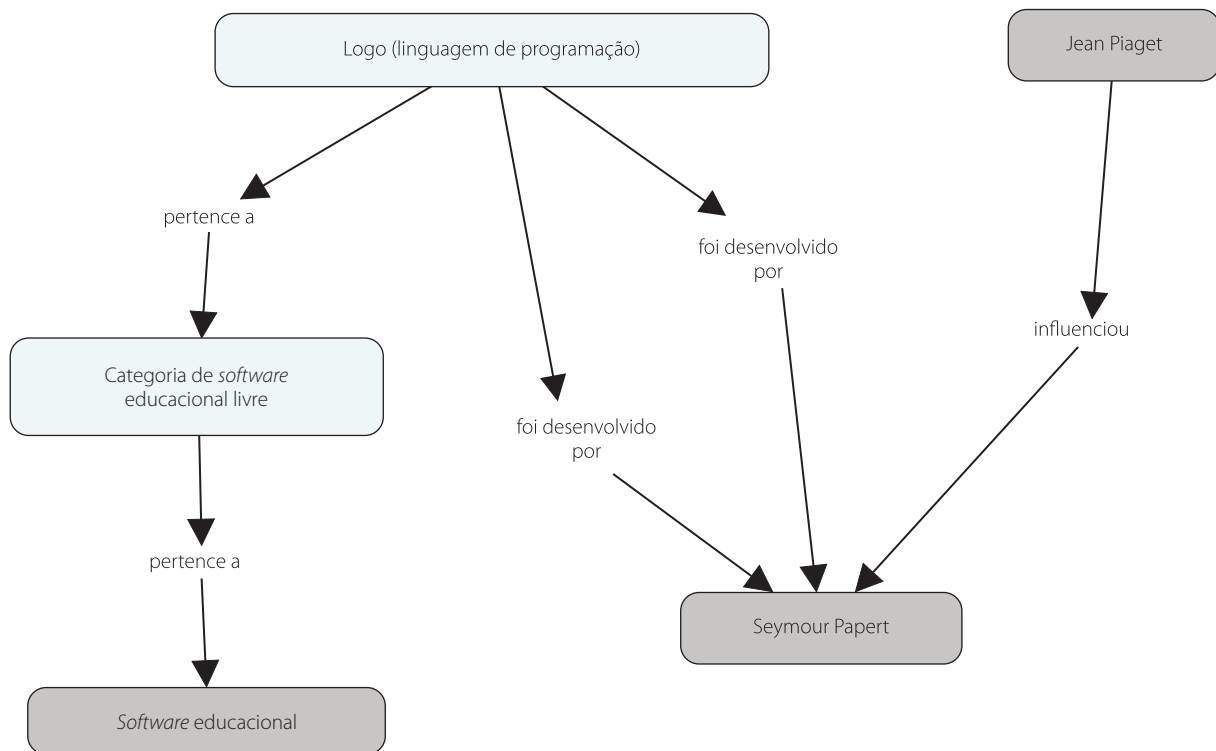


Figura 3. Mapa conceitual resultante dos termos “Jean Piaget”, “Software educacional” e “Seymour Papert”.

Fonte: Elaborada pelos autores (2016) por intermédio do protótipo desenvolvido.

de frequência para selecionar as avaliações com precisão acima de 60%, observou-se uma quantidade de 20 para novos conceitos e 23 para as proposições. E, para precisão entre 20% e 60%, foi identificada uma quantidade de 23 para novos conceitos e 21 para as proposições. Os usuários indicaram 139 proposições faltantes ao todo nas avaliações, porém apenas quatro existiam na base de conhecimento. Baseado nesse último valor e na quantidade de proposições relevantes, a média de todos os valores de revocação das proposições recuperadas foi de 99%. A revocação de cada avaliação foi calculada como a razão da quantidade de todas as proposições recuperadas e relevantes pela quantidade total de proposições relevantes. Esta última foi calculada pela soma da quantidade de todas as proposições recuperadas, com a quantidade de proposições indicadas pelos usuários como relevantes e que não foram recuperadas, apesar de disponíveis na base de conhecimento.

Discussão

O Quadro 1 apresenta duas interpretações da equação de Brookes (1980), $K[S]+\Delta I=K[S+\Delta S]$, no sistema de RI. Na abordagem da equação enquanto interface do usuário com o sistema (segunda coluna do Quadro 1), a estrutura cognitiva do usuário $K[S]$ é modificada quando se relaciona com o mapa conceitual resultante ΔI , provocando uma alteração no estado do usuário ΔS e, conseqüentemente, na sua estrutura cognitiva, que passa a ser representada por $K[S+\Delta S]$.

Na abordagem da equação enquanto núcleo do sistema (terceira coluna do Quadro 1), a rede de informação $K[S]$ formada pelas triplas RDF recuperadas na base de conhecimento e oriundas dos termos de consulta do usuário S , é mesclada com a rede de informação ΔI , formada pelos nós e ligações capazes de unificar todos os termos de consulta do usuário. Em seguida, essa nova rede $K[S]+\Delta I$ passa por um processo de transformação que provoca sua

Quadro 1. Interpretação da equação de Brookes no sistema de Recuperação de Informação.

Elementos Equação	Abordagem enquanto interface do usuário com o sistema	Abordagem enquanto núcleo do sistema
K	Estrutura cognitiva do usuário.	Base de dados ligados.
S	Estado cognitivo prévio do usuário.	Termos de consulta do usuário.
K[S]	Estrutura cognitiva prévia do usuário.	Rede de informação formada pelas triplas RDFs recuperadas na base de conhecimento e derivadas dos termos de consulta do usuário S.
ΔI	Mapa conceitual resultante recebido pelo usuário.	Rede de informação com os nós e ligações capazes de unificar todos os termos de consulta do usuário da rede K[S].
$K[S] + \Delta I$	Relacionamento entre a estrutura cognitiva prévia do usuário K[S] e o mapa conceitual resultante ΔI , que impulsiona a geração de uma estrutura cognitiva modificada.	Mesclagem da rede K[S] com a rede ΔI . Retorna a rede de informação expandida, com único componente conectado.
=	Transformação cognitiva, significando a geração de um novo estado mental a partir do estado prévio.	Transformação da rede, significando o impulso da rede expandida $K[S]+\Delta I$ para a geração de uma rede de informação reduzida (mapa conceitual) $K[S+\Delta S]$.
ΔS	Modificação do estado cognitivo do usuário durante a interação com o sistema.	Novos conceitos selecionados para a rede de informação reduzida (mapa conceitual).
$S+\Delta S$	Novo estado cognitivo do usuário, após a interação com o sistema.	Conjunto de todos os conceitos do mapa resultante (termos de consulta do usuário S e novos conceitos selecionados ΔS).
$K[S+\Delta S]$	Nova estrutura cognitiva do usuário, após o recebimento do mapa conceitual resultante, ΔI .	Rede de informação reduzida, mapeada no mapa conceitual resultante.

Fonte: Elaborado pelos autores (2016).

redução por intermédio de algoritmos de ranqueamento e seleção em redes complexas até a formação do mapa conceitual resultante $K[S+\Delta S]$, que pode ser interpretado como uma rede de informação formada pelos termos de consulta do usuário S e os novos termos ΔS enquanto nós, e as ligações entre eles.

Quanto à validação do modelo, observaram-se melhores avaliações nos aspectos relacionados à continuidade do processo de pesquisa e como auxílio à continuidade de sua construção. Esse resultado pode estar relacionado à tendência do uso da informação recuperada como ponto de partida para continuidade da aprendizagem sobre os relacionamentos existentes entre os termos de consulta, tal como preconiza a equação de Brookes (1980). Isto é, as estruturas cognitivas do usuário K[S] serão modificadas à medida que ele se relaciona com a nova informação recebida ΔI , isto é, o mapa conceitual resultante, formando assim uma nova estrutura de conhecimento $K[S+\Delta S]$ em função da modificação do seu estado ΔS .

A precisão preliminarmente estimada a partir das indicações de relevância dos usuários foi mediana, com 65% para os novos conceitos e 64% para as proposições. É importante destacar que o fato de os usuários já conhecerem os assuntos relacionados aos termos de consulta leva-os a uma expectativa alta quanto à relevância da informação recuperada dentro do que eles já conhecem. Além disso, a indicação de proposições faltantes pode também induzir a uma insatisfação dos usuários, pois eles não tinham conhecimento de que apenas uma quantidade muito pequena delas, quatro das 139 indicadas, de fato existiam na base e não foram recuperadas pelo sistema. Apesar disso, o cálculo da revocação das proposições recuperadas teve um bom índice (99%), ainda que a confiabilidade estatística dos dados não tenha sido estimada.

Assim, devido ao grande número de proposições sinalizadas pelos usuários como relevantes, porém, inexistentes na base de conhecimento DBpedia, é possível inferir que ainda há grande demanda de crescimento para essa base. Outro fator que pode ter influenciado nos resultados é o fato de os usuários não terem utilizado

diretamente o protótipo, pois eles forneciam a um dos pesquisadores os conjuntos de termos, que depois eram traduzidos do português para o inglês e executados no protótipo. Finalizado o processo, o mapa conceitual resultante era traduzido do inglês para o português e entregue ao usuário para avaliação.

Com base em pesquisa bibliográfica junto ao Portal de Periódicos da Capes, foram identificados, analisados e comparados trabalhos que atendiam aos critérios em comum com o presente trabalho. O Quadro 2 apresenta uma síntese comparativa, onde as colunas, numeradas de 1 a 13, representam o atendimento aos critérios usados na comparação, apresentados a seguir, com indicação, entre parênteses, dos principais termos utilizados na identificação, análise e comparação de trabalhos correlatos. Foram usadas combinações dos seguintes termos de consulta, para seleção dos trabalhos: "*information retrieval*", "*knowledge retrieval*", "*text retrieval*", "*semantic web*", "*linked data*", "*complex network*", "*network analysis*", "*concept map*", "*relationship concept*", "*concept map generation*", "*information visualization*", "*knowledge visualization*", e seus correspondentes na língua portuguesa. A partir dessa seleção, foi feita uma verificação analítica em cada um deles e descartados aqueles com menos de quatro indicações de similaridade, conforme atendessem aos seguintes critérios:

- 1) Recupera ou extrai informações.
- 2) Enquanto método de RI, tem a recuperação de texto como forma predominante.
- 3) Aceita como ponto de partida uma lista de termos textuais fornecidos pelo usuário.
- 4) Usa dados ligados da *web* semântica como base de conhecimento.
- 5) Tem como foco a descoberta de relacionamentos existentes entre os termos fornecidos pelo usuário.
- 6) Revela relacionamentos intermediários entre os termos da busca, ainda que estes estejam distantes por alguns nós e ligações ao longo da rede informacional.
- 7) Usa métrica de rede e algoritmos de grafos sobre a rede informacional como parte fundamental para o ranqueamento e seleção de documentos relevantes.
- 8) Considera a topologia da rede informacional para o ranqueamento dos documentos.

Quadro 2. Comparação entre o modelo proposto e os trabalhos correlatos.

Referência para o trabalho	Critérios comparativos												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Modelo proposto no presente trabalho	X	X	X	X	X	X	X	X	X	X	X	X	X
Cañas <i>et al.</i> (2004)	X	X	X							X	X	X	
Lima (2005)	X	X								X	X	X	
Thammasut e Sornil (2006)	X	X					X	X					
Graudina e Grundspenkis (2008)	X									X	X	X	X
Truong <i>et al.</i> (2008)	X	X					X	X					
Heim, Ertl e Ziegler (2010)	X	X	X	X	X								
Lohmann <i>et al.</i> (2010)	X	X	X	X	X	X			X	X			
Guéret <i>et al.</i> (2012)	X			X			X	X					
Valerio, Leake e Cañas (2012)	X	X								X	X	X	X
Paulheim (2013)	X	X	X	X									
McLinden (2013)	X						X	X		X	X	X	X
Cury, Perin e Santos Junior (2014)	X		X							X	X	X	X
Usbeck (2014)	X	X	X	X			X						

Fonte: Elaborado pelos autores (2016).

9) Usa um processo de retroalimentação que, mediante uma única solicitação do usuário, realiza novas buscas na base de conhecimento a partir de documentos já recuperados.

10) Usa um formato visual de rede informacional para apresentar a informação resultante.

11) Usa mapa conceitual para representar a informação resultante.

12) Parte da informação recuperada é apresentada ao usuário como conceitos de um mapa conceitual gerado de forma semiautomática.

13) Parte da informação recuperada é apresentada ao usuário como relações entre conceitos de um mapa conceitual gerado de forma semiautomática.

O trabalho que mais se aproximou da presente proposta é o de Lohmann *et al.* (2010), com oito indicações de similaridade das treze totais. A proposta dos autores consegue, a partir de uma lista de termos do usuário, descobrir relacionamentos entre eles, apresentando-os num formato de rede informacional próximo a um grafo de RDF. Existe uma interface interativa adequada que permite grande flexibilidade nas consultas. Além disso, existe uma retroalimentação que consegue fazer a rede crescer à medida que novos relacionamentos vão sendo descobertos. Contudo, após teste da ferramenta <<http://www.visualdataweb.org/refinder.php>>, observou-se que a descoberta de relacionamentos é mais indicada para aqueles com conexão direta, pois vários testes realizados entre termos que se conectavam apenas indiretamente falharam. Adicionalmente, o trabalho não atende aos critérios de empregar mapas conceituais para apresentação de resultados.

A originalidade do presente trabalho concentra-se, principalmente, no conjunto de elementos de diversas áreas do conhecimento para a construção do modelo de RI em dados ligados, tais como o uso de conhecimentos em redes complexas e mapas conceituais. Os pontos específicos que mais se destacam são o uso de métricas de rede, algoritmos de grafos e análise topológica sobre a rede informacional como parte fundamental para o ranqueamento e seleção dos nós mais relevantes, e a apresentação da informação recuperada no formato de um mapa conceitual.

Conclusão

Um dos maiores desafios deste trabalho foi ter que lidar com um desenvolvimento integrado em várias áreas do conhecimento, em direção ao cumprimento dos seus objetivos. Com foco na busca pelas relações existentes entre os termos fornecidos pelo usuário, e não pela via mais tradicional com buscas de propriedades, definições ou explicações individuais, o presente trabalho abriu possibilidades concretas na RI com o uso de elementos da Ciência das Redes e mapas conceituais no contexto dos dados abertos ligados na *web* semântica. Essa busca por relações abre um leque de possibilidades interessantes em várias áreas do conhecimento e na disponibilização de serviços para a sociedade. Por exemplo, cidadãos podem usufruir de um serviço com essas características para encontrar relacionamentos em informações governamentais, de forma a terem uma postura mais ativa quanto ao acompanhamento de dados num contexto de transparência governamental e combate à corrupção.

Apesar de indícios de um resultado razoável na recuperação da informação, segundo avaliações dos usuários, a realização de validação com confiabilidade estatística é dependente de um quantitativo maior de usuários e de bases de conhecimento diferentes, como as *Linked Open Government Data*. Bases brasileiras, tais como o Portal da Transparência, também podem ser experimentadas. Além disso, a adoção de uma RI interativa no modelo poderia oferecer ao usuário maior flexibilidade na escolha final dos conceitos, isto é, ao invés do recebimento de um mapa conceitual pronto, o usuário interagiria durante o processo de ranqueamento e seleção dos conceitos, podendo voltar quantas vezes necessárias até a formação de um mapa conceitual mais próximo do seu desejo informacional.

Colaboradores

Todos os autores contribuíram na concepção e desenho do estudo, análise de dados e redação final.

Referências

- Araújo, E. A. Equação do impacto informacional: uma proposta paradigmática. In: Encontro Nacional de Pesquisa em Ciência da Informação, 5., 2003, Belo Horizonte. *Anais eletrônicos...* Belo Horizonte: UFMG, 2003. Disponível em: <<http://enancib.ibict.br/index.php/enancib/venancib/paper/view/2125/1260>>. Acesso em: 6 abr. 2016.
- Auer, S. *et al.* DBpedia: A nucleus for a web of open data. In: Aberer, K. *et al.* (Org.). *The Semantic Web*. [Berlin]: Springer Berlin Heidelberg, 2007. p. 722-735. (Lecture Notes in Computer Science). Available from: <http://link.springer.com/chapter/10.1007/978-3-540-76298-0_52>. Cited: Feb. 21, 2016.
- Auer, S. *et al.* Introduction to linked data and its lifecycle on the web. In: Rudolph, S. *et al.* (Org.). *Reasoning web: Semantic technologies for intelligent data access*. Berlin: Springer Berlin Heidelberg, 2013. p. 1-90. (Lecture Notes in Computer Science). Available from: <http://link.springer.com/chapter/10.1007/978-3-642-39784-4_1>. Cited: Feb. 19, 2016.
- Baeza-Yates, R.; Ribeiro-Neto, B. *Modern information retrieval: The concepts and technology behind search*. 2nd ed. New York: Addison-Wesley, 2011.
- Barabási, A. L. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 371, n. 1987, p. 20120375–20120375, 2013. Available from: <<http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2012.0375>>. Cited: Feb. 15, 2016.
- Batista, F. F.; Costa, S. M. S.; Alvares, L. M. A. R. Gestão do conhecimento: a realização da proposta de Brookes para a Ciência da Informação? In: Encontro Nacional de Pesquisa em Ciência da Informação, 8., 2007, Salvador. *Anais eletrônicos...* Salvador: Ancib, 2007. Disponível em: <<http://repositorio.unb.br/handle/10482/1006>>. Acesso em: 30 mar. 2016.
- Bauer, F.; Kaltenböck, M. *Linked Open Data: The essentials. A quick start guide for decision makers*. Vienna: Edition mono/monochrom, 2012. Available from: <<https://www.reeep.org/LOD-the-Essentials.pdf>>. Cited: Oct. 12, 2015.
- Bawden, D. Brookes equation: The basis for a qualitative characterization of information behaviours. *Journal of Information Science*, v. 37, n. 1, p. 101-108, 2011. Available from: <<http://openaccess.city.ac.uk/3130/>>. Cited: Mar. 29, 2016.
- Berners-Lee, T. Long live the web: A call for continued open standards and neutrality. *Scientific American*, v. 303, n. 6, 2010. Available from: <<http://www.scientificamerican.com/article/long-live-the-web/>>. Cited: Feb. 22, 2016.
- Berners-Lee, T. *et al.* The semantic web. *Scientific American*, v. 284, n. 5, p. 28-37, 2001. Available from: <http://iscl2918929391.googlecode.com/svn-history/r347/trunk/RPC/Slides/p01_theSemanticWeb.pdf>. Cited: Feb. 19, 2016.
- Brookes, B. C. The foundations of information science: Part I: Philosophical aspects. *Journal of Information Science*, v. 2, n. 3-4, p. 125-133, 1980. Available from: <<http://jis.sagepub.com/content/2/3-4/125>>. Cited: Mar. 29, 2016.
- Buchanan, M. *Nexus: Small worlds and the groundbreaking science of networks*. New York: WW Norton and Company, 2002.
- Cañas, A. J. *et al.* Mining the web to suggest concepts during concept map construction. In: International Conference on Concept Mapping, 1st., 2004, Pamplona. *Electronic proceedings...* Pamplona: Dirección de Publicaciones de la Universidad Pública de Navarra, 2004. Available from: <<http://eprint.ihmc.us/91/1/cmc2004-284.pdf>>. Cited: Mar. 4, 2016.
- Cristovão, H. M. *Um modelo híbrido de recuperação de informação e conhecimento baseado na síntese de mapas conceituais obtidos por operações de transformação de redes complexas orientadas por busca de relacionamentos entre termos de consulta em bases de dados ligados*. 2016. 320 f. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, 2016. Disponível em: <<http://repositorio.unb.br/handle/10482/22284>>. Acesso em: 28 jan. 2017.
- Cury, D.; Perin, W. A.; Santos Junior, I. A. M. CMPAAS: A platform of services for construction and handling of concept maps. In: International Conference on Concept Mapping, 6th., 2014, Santos. *Electronic proceedings...* Santos: USP, 2014. p. 107-115. Available from: <<http://cmc.ihmc.us/cmc2014Program.html>>. Cited: Apr. 3, 2016.
- Graudina, V.; Grundspenkis, J. Concept map generation from OWL ontologies. In: International Conference on Concept Mapping, 3rd., 2008, Tallinn. *Electronic proceedings...* Tallinn: Tallinn University, 2008. p. 263-270. Available from: <<http://cmc.ihmc.us/cmc2008papers/cmc2008-p263.pdf>>. Cited: Apr. 27, 2016.
- Guéret, C. *et al.* Assessing linked data mappings using network measures. In: International Conference on The Semantic Web: Research and Applications, 9th., Berlin. *Electronic proceedings...* Berlin: Springer-Verlag, 2012. p. 87-102. http://dx.doi.org/10.1007/978-3-642-30284-8_13
- Heim, P.; Ertl, T.; Ziegler, J. Facet Graphs: Complex semantic querying made easy. In: Aroyo, L. *et al.* (Org.). *The semantic web: Research and applications*. Berlin: Springer Berlin Heidelberg, 2010. p. 288-302. (Lecture Notes in Computer Science). Available from: <http://www.sfb716.uni-stuttgart.de/uploads/tx_vispublications/eswc10-heimErtlZiegler.pdf>. Cited: May 4, 2016.
- Hjørland, B. The foundation of the concept of relevance. *Journal of the American Society for Information Science and Technology*, v. 61, n. 2, p. 217-237, 2010. Available from: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.21261/abstract>>. Cited: Feb. 16, 2016.

Le Coadic, Y. F. *A Ciência da informação*. Brasília: Briquet de Lemos Livros, 1996.

Lehmann, J. *et al.* DBpedia: A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web Journal*, v. 6, n. 2, p. 167-195, 2015. Available from: <http://jens-lehmann.org/files/2014/swj_dbpedia.pdf>. Cited: Feb. 21, 2016.

Lima, G. A. B. O. Mapa conceitual como ferramenta para organização do conhecimento em sistema de hipertextos e seus aspectos cognitivos. *Perspectivas em Ciência da Informação*, v. 9, n. 2, p. 134-145, 2004. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/355>>. Acesso em: 9 mar. 2016.

Lima, G. A. B. O. Modelo hipertextual-MHTX: um modelo para organização hipertextual de documentos. In: Encontro Nacional de Pesquisa em Ciência da Informação, 6., 2005, Florianópolis. *Anais eletrônicos...* Florianópolis: IBICT, 2005. Disponível em: <<http://enancib.ibict.br/index.php/enancib/vienancib/schedConf/presentations>>. Acesso em: 9 mar. 2016.

Lohmann, S. *et al.* The RelFinder user interface: Interactive exploration of relationships between objects of interest. International Conference on Intelligent User Interfaces, 15th., 2010, New York. *Proceedings...* New York: ACM, 2010. p. 421-422. <http://doi.org/10.1145/1719970.1720052>

McLinden, D. Concept maps as network data: Analysis of a concept map using the methods of social network analysis. *Evaluation and Program Planning*, v. 36, n. 1, p. 40-48, 2013. Available from: <<http://www.sciencedirect.com/science/article/pii/S0149718912000456>>. Cited: Feb. 25, 2016.

Mika, P. *Social networks and the semantic web*. Boston: Springer, 2007. (Semantic Web and Beyond, v. 5). Available from: <<http://link.springer.com/10.1007/978-0-387-71001-3>>. Cited: Feb. 22, 2016.

Moraes, M. B. A Ciência da Informação nos caminhos do contemporâneo. *PontodeAcesso*, v. 7, n. 2, p. 2-24, 2013. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/5199>>. Acesso em: 30 mar. 2016.

Nascimento, D. M. A abordagem sócio-cultural da informação. *Informação e Sociedade: Estudos*, v. 16, n. 2, p. 25-35, 2006. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/477>>. Acesso em: 30 mar. 2016.

Neill, S. D. Brookes, Popper, and objective knowledge. *Journal of Information Science*, v. 4, n. 1, p. 33-39, 1982. Available from: <<http://jjs.sagepub.com/content/4/1/33>>. Cited: Apr. 7, 2016.

Newman, M. E. J. *Networks: An introduction*. Oxford: Oxford University Press, 2010.

Nooy, W.; Mrvar, A.; Batagelj, V. Exploratory social network analysis with Pajek. 2nd ed. rev. and expanded. Cambridge: Cambridge University Press, 2011. (Structural Analysis in the Social Sciences, 34).

Novak, J. D. *A theory of education*. Ithaca: Cornell University Press, 1977. Available from: <<http://catalog.hathitrust.org/Record/000252496>>. Cited: Feb. 21, 2016.

Novak, J. D.; Cañas, A. J. The universality and ubiquitousness of concept maps. In: International Conference on Concept

Mapping, 4th., 2010, Viña del Mar. *Electronic proceedings...Viña del Mar*: Lom Ediciones, 2010. Available from: <<http://cmc.ihmc.us/cmc/CMCProceedings.html>>. Cited: Feb. 21, 2016.

Novak, J. D.; Gowin, D. B. *Learning how to learn*. Cambridge: Cambridge University Press, 1984.

Open Definition. *Open Knowledge: Source Code*. 2.1. [S.l.: s.n.], 2015. Available from: <<http://opendefinition.org/>>. Cited: Feb. 21, 2016.

Orrantia, J. S. Conocity: videos enriquecidos con mapas para la gestión del conocimiento. In: International Conference on Concept Mapping, 5., 2012, Valleta. *Anais eletrônicos...* Valleta: University of Malta, 2012. Disponível em: <<http://cmc.ihmc.us/cmc2012/CMC2012Program.html>>. Acesso em: 22 feb. 2016.

Paulheim, H. Exploiting linked open data as background knowledge in data mining. In: International Workshop on Data Mining on Linked Data, 2013, Prague. *Electronic proceedings...* Prague: CEUR, 2013. Available from: <<http://ceur-ws.org/Vol-1082/extendedAbstract.pdf>>. Cited: Feb. 24, 2016.

Pereira, F. C. M. A equação fundamental da Ciência da Informação e a importância de Brookes enquanto referência para o campo da Ciência da Informação. *Informação e Informação*, v. 13, n. 1, p. 15-31, 2008. Disponível em: <<http://www.uel.br/revistas/wrevojs246/index.php/informacao/article/view/1761>>. Acesso em: 30 mar. 2016.

Pontes Junior, J.; Carvalho, R. A.; Azevedo, A. W. Da recuperação da informação à recuperação do conhecimento: reflexões e propostas. *Perspectivas em Ciência da Informação*, v. 18, n. 4, p. 2-17, 2013. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/965>>. Acesso em: 29 fev. 2016.

Robredo, J. *Da Ciência da Informação revisitada aos sistemas humanos de informação*. Brasília: Thesaurus, 2003.

Santos Neto, A. L. *et al.* Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. *Transinformação*, v. 25, n. 1, p. 81-87, 2013. Disponível em: <<http://www.scielo.br/pdf/tinf/v25n1/a08v25n1.pdf>>. Acesso em: 30 maio 2016. <http://dx.doi.org/10.1590/S0103-37862013000100008>

Saracevic, T. *Information science: Encyclopedia of Library and Information Sciences*. 3rd ed. New York: Taylor and Francis, 2010. p. 2570-2586. Available from: <<http://cominfo.rutgers.edu/~tefko/SaracevicInformationScienceELIS2009.pdf>>. Cited: Feb. 16, 2016.

Shadbolt, N. *et al.* Web science: A new frontier. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 371, n. 1987, p. 20120512-20120512, 2013. Available from: <<http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2012.0512>>. Cited: Feb. 19, 2016.

Stuckenschmidt, H. Data semantics on the web. *Journal on Data Semantics*, v. 1, n. 1, p. 1-9, 2012. Available from: <<http://link.springer.com/article/10.1007/s13740-012-0003-z>>. Cited: Feb. 17, 2016.

Thammasut, D.; Sornil, O. A graph-based information retrieval system. In: International Symposium on Communications and Information Technologies, 2006, Ladkrabang. *Electronic*

proceedings... Ladrabang: IEEE, 2006. p. 743-748. Available from: <<http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4141327>>. Cited: Apr. 14, 2016.

Todd, R. J. Back to our beginnings: Information utilization, Bertram Brookes and the fundamental equation of information science. *Information Processing and Management*, v. 35, n. 6, p. 851-870, 1999. Available from: <<http://www.sciencedirect.com/science/article/pii/S0306457399000308>>. Cited: Mar. 29, 2016.

Truong, Q. D. *et al.* Information retrieval model based on graph comparison. In: Journées Internationales D'Analyse Statistique des Données Textuelles (JADT), 2008, Lyon. *Electronic proceedings...* Lyon: Laboratoire, 2008. Available from: <http://www.irit.fr/publis/SIG/2008_JADT_TDMC.pdf>. Cited: Apr. 14, 2016.

Usbeck, R. Combining linked data and dtatistical information retrieval. In: European Semantic Web Conference, Lecture Notes in Computer Science, 05., 2014, Anissaras. *Electronic proceedings...* Cham: Springer, 2014. p. 845-854. Available from: <https://link.springer.com/chapter/10.1007/978-3-319-07443-6_58>. Cited: Apr. 14, 2016.

Valerio, A.; Leake, D. B.; Cañas, A. J. Using automatically generated concept maps for document understanding: A human subjects experiment. In: International Conference on Concept Mapping, 5th., 2012, Valleta. *Electronic proceedings...* Valleta: University of Malta, 2012. Available from: <<http://cmc.ihmc.us/cmc/CMCProceedings.html>>. Cited: Feb. 25, 2016.

Vekiri, I. What is the value of graphical displays in learning? *Educational Psychology Review*, v. 14, n. 3, p. 261-312, 2002. Available from: <<http://link.springer.com/article/10.1023/A:1016064429161>>. Cited: Feb. 21, 2016.

Wasserman, S.; Faust, K. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press, 1994.

Wersig, G.; Neveling, U. The phenomena of interest to information science. *The Information Scientist*, v. 9, n. 4, p. 127-140, 1975. Available from: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.232.5319&rep=rep1&type=pdf>>. Cited: Feb. 16, 2016.

Zhang, J. *Visualization for information retrieval*. Berlin: Springer, 2008. (The Information Retrieval Series).