

VISUALIZAÇÃO DA INFORMAÇÃO NA CONSTRUÇÃO DE INTERFACES AMIGÁVEIS PARA SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO.

Jessica Monique de Lira Vieiraⁱ
Renato Fernandes Correaⁱⁱ

Resumo: As informações representadas por meio de recursos visuais auxiliam os Sistemas de Recuperação de Informação (SRIs) a exercerem seu objetivo principal de recuperar informações relevantes que atendam a necessidade informacional de seus usuários. O objetivo deste artigo é descrever e analisar técnicas propostas pela área de Visualização da Informação e modelos de interface discutidos na literatura de Ciência da Informação que, aplicados na construção de interfaces gráficas, proporcionam melhor apropriação da informação pelos usuários dos SRIs, auxiliando-os na busca, navegação e recuperação de informações. A metodologia consiste em revisão de literatura com foco na potencial contribuição da representação visual da informação na construção de interfaces amigáveis para SRIs, bem como na identificação e análise de visualizações utilizadas por SRIs. O uso de visualizações é de grande importância na comunicação dos SRIs com os usuários, pois ele potencializa a descoberta de novos conhecimentos, e as informações disponibilizadas por meio de representação visual são mais facilmente entendidas.

Palavras-Chave: Sistemas de Recuperação de Informação. Visualização da informação. Interface visual.

INFORMATION VISUALIZATION TO USER-FRIENDLY INTERFACE CONSTRUCTION FOR INFORMATION RETRIEVAL SYSTEMS.

Abstract: The information presented through visualization help the Information Retrieval System (IRS) to reach its main goal: to retrieve relevant information that meets the informational needs of its users. The objective of this article is to describe and analyze techniques proposed by the Information Visualization area and interface models discussed in Information Science Literature, which applied to graphical interface construction would facilitate the appropriation of information by the users of IRS and would help them to search, browse and retrieve information. The methodology consists of a literature review focusing on the potential contribution of the visual representation of information in the development of user-friendly interfaces to IRS, as well as identification and analyses of visualizations used as interfaces by IRS. The use of visualizations is of great importance in the communication between SRI and users, because the information presented through visual representation are better understood by user and allow the discovery of new knowledge.

Keywords: Information Retrieval Systems. Information visualization. User-friendly graphical interface.



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).

ⁱ Universidade Federal de Pernambuco. jessicalirav@yahoo.com.br.

ⁱⁱ Professor do departamento de Ciência da Informação da UFPE. fc_renato@yahoo.com.br.

Recebido em: 16/11/2010; aceito para publicação em: 01/09/2011.

1 INTRODUÇÃO

A sociedade atual, dita Sociedade da informação, tem sido caracterizada por um crescimento exponencial na quantidade de informações produzidas e disponibilizadas diariamente aos usuários, nos mais diversificados domínios de aplicação, o que gera um grande volume de informações de diversas origens e formatos (VALIATI, 2008).

Nesse contexto, os Sistemas de Recuperação de Informação (SRIs) são *softwares* que visam satisfazer as necessidades informacionais dos usuários, auxiliando-os a recuperar conteúdo pertinente no grande volume de informações disponíveis em coleções de documentos.

Os SRIs enfrentam uma série de dificuldades na recuperação de informações relevantes para o usuário (SOUZA, 2006). As principais dificuldades são a sobrecarga de informações ou excesso de documentos retornados; a dificuldade do usuário em expressar ao sistema sua real necessidade por meio de palavras-chave; e a dificuldade do sistema em compreender a necessidade expressa e recuperar para o usuário os documentos que satisfaçam essa necessidade.

Tais dificuldades poderiam ser minimizadas, se os SRIs possuísem uma interface gráfica amigável que permitisse uma interação efetiva do usuário com representações das informações disponibilizadas pelo sistema, facilitando a busca, navegação e recuperação da informação. Uma interface baseada na visualização possibilita a compreensão de como está organizada a informação no sistema, oferecendo uma visão geral dos assuntos tratados, contextualizando o resultado da busca e permitindo o descobrimento de novos conceitos e relações entre eles. Como afirma Valiati (2008), a representação das informações de forma gráfica busca não apenas divulgar ou apresentar dados e resultados, mas prover mecanismos que possibilitem ao usuário interagir com a informação, explorando-a visualmente.

O objetivo deste artigo é descrever e analisar técnicas propostas pela área de Visualização da Informação e modelos de interface discutidos na literatura de Ciência da Informação que, aplicados na construção de interfaces gráficas, proporcionam melhor apropriação da informação pelos usuários dos SRIs, auxiliando-os na busca, navegação e recuperação de informações.

Este artigo se encontra organizado da seguinte maneira: a seção 2 apresenta a metodologia utilizada neste trabalho; a seção 3 discute os Sistemas de Recuperação da Informação; a seção 4 discute a representação visual de informações e sua utilização em SRIs, sendo abordados a área Visualização da Informação e suas técnicas, e modelos de interface

gráfica discutidos na área de Ciência da Informação; a seção 5 apresenta discussões sobre como devem estar estruturadas as informações nas bases de dados dos SRIs para que as visualizações possam ser construídas; a seção 6 traz as conclusões do presente trabalho.

2 METODOLOGIA

A metodologia consiste em uma pesquisa de caráter exploratório com abordagem qualitativa (GIL, 2010), com o propósito de buscar referencial teórico sobre interfaces gráficas para Sistemas de Recuperação de Informação baseadas em visualizações, isto é, representação visual das informações.

Através de revisão de literatura, foram analisadas algumas das técnicas de visualização da informação, bem como modelos de interface visual discutidos na literatura de Ciência da Informação. Foram analisadas referências teóricas extraídas de artigos, teses, dissertações e livros que versem sobre visualização da informação, interface visual e recuperação de Informação. Na busca por artigos presentes nos periódicos da área de Ciência da Informação, foram consultadas as bases de dados BRAPCI, Scielo e Google Acadêmico. Foram pesquisadas também teses e dissertações, as quais foram acessadas por meio do portal nacional da Biblioteca Digital de Teses e Dissertações (BDTD) mantido pelo IBICT. Outras publicações científicas foram pesquisadas nos anais do ENANCIB de 2005 a 2010 e no Google Acadêmico.

Paralelamente, através dos trabalhos revisados e de pesquisas em mecanismos de busca na *web*, buscou-se levantar SRIs que utilizassem recursos visuais em sua interface. Esse levantamento permitiu filtrar as representações visuais utilizadas por SRIs atuais, demonstrar a viabilidade prática do uso de interfaces visuais por SRIs, bem como fazer análise desses usos.

3 SISTEMA DE RECUPERAÇÃO DE INFORMAÇÃO

Os Sistemas de Recuperação de Informação (SRIs) são sistemas ou *softwares* que tratam, essencialmente, de indexação, busca e classificação de documentos com o objetivo de satisfazer necessidades informacionais expressas através de consultas (GONZÁLES; LIMA,

2003). Como afirma Araújo Júnior (2007), os SRIs dizem respeito a um sistema de operações interligadas para identificar, dentre um grande conjunto de informações, aquelas que são de fato úteis, ou seja, que estão de acordo com a necessidade expressa pelo usuário.

Segundo Lunardi (2008), o principal objetivo de um SRI é a seleção, dentro de um universo de documentos disponíveis, do conjunto de documentos relevantes para uma necessidade de informação do usuário.

Os assuntos tratados nos documentos são representados no sistema através do processo de indexação. A indexação consiste na atribuição de termos (palavras-chave) a documentos, com a finalidade de representar as informações contidas neles, bem como facilitar a posterior recuperação desses documentos através de tais termos. A indexação pode ser realizada de forma intelectual ou automática. Na indexação intelectual, também chamada indexação manual, seres humanos elegem termos capazes de representar as informações nos documentos através do processo intelectual de análise de assunto. Na indexação automática, *softwares* atribuem termos julgados relevantes para representar as informações presentes nos documentos através de análise estatística da ocorrência de palavras ou identificação de sequência de palavras com determinadas funções linguísticas.

A busca, ou consulta, consiste no momento em que o usuário expressa sua necessidade ao sistema, e, a partir daí, o SRI realiza uma seleção de documentos que satisfaçam a busca do usuário. Para que isso ocorra, é necessário que haja uma interface, por meio da qual o usuário irá traduzir sua necessidade de informação em uma consulta (SOUZA, 2006).

Os SRIs fornecem dois modos através dos quais o usuário pode realizar a busca: recuperação e navegação. Para cada um desses modos existem modelos específicos (BAEZA-YATES; RIBEIRO-NETO, 1999). Na recuperação, o usuário expressa sua necessidade ao sistema em forma de questões ou palavras-chave. Na navegação, o usuário não propõe uma questão ao sistema, mas navega por categorias e em documentos alocados nas categorias, em busca de informações pertinentes.

A estrutura que permite aos SRIs selecionarem documentos que satisfaçam uma busca se chama índice. O índice geralmente consiste de uma estrutura de dados chamada arquivo invertido (BAEZA-YATES; RIBEIRO-NETO, 1999). Nessa estrutura, palavras apontam para os documentos nos quais ocorrem, e adicionalmente são armazenadas também a frequência de ocorrência da palavra e as posições em que ela ocorreu em cada documento.

O índice revela importantes características da coleção de documentos e do SRI, como o vocabulário do sistema, as relações de coocorrência das palavras e as relações de

similaridade entre documentos. Essas informações representadas visualmente e apresentadas ao usuário permitem uma interação mais efetiva com o SRI, auxiliando na formulação da estratégia de busca e no encontro de documentos relevantes.

4 REPRESENTAÇÃO VISUAL DE INFORMAÇÕES

Nesta seção serão abordadas a área de Visualização da Informação e algumas de suas técnicas, além de modelos de interface discutidos na área de Ciência da Informação. Visualizações utilizadas em interfaces de SRIs são analisadas a fim de ilustrar e demonstrar a importância da representação visual de informações na interação com o usuário e a contribuição das visualizações no processo de busca e recuperação de informação.

4.1 Visualização da informação

A Visualização da Informação pode ser definida como uma área da ciência que tem por objetivo o estudo das principais formas de representações gráficas para apresentação de informações de maneira a contribuir para uma melhor percepção e entendimento delas, além de possibilitar a dedução de novos conhecimentos com base no que está sendo apresentado (FREITAS et al., 2001). Consiste de uma área que combina aspectos de computação gráfica, interação humano-computador, cartografia e mineração de dados.

O processo de visualização de informação está relacionado com a transformação de dados abstratos em gráficos ou imagens, denominados visualizações. O objetivo na construção das visualizações é auxiliar no entendimento de determinado assunto minimizando o esforço cognitivo nessa compreensão, visando à inclusão informacional dos usuários ou consumidores das visualizações.

Na maioria dos casos, o uso de recursos gráficos na apresentação de informações produz a compreensão da mensagem transmitida de forma mais natural e com menos esforço cognitivo. Segundo Nascimento e Ferreira (2005), uma simples visualização pode condensar uma grande quantidade de informações, facilitando a compreensão dessas informações, já que a visão é o sentido humano que possui maior capacidade de captação de informações por unidade de tempo.

Resumidamente, as técnicas de Visualização da Informação procuram representar graficamente dados de um determinado domínio de aplicação de modo que a representação visual gerada explore a capacidade de percepção do homem e este, a

partir das relações espaciais exibidas, interprete e compreenda as informações apresentadas e, finalmente, deduza novos conhecimentos (FREITAS et al., 2001, p. 144).

Como afirma Card et al. (1999 apud PRESSER; SILVA; SANTOS, 2010), a visualização da informação procura reunir o poder da percepção visual humana e o poder de processamento do computador para analisar e compreender um conjunto de informações, de forma a otimizar o uso da capacidade visual humana para compreensão das informações transmitidas.

O avanço das tecnologias de informação, dos equipamentos de processamento de imagens e dos computadores, tem permitido a construção de sistemas cada vez mais complexos e ricos na qualidade da apresentação de informações através de recursos gráficos.

Ware (2000 apud LUNARDI, 2008) define cinco vantagens que a visualização de informações propicia quando utilizada de forma eficiente: compreensão - a visualização permite a compreensão de grande quantidade de informação; percepção - a visualização revela propriedades do dado que não podem ser antecipadas; controle de qualidade - a visualização permite o controle de qualidade dos dados, porque os problemas se tornam imediatamente aparentes; foco mais contexto - a visualização facilita a compreensão de um aspecto dentro do contexto geral dos dados em que esse encontra; e interpretação - a visualização apoia a formação de hipóteses que propiciam futuras investigações.

Autores em várias áreas reconhecem a importância do uso de estruturas de visualização da informação. A seguir são discutidos dois trabalhos sobre a temática no contexto dos sistemas de informação.

Santos e Kobashi (2009) abordam o uso de algumas técnicas de visualização da informação em apoio aos métodos estatísticos e matemáticos utilizados para mapear informações, enfatizando que essas técnicas são importantes para melhor perceber e compreender dados bibliográficos manipulados por métodos estatísticos, com a finalidade de elaborar mapas que possam representar adequadamente os aspectos quantitativos e cognitivos da ciência.

Na dissertação de Dias (2007) é enfatizada a contribuição da Visualização da Informação na disponibilização e na apropriação de conhecimento por parte dos usuários, demonstrando que a utilização de recursos visuais facilita a compreensão acerca das informações armazenadas principalmente em meios digitais. São apresentadas as estruturas de visualizações mais utilizadas e como estas podem auxiliar na representação de informações

permitindo a criação de recursos gráficos inovadores, interativos e dinâmicos, garantindo a inclusão informacional dos usuários.

4.1.1 Desenvolvimento e avaliação de visualizações

Uma metodologia para aplicação de técnicas de visualização já existentes ou para projeção de novas visualizações foi criada por Nascimento e Ferreira (2005). Essa metodologia é composta de cinco passos: estudar o problema e os dados a serem visualizados; construir visualização; definir mecanismos de interação; implementar um sistema protótipo; e avaliar as visualizações.

Para que um sistema de visualização de informação possa ser considerado útil, ele deve abordar dois conceitos: expressividade e efetividade (NASCIMENTO; FERREIRA, 2005).

Uma visualização é considerada expressiva quando é capaz de expressar todos os dados de interesse do usuário e nenhum dado a mais ou a menos do contexto desejado.

É considerada efetiva a visualização que facilita a compreensão dos dados apresentados pela estrutura, de forma que o usuário localize a informação desejada sem erros de interpretação.

Expressividade e efetividade são aspectos importantes que devem estar presentes em todo sistema de visualização de informação, pois sem eles uma visualização pode não ser eficaz, agravando a sobrecarga de informação, não trazendo todos os dados relevantes ou levando a erros de interpretação. Além disso, uma visualização pode ser também de difícil entendimento ou dificultar a localização da informação desejada pelo usuário.

4.1.2 Técnicas de visualização

A seguir algumas técnicas de visualização são discutidas, sendo abordadas as mais utilizadas atualmente em interfaces de SRIs.

Desenho de Grafos

Grafos são modelos matemáticos formados por estruturas simples que consistem de um conjunto de vértices e um conjunto de arestas. Os vértices geralmente representam objetos concretos ou abstratos em diversas áreas do conhecimento humano, e arestas indicam relações

entre esses objetos. No desenho de um grafo, adota-se uma representação gráfica para os vértices, e define-se uma localização no plano para cada vértice e aresta. (NASCIMENTO; FERREIRA, 2005).

No contexto dos SRIs, os vértices podem representar termos, palavras ou documentos, e as arestas podem representar as relações semânticas ou de coocorrência entre palavras ou termos, ou similaridade entre documentos.

A visualização utilizada no SRI do site da Queens Library baseia-se na técnica de desenho de grafos, em que os termos são representados por vértices, e as relações entre eles por arestas. Na visualização, o termo utilizado na busca permanece no centro da figura, ligado por meio das arestas aos termos relacionados. O tipo de relação entre os termos está representado pela cor da aresta, podendo ser associações, traduções, variações fonéticas ou termos do tesouro, conforme a Figura 1. Essa visualização permite ao usuário conhecer opções para refinar sua estratégia de busca, explorar o vocabulário controlado e utilizado pelo SRI e incorporar termos à busca por meio do uso do *mouse*.



Figura 1: Desenho de grafos da Queens Library
Fonte: <http://aqua.queenslibrary.org/>

Essa visualização é expressiva, pois mostra apenas as relações das palavras digitadas pelo usuário com outras contidas no vocabulário do sistema. Além disso, é efetiva, pois permite explorar a relação entre palavras e selecionar as mais convenientes para formulação de uma busca.

Desenho de Árvores

Das classes estudadas em desenho de grafos, o desenho de árvores está entre as primeiras, em função da sua popularidade e da simplicidade de sua estrutura (NASCIMENTO; FERREIRA, 2005).

O desenho de árvores é uma particularização do desenho de grafos. As técnicas para desenhar árvores consideram dois casos distintos: o desenho de árvores com raiz (do inglês *rooted trees*) que representam hierarquias, tais como diagramas organizacionais, árvores de diretórios e taxonomias; e o desenho de árvores livres (do inglês *free trees*) em que as árvores não possuem raízes e são representadas por um desenho radial em que o vértice raiz é colocado no centro da figura, e os demais vértices são posicionados em círculos concêntricos de acordo com os seus níveis. Em ambos os casos, é desejável a obtenção de desenhos planares em linhas poligonais (NASCIMENTO; FERREIRA, 2005).

A Figura 2 mostra uma visualização criada pelo mecanismo de busca Yippi, que faz uso do desenho de árvores de diretórios para organizar os documentos retornados para uma busca. Nessa visualização, os documentos resultantes da busca são agrupados em tópicos, de acordo com a semelhança entre eles. Tal semelhança é definida como ocorrência de determinado tópico (frases em comum). Ao selecionar com o *mouse* um tópico, os documentos que o contêm são exibidos à direita da visualização. Tópicos podem conter subtópicos, que funcionam com elementos de discriminação dos documentos dentro de determinado tópico. Tópicos e subtópicos funcionam como elementos de refinamento da busca quando selecionados com o *mouse*. O número de documentos aparece, entre parênteses, ao lado de cada tópico/subtópico.

Essa visualização minimiza a sobrecarga de informação sobre o usuário causada pela longa lista de resultados comuns no contexto dos mecanismos de busca na *web*. A visualização gerada pelo Yippi permite explorar os resultados da busca por meio de uma lista de tópicos ordenada alfabeticamente, possibilitando ao usuário selecionar exatamente o que procura ou descobrir relações entre tópicos e grupos de documentos.

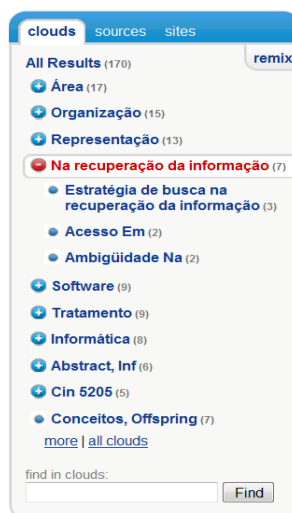


Figura 2: Desenho de árvore do mecanismo de busca Yippy
Fonte: <http://search.yippy.com/>

A visualização da Figura 2 é expressiva, pois mostra apenas tópicos contidos nos documentos presentes na lista de resultados retornados de uma busca realizada pelo usuário, além de possibilitar explorar subtópicos que permitiriam a discriminação entre os documentos de determinado tópico. Em parte é efetiva, pois permite explorar a relação entre tópicos e grupos de documentos e selecionar os grupos de documentos mais relevantes. Entretanto, a efetividade é comprometida, pois os tópicos extraídos dos documentos nem sempre são significativos.

A - Browser Hiperbólico

Técnica que utiliza o princípio do foco mais contexto. O objetivo do browser hiperbólico é apresentar uma visão geral dos dados a serem visualizados destacando uma área específica de interesse (foco) através de uma ampliação suave, enquanto há uma compactação do resto da imagem ou contexto (NASCIMENTO; FERREIRA, 2005). Faz uso também do desenho radial de árvores, em que a estrutura de uma árvore é mapeada em uma representação no plano hiperbólico a ser projetado na tela através de uma circunferência de um círculo cujo centro coincide com o centro da tela e cresce exponencialmente com o seu raio. Essa técnica consegue disponibilizar cerca de dez vezes mais vértices de uma árvore do que quando se utiliza uma visualização no plano cartesiano, a navegação também se torna mais efetiva e pode ser realizada através de simples movimentos do *mouse*, selecionando o vértice de maior interesse e arrastando-o para o centro para ter o foco desejado.

Na técnica browser hiperbólico, os conceitos ou categorias estão organizados numa hierarquia na qual o assunto mais geral é centralizado, e ligados a ele por aresta estão os

assuntos específicos em uma estrutura de árvore. Dentro de cada assunto específico, estão dispostos os termos que se ligam a ele através de ramificações que condensam os assuntos cada vez mais específicos, sendo possível, através dessa estrutura, visualizar desde os assuntos mais abrangentes até aqueles menos inclusivos dentro da árvore, o que possibilita a compactação de grandes quantidades de informações à medida que se aproxima da periferia do desenho.

O serviço de Informação Tecnológica em Agricultura (Infoteca-e) utiliza a técnica browser hiperbólico para reunir e permitir acesso a informações sobre tecnologias produzidas pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa), as quais se relacionam às áreas de atuação de seus demais centros de pesquisa. Suas coleções são formadas por conteúdos editados na própria instituição (em forma de cartilhas, livros, revistas, páginas eletrônicas, programas de rádio e de televisão), com linguagem adaptada de modo que produtores rurais, extensionistas, técnicos agrícolas, estudantes e professores de escolas rurais, cooperativas e outros segmentos da produção agrícola possam assimilá-los com maior facilidade e, assim, apropriarem-se de tecnologias geradas pela Embrapa. A Figura 3 mostra a visualização criada pela Embrapa.

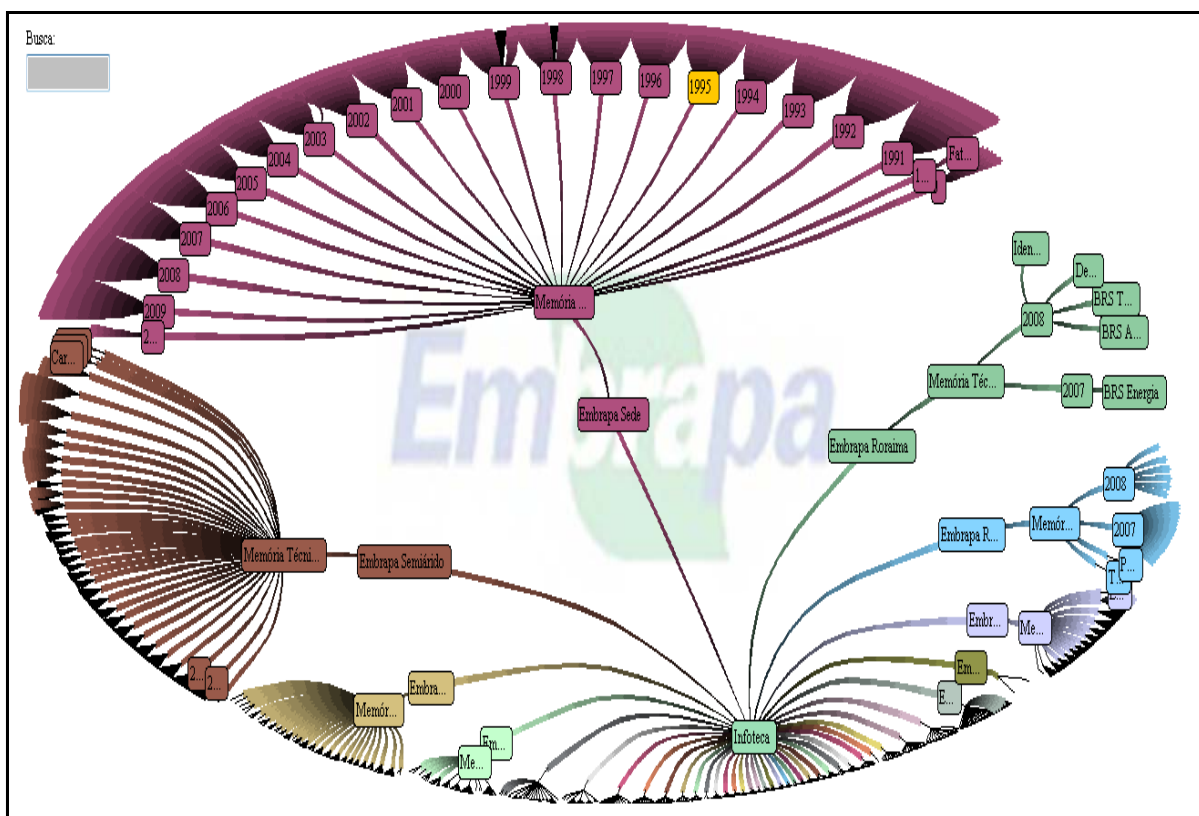


Figura 3: Browser hiperbólico da Infoteca-e/Embrapa

Fonte: <http://www.infoteca.cnptia.embrapa.br/hypertree/hypertree.html>

O uso da técnica browser hiperbólico no *site* da Infoteca-e permite visualizar, de forma gráfica, como está estruturado o acervo em termos de local (sede), tipo de produção, ano e temas organizados hierarquicamente em árvore. A visualização da Figura 3 é efetiva e pode ser realizada através de simples movimentos do *mouse*, selecionando vértices de maior interesse e mudando o foco desejado, facilitando a busca por meio de uma navegação amigável. Entretanto, a expressividade é comprometida por apresentar muitos conceitos ou elementos descritivos dos documentos. Tentando minimizar esse problema, é fornecida uma caixa de texto na qual o usuário pode digitar palavras-chave, e todas as arestas que levam até documentos que casam com a busca digitada serão destacadas a partir do nodo ou vértice central da árvore.

4.2 Modelos de interface visual

A seguir serão discutidos quatro modelos de interface presentes na literatura de Ciência da Informação, que são utilizados em interfaces gráficas de SRIs.

4.2.1 Mapa conceitual

Os mapas conceituais são ferramentas gráficas usadas para organizar e representar o conhecimento a partir de diagramas que indicam relações entre conceitos. Como afirma Silva (2006):

Mapas conceituais são interfaces para a organização e representação do conhecimento. Eles colocam os conceitos geralmente dentro de círculos ou retângulos de algum tamanho, e os relacionamentos entre os conceitos ou as proposições são indicados por uma linha que conecta os dois conceitos. Palavras sobre as linhas especificam as relações entre os dois conceitos.

Segundo Lima (2004b), os mapas conceituais, enquanto ferramentas de visualização, facilitam a estruturação de documentos publicados em forma hipertextual, disponibilizando de uma maneira amigável o conhecimento a ser recuperado e proporcionando uma navegação orientada com código de cores que representa cada nível hierárquico do mapa e explicita ao usuário como todo o conteúdo semântico é organizado e como se conecta internamente.

A representação gráfica e concisa do conhecimento através do mapa conceitual e a possibilidade de ligação entre conceitos, representados por termos ou palavras-chave, resultam em um ambiente ideal para se definir um sistema de navegação, no qual os usuários podem encontrar a informação que buscam, bem como navegar pelo sistema pesquisando

outros temas. Dessa forma, o mapa conceitual auxilia na recuperação de informações relevantes e na construção de novos conhecimentos baseados na análise, interpretação e compreensão da relação existente entre os conceitos (LIMA, 2004a).

As duas importantes características do mapa conceitual, Segundo Novak e Cañas (2008), são: a representação dos conceitos de forma hierárquica, com o conceito mais geral no início do mapa e depois os mais específicos arranjados hierarquicamente; a existência de referências cruzadas que permitem verificar como é representada a relação dos conceitos no domínio do conhecimento.

A Figura 4 mostra o mapa conceitual utilizado pelo *site* IHMC CmapTools. CmapTools é um *software* que permite construir, navegar, compartilhar e criar modelos de conhecimento representados como mapas conceituais. Essa ferramenta pertence ao *Institute for Human & Machine Cognition* (IHMC), que é um instituto de pesquisa sem fins lucrativos do Sistema Universitário da Flórida.

No *site* do CmapTools o mapa conceitual é utilizado como estrutura hierárquica para descrever o próprio programa. A utilização do mapa conceitual, enquanto ferramenta de visualização, facilita a apropriação dos conceitos e suas relações; explicita como todo o conteúdo semântico é organizado e como se conecta internamente; permite ter acesso aos documentos relacionados a cada conceito através de *links*; e permite aos usuários encontrarem as informações que buscam através da navegação.

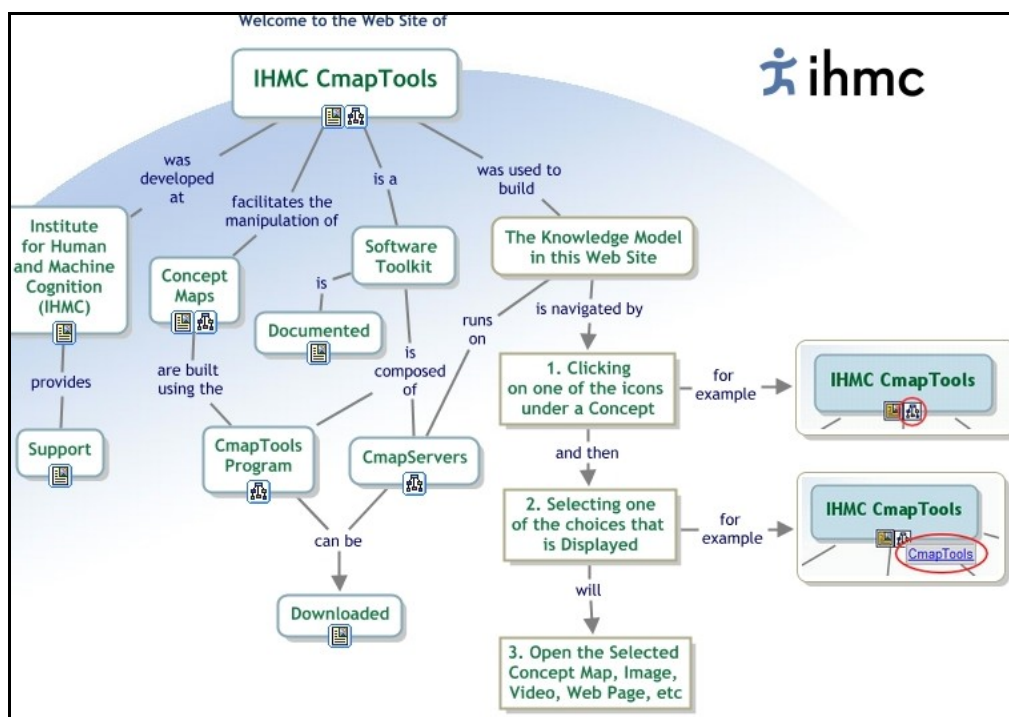


Figura 4: Mapa conceitual do CmapTools
Fonte: <http://cmap.ihmc.us/conceptmap.html>

O mapa conceitual enquanto visualização tem boa efetividade e expressividade quando o número de conceitos, relações e documentos é pequeno.

4.2.2 Cartografia temática

Segundo Eliel (2007), a cartografia temática consiste em um mecanismo de representação das temáticas inscritas em recursos informacionais, tais como livros, periódicos, teses etc. Pode ser entendida como mapas gerados a partir de dados bibliográficos, os quais proporcionam uma visualização gráfica desses dados, oferecem uma visão global de conjuntos de informações e evidenciam relações e estruturas entre elas.

A utilidade da visualização de informações por meio de mapas ou cartografia temática baseia-se em estudos sobre a percepção que mostram que o ser humano tem primeiro uma percepção global de uma cena antes de atentar para os detalhes. Tais estudos abriram caminhos para explorar as características da percepção global e, conseqüentemente, para aplicá-las aos sistemas de informação, particularmente aos aspectos relacionados à recuperação de informação, tais como (KOBASHI; SANTOS, 2008):

- Exploração rápida de conjuntos de informações desconhecidas.
- Evidenciação de relações e estruturas nas informações.
- Fornecimento de alternativas de acesso a informações pertinentes.
- Classificação interativa de informação.

A cartografia temática é, portanto, uma técnica bibliométrica avançada, que facilita o resgate de informações, de forma que novas informações possam ser produzidas a partir delas. Essa técnica facilita a compreensão das relações existentes entre os dados, quando eles são expressos por meio de mapas ou cartografias, e configura-se como interface poderosa para uso em Sistemas de Recuperação de Informação. (KOBASHI; SANTOS, 2008).

O mapa gerado pelo WEBSOM, apresentado na Figura 4, é um exemplo de cartografia temática. Esse mapa proporciona uma visão global dos tópicos em uma coleção de documentos e evidencia relações entre esses tópicos, possui uma navegação simples, possibilitando a localização de informações.

WEBSOM é um método para organizar automaticamente coleções de documentos de texto e gerar mapas visuais dos assuntos tratados a fim de facilitar a mineração de conhecimento e recuperação da informação. Os documentos estão agrupados em pontos do mapa, de acordo com a similaridade de conteúdo. Os documentos agrupados em cada ponto são listados, e o seu conteúdo pode ser consultado clicando nos pontos visíveis no nível de

menor *zoom* do mapa. A cor indica no mapa a densidade ou a tendência de agrupamento dos documentos, o amarelo indica maior densidade de documentos, e as áreas escuras em vermelho, as áreas de menor densidade. As palavras no mapa representam os tópicos ou assuntos e indicam o contexto dos grupos de documentos.

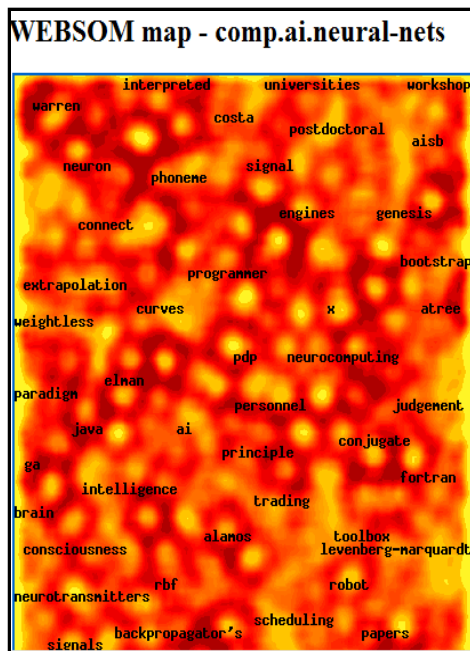


Figura 5: Cartografia temática do WEBSOM

Fonte: <http://websom.hut.fi/websom/milliondemo/html/root.html>

A Figura 5 é uma visualização gerada para fins de recuperação por navegação. Essa visualização é efetiva para o usuário que deseja explorar a coleção de documentos, conhecer os tópicos mais discutidos, tópicos relacionados e os documentos que tratam de cada tópico. Uma grande quantidade de palavras no mapa pode comprometer a expressividade da visualização agravando a sobrecarga de informação. A visualização é pouco efetiva, pois pode levar a erros de interpretação, já que, ao clicar em uma palavra, não se terá todos os documentos a ela relacionados, mas somente os documentos localizados na região selecionada.

4.2.3 Nuvem de tags

Nuvem de tags, ou nuvem de etiquetas (do inglês *Tag Cloud*), consiste numa interface textual usada por *websites* e *blogs*, na qual palavras-chave são dispostas em lista visual, com o objetivo de representar os itens de conteúdo de um *site*, mostrando aos usuários quais são os assuntos principais abordados por esse *site*. Essas palavras ou etiquetas têm a função de facilitar a navegação e seleção de documentos relacionados. Numa nuvem de tags, são

especificadas as *tags* mais utilizadas como descritores e/ou visitadas pelos internautas na coleção de documentos, ou seja, as palavras mais comuns ou mais cobiçadas (MOREIRA; ROMÃO, 2009).

Segundo Moreira e Romão (2009), a nuvem de tags funciona de dois modos: primeiro, promove a condensação dos dizeres que estão dispersos em vários outros arquivos eletrônicos; segundo, esse mesmo campo pode inscrever efeitos de espalhamento de sentidos sobre a página digital, deslocando-os para outros arquivos de sentido semelhante.

O conjunto de etiquetas que formam a nuvem de tags pode ter sua origem em um processo de indexação utilizando folksonomia. Folksonomia é a prática de indexação colaborativa das informações no meio digital realizada pelos usuários, que utilizam o seu próprio vocabulário (linguagem natural) de forma livre, sem preocupação com padronização, para indexar páginas ou informações de seu interesse contidas na *web* (MOREIRA; ROMÃO, 2009). A folksonomia visa ao compartilhamento das informações entre os leitores na *web*.

Os assuntos mais importantes da página acabam sendo indexados através das *tags* criadas pelos próprios internautas e expostas no formato de nuvem, podendo funcionar até como uma consulta pronta para os próximos usuários (MOREIRA; ROMÃO, 2009).

A quantidade de conteúdos que o *site* apresenta sobre o assunto de cada etiqueta é representada proporcionalmente pelo tamanho da fonte. As *tags* que estiverem em destaque dentro da nuvem são as de maior ocorrência no momento, representando, assim, os assuntos e/ou informações que estão em foco dentro do *site*. As etiquetas disponibilizadas na nuvem são *links* que levam a coleções de itens relacionados. Através de um clique do *mouse* em uma etiqueta, é possível ter acesso a uma página em que há vários *links* para documentos que foram indexados por essa etiqueta.

A nuvem de *tags* facilita a navegação, recuperação e compartilhamento de informações pelo usuário, visto que sua estrutura, como componente de busca dentro do *site*, situa os assuntos mais procurados e visitados através de uma interface visual amigável.

A Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) faz uso da nuvem de *tags* em sua interface. A BRAPCI é um produto de informação, fruto de projeto de pesquisa criado pelo Departamento de Ciência e Gestão da Informação da Universidade Federal do Paraná (UFPR). A base de dados é constituída pelos artigos dos periódicos da área de Ciência da Informação (CI). Atualmente disponibiliza referências e resumos de 7.969 textos publicados em 30 periódicos nacionais impressos e eletrônicos da área de CI.

A BRAPCI faz uso da nuvem de tags como interface visual, conforme Figura 6, para guiar o usuário, sugerindo termos de busca. A nuvem de tags utilizada no SRI é composta pelas etiquetas mais consultadas, isto é, o tamanho da fonte das *tags* representa proporcionalmente a quantidade de utilizações da etiqueta nas buscas dos usuários.



Figura 6: Nuvem de tags - termos mais consultados do BRAPCI

Fonte: <http://www.brapci.ufpr.br/>

Nesse tipo de interface visual não estão presentes todos os termos, mas apenas os mais discutidos ou acessados, servindo como sugestão de busca para os usuários do SRI. Esse tipo de visualização é expressivo e efetivo para usuários que querem explorar a coleção de documentos através dos assuntos mais tratados ou mais procurados e respectivos documentos. A ordem alfabética das etiquetas e a limitação do seu número à ordem de dezenas permitem a obtenção de uma boa expressividade.

5 REQUISITOS PARA A CONSTRUÇÃO DE VISUALIZAÇÕES

A representação das informações de maneira visual nas interfaces dos SRIs facilita a contextualização, localização e compreensão dessas informações pelo usuário. Entretanto, para a construção de visualizações efetivas e expressivas, as informações precisam estar organizadas no índice dos SRIs, o que requer investimento por parte das organizações no processo de organização da informação.

Para a construção de visualizações que visam demonstrar os termos e suas relações através das técnicas de desenho de grafos, desenho de árvores ou browser hiperbólico, torna-se necessária a existência de um tesouro ou taxonomia, como instrumento de organização da informação, que deixe claro as relações existentes entre os termos e que seja utilizado na indexação dos documentos.

O mapa conceitual é mais adequado na estruturação de um pequeno conjunto de documentos, hipertexto, por meio de conceitos e proposições que ligam conceitos, do que

milhares de documentos de um SRI típico. Para utilizar o mapa conceitual como interface visual, a própria estrutura do mapa conceitual deve ser construída, e, após isso, os documentos são associados aos conceitos ou proposições pertinentes.

A cartografia temática permite visualizar as relações entre os termos e documentos, com base na coocorrência estatística das palavras como indicador da proximidade ou similaridade entre eles. Os termos e documentos considerados semelhantes tematicamente são agrupados em blocos, numa estrutura de mapa visual em que são dispostos como pontos num espaço cartesiano, ou numa estrutura hierárquica de dendograma. Se forem consideradas apenas palavras isoladas, o próprio índice do SRI já é suficiente para construção desse tipo de visualizações. Entretanto, termos são melhores descritores que palavras isoladas e contribuem para uma melhor discriminação da informação. A utilização de termos requer a construção e utilização na indexação de vocabulários controlados.

Na nuvem de tags, as etiquetas que denotam termos são dispostas em forma de nuvem e estão organizadas de acordo com a quantidade de uso/acesso, ou seja, as *tags* contidas na nuvem representam os principais assuntos abordados ou mais consultados no momento. Não há relações definidas entre as *tags* presentes na nuvem, não existindo relações hierárquicas, associativas ou de qualquer tipo entre as *tags*. Para utilização dessa interface visual, seria necessária a adoção de uma lista de termos de indexação ou vocabulário controlado.

6 CONCLUSÃO

No contexto atual em que os Sistemas de Recuperação de Informação (SRIs) têm enfrentado dificuldades em exercer o objetivo de recuperar informações relevantes para seus usuários, a representação visual de informações se apresenta como alternativa para potencializar a comunicação com o usuário e tornar os SRIs mais efetivos.

Os usuários, ao utilizarem um sistema de Recuperação de Informação, geralmente apresentam incerteza quanto ao que informar ao sistema no momento da busca e na seleção do que é pertinente para atender a sua necessidade diante da lista de documentos retornados. Levando em conta esse fato, o presente trabalho buscou contribuir demonstrando que as visualizações podem auxiliar o usuário a entender a estrutura e organização dos dados no sistema e possibilitam a interação com as informações e a descoberta de conhecimento. Além

disso, este trabalho contribui no levantamento de requisitos necessários para construção de visualizações para SRIs.

As visualizações, para serem úteis no contexto dos SRIs, devem ser construídas de maneira que sejam efetivas e expressivas. A obtenção de boas visualizações decorre não só do emprego correto de técnicas de visualização de informação e uso de modelos de interfaces visuais, mas carece de grande investimento na organização da informação, principalmente na construção e uso de vocabulários controlados.

Foram encontradas dificuldades em localizar trabalhos com o foco específico na aplicação da visualização da informação na construção de interface para SRIs, pois na literatura de Ciência da Informação são poucos os trabalhos que discutem as visualizações neste domínio de aplicação. Devido a esse fato, foi necessário utilizar referências de outras áreas, como a Ciência da Computação, na tentativa de obter informações mais específicas.

Como trabalhos futuros, pretende-se analisar artigos da literatura internacional, tendo o objetivo de encontrar métodos que permitam construir visualizações capazes de acelerar a apropriação de conhecimento por parte dos usuários de SRIs.

REFERÊNCIAS

ARAÚJO JUNIOR, R. H. **Precisão no processo de busca e Recuperação da informação**. Brasília: Thesaurus, 2007.

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

DIAS, M. P. **A contribuição da Visualização da Informação para a Ciência da Informação**. 2007 Dissertação (Mestrado)-Pontifícia Universidade Católica de Campinas, Centro de Ciências Sociais Aplicadas, Pós-Graduação em Ciência da Informação, Campinas, 2007. Disponível em: < http://www.bibliotecadigital.puc-campinas.edu.br/tde_arquivos/2/TDE-2007-08-02T080844Z-1353/Publico/Mateus%20Pereira%20Dias.pdf > Acesso em: 14 ago. 2011.

ELIEL, O. **Cartografia temática de artigos de periódicos nacionais da Ciência da Informação (período de 1986 a 2005): contribuições ao campo teórico da área**. 2007. 120 f. Campinas: PUC- Campinas, 2007. Dissertação (Mestrado)- Programa de Pós- graduação em Ciência da informação, Pontifícia universidade Católica de Campinas, Campinas, 2007. Disponível em: < http://www.bibliotecadigital.puc-campinas.edu.br/tde_busca/arquivo.php?codArquivo=278 >. Acesso em: 24 fev. 2010.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010. 185 p. ISBN 9788522458233.

FREITAS, C. M. D. S.; CHUBACHI, O. M.; LUZZARDI, P. R. G.; CAVA R. A. Introdução à Visualização de Informações. **RITA** – Revista de Informática Teórica e Aplicada, Instituto de Informática, v. 8, p. 143-158, 2001.

GONZALES, M.; LIMA, V. L. S. Recuperação de Informação e Processamento da Linguagem Natural. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 23. , 2003, Campinas. **Anais do III Jornada de Mini-Cursos de Inteligência Artificial**, Volume III, Rio de Janeiro: [s.n.], 2003. p.347-395.

KOBASHI, N.Y.; SANTOS, R. N. Macedo dos. Arqueologia do trabalho imaterial: uma aplicação bibliométrica à análise de dissertações e teses. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis,, n. esp., 1º sem. 2008. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/view/1130>>. Acesso em: 17 nov. 2009.

LIMA, G. A. B. O. Mapa Conceitual como ferramenta para organização do conhecimento em sistema de hipertextos e seus aspectos cognitivos. **Perspectiva em ciência da informação**, Belo Horizonte, v.9 , n.2, p. 134-145, jul./dez. 2004.

LIMA, G. A. B. O. **Mapa Hipertextual (MHTX)- um modelo para organização hipertextual de documentos.**2004. 199f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004. Disponível em:< <http://www.gercinalima.com/glima/uploads/tesegercina.pdf>>. Acesso em: 24 fev.2010.

LUNARDI, M. S. **Visualização em nuvens de texto como apoio à busca exploratória na web.** São Paulo. 2008. Dissertação (Mestrado)- Escola Superior de Desenho Industrial, Universidade do Estado do Rio de Janeiro, São Paulo, 2008. Disponível em: < http://www.bdtd.uerj.br/tde_busca/arquivo.php?codArquivo=1522 > Acesso em: 11 ago. 2011.

MOREIRA, Vivian Lemes; ROMÃO, Lucilia Maria Sousa. O funcionamento discursivo das nuvens de *tags* na rede eletrônica: sentidos sobre Capitolina. **DataGramZero**, v. 10, n. 6, nov./dez., 2009. Disponível em: <http://www.dgz.org.br/dez09/Art_02.htm> em:<<http://www.portalppgci.marilia.unesp.br/enancib/viewpaper.php?id=230>>. Acesso em: 03 mar. 2010.

NASCIMENTO, H. A. do; FERREIRA, C. B. R. Visualização de Informações - Uma Abordagem Prática. In: UNISINOS, 2005. Rio Grande do Sul. XXV Congresso da Sociedade Brasileira de Computação. **Anais ...** Rio Grande do Sul: São Leopoldo, 2005, p. 1262-1312.

NOVAK, J. D.; CAÑAS, A. J. The Theory Underlying Concept Maps and How to Construct Them. *Florida Institute for Human and Machine Cognition*, 2008, 36 p. Disponível em:< <http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf>>. Acesso em: 23 ago. 2011.

PRESSER, N. H.; SILVA, E. L. da; SANTOS, R. N. M. dos. RECURSOS DE FORMULAÇÃO E VISUALIZAÇÃO DE INDICADORES PARA APOIAR PROCESSOS DE GESTÃO EDUCACIONAL EM IESS. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 7, n.2, p. 247-259, jan./jun. 2010.

SANTOS, Raimundo Nonato Macedo dos; KOBASHI, Nair Yumiko. BIBLIOMETRIA, CIENTOMETRIA, INFOMETRIA: conceitos e aplicações. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 2, n. 1, p. 155-172, 2009.

SILVA, M.F. Diagramas hierárquicos, mapas conceituais e mapas hiperbólicos: um estudo comparativo entre interfaces hipertextuais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 7., 2006, Marília/ SP. **Anais Eletrônicos...** Marília: ANCIB, 2006. 12 p.

SOUZA, R. R. Sistemas de Recuperação de Informação e Mecanismos de Busca na web: panorama atual e tendências. **Perspectivas em ciência da informação**. Belo Horizonte, V.11, n. 2, p. 161-173, maio/ago. 2006.

VALIATI, Eliane Regina de Almeida. **Avaliação de Usabilidade de Técnicas de Visualização de Informações Multidimensionais**. Tese (Doutorado) - Programa de Pós-Graduação em Ciência da Computação. Universidade Federal do Rio grande do Sul. Porto Alegre, 2008.