

RAPHAEL FIGUEIREDO XAVIER

**ANÁLISE DE MÉTODOS DE PRODUÇÃO DE
INTERFACES VISUAIS PARA RECUPERAÇÃO DA
INFORMAÇÃO**

MARÍLIA
2009

RAPHAEL FIGUEIREDO XAVIER

ANÁLISE DE MÉTODOS DE PRODUÇÃO DE INTERFACES VISUAIS PARA RECUPERAÇÃO DA INFORMAÇÃO

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista, UNESP – Campus de Marília, como requisito parcial para obtenção do título de Mestre.

Área de concentração: Informação, Tecnologia e Conhecimento.

Linha de pesquisa: Informação e Tecnologia

Orientador: Edberto Ferneda

MARÍLIA
2009

Xavier, Raphael Figueiredo

X3A Análise de Métodos de Produção de Interfaces Visuais para Recuperação da Informação / Raphael Figueiredo Xavier. – Marília, 2009.

78 f. ; 30 cm.

Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, 2009.

Bibliografia: f. 74-78.

Orientador: Dr. Edberto Ferneda.

1. Recuperação da Informação. 2. Visualização da Informação. 3. Interfaces Visuais para Recuperação de Informação. I. Autor. II. Título.

CDD 006.6

RAPHAEL FIGUEIREDO XAVIER

ANÁLISE DE MÉTODOS DE PRODUÇÃO DE INTERFACES VISUAIS PARA RECUPERAÇÃO DA INFORMAÇÃO

BANCA EXAMINADORA:

Presidente e Orientador

Dr. Edberto Feneda
Programa de Pós Graduação em Ciência da Informação
Universidade Estadual Paulista – UNESP – Campus de Marília

Membro Titular

Dr. Guilherme de Ataíde Dias
Departamento de Ciência da Informação
Universidade Federal da Paraíba – UFPB

Membro Titular

Dra. Silvana Aparecida Borseti Gregorio Vidotti
Programa de Pós Graduação em Ciência da Informação
Universidade Estadual Paulista – UNESP

Local: Universidade Estadual Paulista
Faculdade de Filosofia e Ciências
UNESP – Campus de Marília

Data da defesa: 29/09/2009

*Dedico este trabalho a Gabriela e
Beatriz. Mulheres que me possibilitaram
a felicidade cotidiana.*

AGRADECIMENTOS

O período em que estive matriculado no programa de Pós-Graduação fazendo o mestrado foi um período de completa mudança em minha vida. Algumas adoráveis, outras, entretanto, um tanto amargas. Inúmeras pessoas passaram por esse período me ajudando, apoiando e cobrando. Todas elas merecem ser agradecidas por mais essa fase.

Primeiramente, gostaria de agradecer a Gabriela e a Beatriz, que estão ao meu lado em dias difíceis, dias fáceis, nas vitórias e derrotas que acumulamos. Que crescem ao meu lado. Amo vocês.

Ao Renan e Josi que estiveram morando comigo no fim da minha graduação e começo da Pós. Um período extremamente difícil e de muitas incertezas. Saudades.

Meu ingresso a pós-graduação se deve ao Professor Edberto Ferneda, que apostou no meu projeto e que me deu total liberdade de trabalho até o fim desta pesquisa. Obrigado por sua confiança e amizade.

Aos amigos em que dividi a sala nas disciplinas no primeiro ano: Emanuel, Marta, Lucilene, Elvis, César, Zeca, Fabiano, Rachel, Maria José, Aldinar, Luana, Maria Carolina, Fabio, Simone e os outros que minha memória me trai agora.

Ao grupo Intersemioses e comunidade Ciam que e resgataram um espírito artístico que andava meio longe de mim: Plácida, Maria José, Rachel, Aldinar, Fabiano, Lis Helena, Ana Carolina (Karu), Luciana, Wellington, Ângela, Etine e outros.

À Profa. Plácida que me ajudou e socorreu diversas vezes tanto na área da Ciência da Informação como em minha vida pessoal (a bendita alergia da Beatriz...)

Ao Professor Izidoro Gil Leiva e sua esposa Margarida, pela grande amizade construída no período em que estiveram no Brasil.

Aos alunos do grupo de circo que a cada dia sonhavam mais alto na possibilidade de voar. *Halle Up!*

À Julia, pelas melhores aulas de inglês que já tive e pelo socorro na nossa mudança da “casa molhada” para o “apartamento 1”. Adoramos você!

Às vizinhas mais legais que poderíamos ter enquanto moramos no “apartamento 1”: Ciça e Aline. Saudades dos churrascos na varanda.

Ao Carlos, Luana e Camila, pela amizade sincera, apoio e empurrões. Me orgulho muito de ser amigo de vocês!

Ao professor Antonio Tobias e UNIFLOR pela proposta de trabalho no Mato Grosso na qual pude aprender muitos dos meandros da profissão que escolhi. Esta mudança proporcionou novos rumos, desafios, frustrações e esperanças na qual estou tentando me sintonizar até hoje. Disseram-me que isso é ser adulto.

Ao Walter, que me deu toda ajuda possível no meu processo de mudança geográfica e pessoal, além de ser daqueles amigos que estão para “o que der e vier”.

Aos amigos feitos na Uniflor: Lucicléia, Gisele, Letícia, Luciana, Vinícius, Farol, Gláucia, Leandro, Marina, Lídio, Fabiana, Hélio, Loiva, Marcão e todos outros professores e funcionários.

Aos meus alunos da Uniflor pela amizade e paciência comigo. Um grande abraço a turma da Base.

À Aline que através de sua ida para Peixoto de Azevedo (E você foi mesmo!), estreitou mais ainda nossa amizade.

À Elaine que apesar da atual distância, continua presente e uma grande amiga.

Ao Professor Alexandre Galheta e Aidênia França, que permitiram minha volta ao sudeste no momento em que eu precisava voltar.

À minha mãe e meu pai Mauro que tiveram mais uma vez paciência e fé em mim e me ajudaram na volta pra casa.

Ao meu pai Xavier pelo seu carinho e dedicação com os filhos, mesmo que um pouco atrasado.

Aos amigos feitos na Faculdade Pitágoras: Viviane, Lenildo, Luana, Lucila, Larissa, Liliane, Merlin, Fábila, Vanessa, Thaís, Hernandez, Kalleo, Carol, Lázaro, Patrícia e Verônica.

Aos antigos e eternos amigos que agora posso ver com muito mais frequência: Daniel, Limão, Milton, Daniele, Ettore, Jorge, Beto, Orc, Paraíba.

Por fim, e em especial, a Profa. Silvana Vidotti, que apesar do distanciamento, tenho a mais profunda estima. Pelas oportunidades oferecidas, pela força compartilhada, pela sinceridade acima de tudo. Sempre que tenho que tomar uma decisão difícil, penso em como você lidaria com ela. Obrigado!

*Visão é a arte de enxergar coisas
invisíveis
(Jonathan Swift)*

XAVIER, Raphael Figueiredo. **Análise de Métodos de Produção de Interfaces Visuais para Recuperação da Informação**. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2009.

RESUMO

O advento da *Web* e o conseqüente aumento no volume de informações eletrônicas acarretaram muitos problemas em relação ao acesso, busca, localização e recuperação de informação em grandes volumes de dados. O presente trabalho realiza uma revisão dos diferentes modelos, métodos e algoritmos existentes para a geração de Interfaces Visuais para Recuperação da Informação, classificados segundo ao seu processo de produção: Análise e Transformação dos Dados, Aplicação de Algoritmos de Classificação e Distribuição Visual e Aplicação de Técnicas de Transformação Visual. Os resultados pretendem servir a outros investigadores como ferramenta para a eleição de uma ou outra combinação metodológica no desenvolvimento de propostas específicas de Interfaces Visuais para Recuperação da Informação, além de sugerir a necessidade de maiores investigações sobre novas técnicas de transformação visual.

Palavras-chave: Recuperação de Informação; Visualização da Informação; Interfaces Visuais para Recuperação de Informação

XAVIER, Raphael Figueiredo. **Análise Métodos de Produção de Interface Visuais para Recuperação da Informação.** Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2009.

ABSTRACT

The advent of the Web and the consequent increase in the volume of electronic information had caused many problems about access, search, location and retrieval of information in large volumes of data. This work is a revision of the different models, methods and algorithms to create interfaces for Visual Information Retrieval, classified according to their production process: Analysis and Data Processing, Implementation of algorithms for classification and distribution of Visual and Application Processing Techniques of Visual. The results of other researchers want to serve as a tool for the election of one or another combination methodology in the development of specific proposals for visual interfaces for information retrieval, and suggest the need for more research into new techniques for processing visual.

Keywords: Information Retrieval; Information Visualization; Visual Information Retrieval Interface

LISTA DE FIGURAS

Figura 1: Processo do modelo clássico de recuperação da informação	16
Figura 2: Processo de recuperação baseado no sistema <i>Querying</i>	17
Figura 3: Pesquisa baseada no <i>Browsing</i>	18
Figura 4: Recuperação da informação e seus 2 níveis.....	19
Figura 5: Representação vetorial de um documento com dois termos de indexação	26
Figura 6: Representação vetorial de um documento com três termos de indexação.	27
Figura 7: Representação de um conjunto documental com a vetorização.	27
Figura 8: Modelo de Geração de Estrutura de Visualização da Informação.....	36
Figura 9: Representação na técnica Bifocal Display (a) e Perspective Wall (b).	38
Figura 10: Aplicação do Perspective Wall.....	39
Figura 11: Esquema básico da técnica cone tree, com as ramificações dispostas na base de um cone, com nodo principal no topo.....	40
Figura 12: Interface tridimensional da Cam Tree (com desenvolvimento horizontal)	40
Figura 13: Esquema básico da técnica tree-maps.	41
Figura 14 : Modelo Treemap apresentada pelo MIT.....	42
Figura 15: Esquema básico da técnica hyperbolic tree.	43
Figura 16: Modelo <i>Hiperbolic tree</i> apresentada pela Inxight.....	43
Figura 17: Tipos de minerações na WEB.....	52
Figura 18: Esquema de co-citação entre documentos	54
Figura 19: Esquema de Acoplamento Bibliográfico entre documentos	55
Figura 20: Esquema de Co-acesso entre documentos.....	56
Figura 21: Aplicação de <i>Clusterings</i>	58
Figura 22: Representação em <i>Pathfinder</i>	59
Figura 23: Aplicação de uma representação MDS	61
Figura 24: Análise de co-ocorrência baseado no WebSOM	62
Figura 25: Estruturas preservadas por cada técnica	64
Figura 26: Aplicação da técnica <i>detail+overview</i> no <i>Google Earth</i>	66
Figura 27: Aplicação da visão olho de peixe (Fisheye View) com a ferramenta “aiSee”.....	68
Figura 28: Combinações entre técnicas de desenvolvimento de interfaces visuais para recuperação da informação	70

Sumário

1. Introdução	14
1.1 Objetivos.....	21
1.2 Metodologia	22
2. Recuperação da Informação.....	24
2.1 Modelo Vetorial	25
2.2 <i>Browsing</i>	31
3. Visualização da Informação.....	35
3.1 Modelos de Interfaces Visuais.....	37
4. Produção Automatizada de Interfaces Visuais para Recuperação da Informação .	46
4.1 Estabelecimento de modelos visuais para recuperação da informação.....	46
4.2 Desenvolvimento de Interfaces Visuais para Recuperação da Informação.....	50
4.2.1 Análise e Transformação de Dados.....	51
4.2.2 Aplicações de Algoritmos de Classificação e Distribuição Visual.....	56
4.2.3 Aplicações de Técnicas de Transformação Visual	65
4.3 Análise dos Métodos	70
5. Considerações Finais	72

1. Introdução

O valor de um sistema de informação está condicionado não somente à qualidade e à quantidade da informação disponível, mas também por sua capacidade ou facilidade de encontrar determinada informação, facilidade esta que naturalmente diminuirá conforme aumente a quantidade de informação. Não é de se estranhar, portanto, que a investigação em técnicas de acesso e recuperação de informação tenha experimentado seu maior auge com o surgimento de um sistema de informação de dimensões gigantescas, como a *World Wide Web*. (HASSAN-MONTERO, 2006)

Pesquisadores em recuperação da informação há tempos vêm buscando encontrar formas que tornem sistemas mais acessíveis aos usuários finais. Têm ainda procurado novos meios dos usuários explorarem os dados em uma base de dados. Técnicas de visualização - métodos computacionais para exibição gráfica de uma grande quantidade de informação (CARD, 1999), parecem promissoras como um meio para atingir estes objetivos. (MORSE; LEWIS, 1997)

Avanços da Visualização da Informação foram impulsionados significativamente pelas pesquisas em Recuperação da Informação. Um dos problemas centrais para pesquisadores e profissionais ligados à recuperação da informação é melhorar a eficácia e a eficiência do processo de recuperação. De modo geral, quanto mais usuários conheçam seu espaço de pesquisa, mais as suas buscas se tornarão efetivas (BÖRNER, 2003). Diversos sistemas de visualização da informação descrevem estruturas semânticas globais da coleção de documentos. Usuários utilizam dessa “visualização estrutural” como uma base para pesquisas posteriores.

Recuperação da Informação (RI) é um campo de pesquisa que se refere ao processo de busca, exploração e descoberta de informação em repositórios de dados organizados a fim de satisfazer às necessidades informacionais de seus usuários. Zhang (2008) apresenta dois componentes fundamentais na recuperação da informação: *Information Retrieval* e *Information Organization*. Na perspectiva dos usuários, a *Information Organization* é um processo interno nos sistemas de recuperação de informação, sendo essencial para a *Information Retrieval*, porém, invisível e não transparente. Na perspectiva do sistema, a *Information Organization* é indispensável, vital e fundamental. O método da *Information Organization* afeta diretamente o método da *Information Retrieval*. E isto não é uma exceção para os sistemas de visualização para recuperação da informação. Estes sistemas necessitam processos de *Information Organization* apropriados que suportem apresentações visuais e recursos de recuperação (ZHANG, 2008).

Adequando-se a este esquema, analisemos duas frentes da recuperação da informação: o *Browsing* e o *Query searching* (ou simplesmente *Querying*). Análise esta, que nos ajudará a compreender as necessidades da aplicação da visualização da informação na área de Recuperação da Informação.

Estes dois paradigmas básicos da recuperação da informação refletem dois tipos de comportamento na busca de informação. Cada um deles possui vantagens e desvantagens, e complementam-se na RI (ZHANG, 2008).

O *Querying* é considerado uma pesquisa analítica baseada no modelo clássico, representado na Figura 1 (BELKIN, CROFT, 1987). Neste modelo há um conjunto de documentos (*corpus*) a serem buscados, representados através de termos de indexação (palavras-chave). O usuário formula uma busca expressando sua necessidade informacional para que o sistema verifique através de seu vocabulário controlado os documentos que satisfazem a expressão de busca. A partir disso, o usuário inspeciona os documentos recuperados e seleciona os que lhe interessa (COX, 1992).

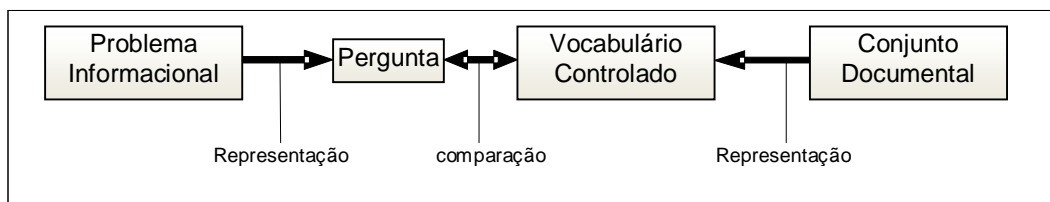


Figura 1: Processo modelo clássico de recuperação da informação

Fonte: adaptado de Cox, 1992

Ainda que este modelo tenha funcionado bem e se adaptado às mudanças tecnológicas, diversos problemas permanecem existindo em suas aplicações. (BATES, 1989; COOPER, 1988; WILLETT e WOOD, 1989). Alguns podem ser listadas como:

- Dificuldade em formular a expressão de busca
- Resultado vazio na busca
- Adaptação para outros objetos além de documentos (como sons e imagens)
- Definição de termos que representem o objeto

O *Querying* é uma complexa tarefa na qual se envolve a articulação entre a necessidade informacional formulada em uma expressão de busca com a representação dos objetos pelo sistema através de termos de um vocabulário controlado (ZHANG, 2008). A possibilidade de uma pesquisa através do *Querying*, onde a expressão do usuário não coincida com a representação dada pelo sistema faria com que a busca falhasse em seus objetivos.

Diversos autores, a fim de contornar esta dificuldade, sugerem o *Browsing* como um sistema alternativo capaz de se utilizar da capacidade de reconhecimento de usuários humanos, podendo, ainda, ajudá-los a superar algumas dificuldades na pesquisa (BATES, 1989; COVE e WALSH, 1987; MARCHIONINI, 1987; HILDRETH, 1982; ODDY, 1977).

O *Browsing* refere-se à visualização, olhar global, visão geral e exploração da informação no conjunto documental. É um importante recurso

para explorar e descobrir informação. O ambiente informacional é indispensável, essencial e vital para o funcionamento do *Browsing* (HASSAN-MONTERO, 2006). Porém, Zhang (2008) lembra que a capacidade do *Browsing* ainda é limitada quando aplicada em ambientes informacionais que se apresentam seus resultados de busca através de listas lineares.

O *Browsing* pode ser definido como “uma atividade interativa de busca na qual a direção da pesquisa é determinada pelo usuário com base no *feedback* imediato do que o sistema apresenta” (COX, 1992).

Baseando-se no processo de recuperação por *Querying* apresentado na Figura 1, Cox (1992) representa na Figura 2 o processo de recuperação de objetos examinados em um pequeno espaço documental minimamente estruturado.

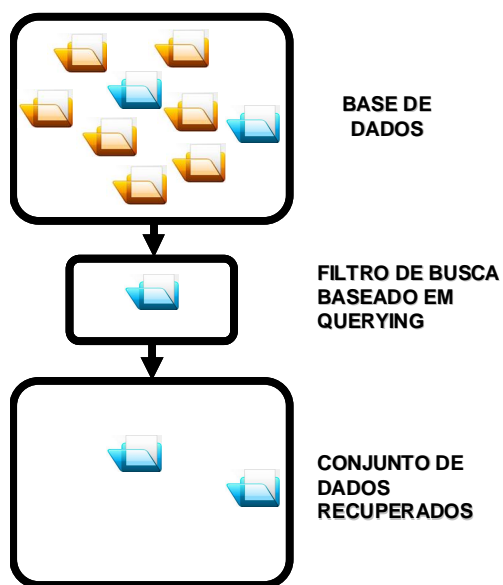


Figura 2: Processo de recuperação baseado no sistema *Querying*

Fonte: adaptado de Cox, 1992

Na busca baseada no *Browsing* é necessário que o usuário se posicione na base de dados já estruturada. O usuário se move dentro desta estrutura com base nas informações recebidas do sistema (Figura 3). Em nenhum momento se isolam os objetos para uma análise independente. O usuário desloca-se livremente para as suas áreas de interesse.

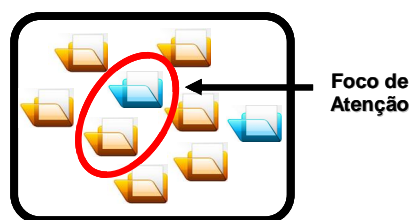


Figura 3: Pesquisa baseada no *Browsing*

Fonte: adaptado de Cox, 1992

Ajudando a contextualizar os dois modelos de recuperação, Zhang (2008) aponta dois diferentes graus de informações a serem recuperadas: um de nível micro e outro de nível macro. Informações de nível micro referem-se a objetos individuais, ou se tratando de um documento, elementos individuais (por exemplo, palavras-chave e resumo) enquanto que o nível macro da informação se refere a conjuntos de objetos agregados em uma coleção de dados. Enquanto a informação de nível micro é óbvia e direta, a de nível macro é indireta e sofisticada. Vale ressaltar que as informações agregadas são geradas a partir da junção dos objetos individuais que formam o micro-nível informacional. Os dois níveis se distinguem em sua natureza. O valor dado às informações agregadas prove da conexão de objetos, tendências e padrões que surgem a partir dos objetos individuais no micro-nível. A informação no nível macro também ajuda usuários explicar a localização de informações relacionadas a outros itens ilustrando visões gerais holísticas e contextos heurísticos.

O objetivo final de um sistema de recuperação de informação é prover a usuários resultados precisos, relevantes e confiáveis. Portanto, certamente o sistema não deveria incluir somente informações de nível micro. Para este objetivo, os dois níveis de informação devem estar disponíveis e acessíveis aos usuários. Na maioria dos sistemas é perceptível a predominância do *Querying* na recuperação de informações de nível micro através da comparação entre expressão de busca e os termos que representam os objetos individualmente. A informação agregada (nível macro) é pouco utilizada, se não totalmente ignorada neste modelo. O *Browsing*, porém se foca tanto nos objetos isolados como também as de nível macro conforme a figura 4.

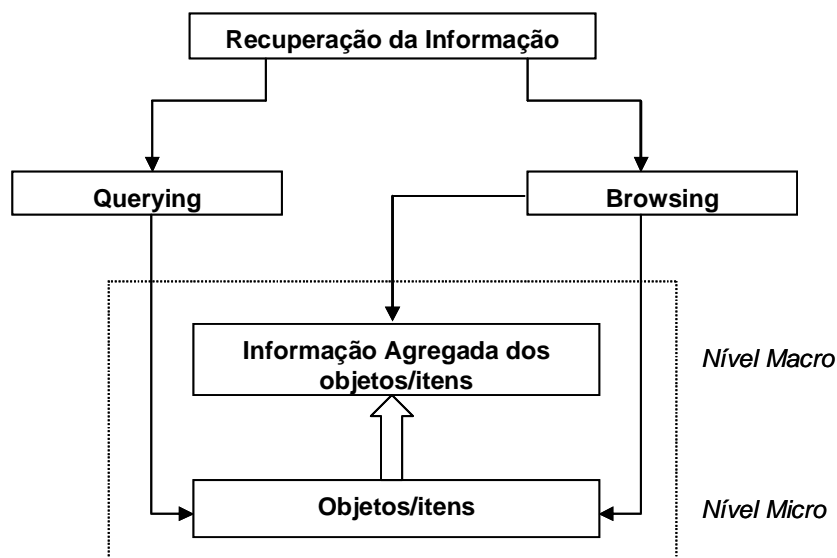


Figura 4: Recuperação da informação e seus 2 níveis

Fonte: adaptado de Zhang, 2008

Encaminhando a discussão para a visualização da informação, são necessários alguns comentários acerca das características espaciais de conjuntos documentais já mencionados.

O espaço informacional é multidimensional, abstrato e invisível, possuindo duas relações básicas: semântica e espacial (ZHANG, 2008). Características semânticas são perceptivas através da organização do conjunto os dados, revelando suas relações. As características espaciais não são tão óbvias como as semânticas. Por si só, informações abstratas não tem forma definida (KOIKE, 1993), e, por isso, não constituem um espaço. São as relações semânticas entre os dados/informações que estabelecem a estrutura espacial do espaço documental.

Um espaço informacional pode ser constituído por características intrínsecas como palavras-chave, citações, hiperlinks e autores; ou extrínsecas: como catálogos de assuntos e Tesouros. Pode ainda possuir ambos os atributos. Uma importante propriedade do espaço é a distância entre dois objetos no conjunto documental capaz de, por exemplo, definir caminhos mais curtos entre links em sistemas baseados em hiperlink.

O *Browsing* é dependente deste ambiente informacional e está claramente associado à direção, distância, posição, e outras propriedades espaciais fundamentais, constituindo uma série de movimentos espaciais de um foco de atenção a outro. Quando permanecem no mesmo foco, usuários examinam o conteúdo do objeto e seu contexto, julgam a relevância de documentos, identificam tendências, analisam *clusters* significativos, comparam padrões, avaliam os resultados da pesquisa ou reformulam suas estratégias de busca. Os padrões espaciais do *Browsing* são forçados diretamente nestes pontos.

Portanto, as interfaces visuais para recuperação da informação referem-se ao processo de transformação de dados abstratos invisíveis e suas relações semânticas em um conjunto de dados dentro de um *display*, e a possibilidade de visualização do processo interno de recuperação por parte dos usuários (ZHANG, 2008). Basicamente, podemos compreender a visualização para recuperação da informação em dois componentes: *apresentação* visual da informação e a *recuperação* visual da informação, onde a apresentação visual oferece uma plataforma onde a recuperação visual é executada e conduzida (LIN, 1997).

Criar e desenvolver ambientes visuais para recuperação da informação é um complicado processo que é afetado por diversas variáveis devido à multiplicidade de possíveis estruturas visuais, objetos a serem visualizados, métodos de organização, abordagens de apresentações e controles de busca.

Em trabalho de Börner et al (2003), os autores realizaram revisões de diferentes etapas e técnicas para a produção de interfaces visuais para recuperação da informação, firmando bases teóricas e metodológicas para a área.

Após a contextualização do tema abordado, caracterizamos o problema que motivou o desenvolvimento de nossa pesquisa: a pouca produção científica e bibliográfica sobre representações visuais que permitam novas abordagens na Recuperação da Informação. Isso sem falar em trabalhos produzidos no Brasil, que ainda tem mínima participação neste próspero escopo de pesquisa. Investidores em Recuperação da informação têm buscado maneiras para

tornar sistemas de buscas mais acessíveis para seus usuários, e ainda, novas formas de exploração de dados/informação. Modelos de visualização se colocam promissores como meio de atingir estes objetivos.

O instigante das representações visuais para recuperação da informação é a diversidade de possíveis modelos. A alta dimensionalidade dos atributos dos objetos e suas sofisticadas relações no espaço documental juntamente com a limitação espacial das interfaces visuais, sugerem uma redução dimensional dos objetos para que eles possam se ajustar na interface visual. Com isso, atributos significativos são identificados e preservados enquanto que atributos irrelevantes são sacrificados. Sendo assim, podemos chegar de várias formas a identificação de atributos principais dos objetos e diversas formas de desenvolver modelos visuais que os apresentem. Embora haja uma variedade de modelos e sistemas visuais para recuperação da informação, um procedimento básico para seu desenvolvimento é comum a todos. Desenvolver mecanismos interativos de recuperação da informação é crucial para que usuários finais possam explorar espaços informacionais visuais.

1.1 Objetivos

Como já dito, diversas variáveis influenciam diretamente na elaboração de modelos visuais para RI (estruturas visuais, objetos a serem visualizados, métodos de organização visual, abordagens de apresentações, controles de busca, etc.). Para estabelecê-las, uma série de passos devem ser levados em conta. A pesquisa objetiva detalhar estes processos, destacar suas potencialidades e discuti-los.

Para tanto, consideram-se como objetivos específicos:

1. Pontuar aspectos teóricos da Visualização da Informação aplicada à Recuperação da Informação;
 2. Apresentar técnicas de interfaces visuais que representaram avanços significativos para a área;
 3. Descrever combinações de técnicas de desenvolvimento de interfaces visuais para recuperação da informação.
-

A presente investigação se insere na linha de pesquisa “Informação e Tecnologia do Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências UNESP - Campus de Marília, na área de concentração “Informação, Tecnologia e Conhecimento”, cuja delimitação compõe a análise de interfaces visuais para recuperação de informação.

1.2 Metodologia

Esta pesquisa caracteriza-se como exploratória e descritiva, com levantamento e análise documental sobre o tema e análise das etapas e procedimentos de modelos de interfaces visuais. O levantamento bibliográfico baseou-se na literatura registrada em livros, artigos, teses e dissertações capazes de abarcar a temática desenvolvida. Pouca bibliografia foi encontrada em língua portuguesa. Os textos, em sua maioria, são preponderantes em língua inglesa e espanhola e vêm das áreas de Ciência da Informação, Tecnologia e Ciência da Computação. O processo de recuperação da informação, incluindo seus dois modelos de Browsing e Querying (FERNEDA, 2003; MARCHIONINI, 1995; HASSAN-MONTERO, 2006; RIJSBERGEN, 1975), a área de Visualização da Informação (CARD, 1999; ZHANG 2008; CHEN, 2006) e os processos de criação de interfaces visuais para recuperação da informação (HERRERO-SOLANA, 2000; HERRERO-SOLANA E HASSAN-MONTERO, 2006; LIN, 1997; ZHANG, 2008; CHUNG *Et al*, 2003; POLANCO E ZARTL, 2002; BÖRNER *Et al.*, 2003) subsidiaram o levantamento bibliográfico no desenvolvimento de uma base teórica adequada à pesquisa.

A dissertação está organizada em cinco capítulos, entre os quais se inclui o introdutório (**Capítulo 1**) no qual é apresentado o universo da pesquisa, seu problema, proposta e objetivos. Bem como a metodologia e justificativa do trabalho desenvolvido.

No **Capítulo 2** são apresentados aspectos relacionados diretamente a recuperação da informação. Diversos modelos e teorias em representações visuais para recuperação da informação dependem diretamente das teorias, modelos e conceitos básicos da Recuperação da Informação. Em outras

palavras, não podemos explicar visualização no contexto proposto sem fundamentarmos a Recuperação da informação e seus dois paradigmas básicos: o Browsing e o Querying.

Os conceitos de visualização da informação, representação visual e seu processo de abstração são descritos no **Capítulo 3**. Estes conceitos tornam-se fundamentais para a construção e processo de criação de interfaces visuais para recuperação da informação. São apresentadas, ainda, técnicas de visualização que representaram avanços significativos para o desenvolvimento da área.

No **Capítulo 4** é descrito o processo de desenvolvimento de interfaces visuais para recuperação da informação. Primeiramente é discutido uma série de passos que devem ser levados em consideração na escolha de modelos visuais antes mesmo de sua criação, como por exemplo: a determinação do paradigma de recuperação (Querying e Browsing); a identificação dos objetos a serem expostos; a extração de atributos dos objetos; a definição estrutural do espaço visual; dentre outros. Ao todo, são descritos oito passos. Após essa descrição é apresentada três etapas ou fases que compõe diversos esquemas metodológicos propostos para a produção automatizada de interfaces visuais para a recuperação da informação: Análise e transformação de dados; Aplicação de algoritmos de classificação e distribuição visual; e Aplicação de técnicas de transformação visual. São comparados e discutidos os diferentes métodos de cada etapa.

Por fim, no **Capítulo 5** são apresentadas as considerações finais pontuando os principais aspectos da pesquisa em relação aos resultados obtidos e objetivos traçados, com destaque a suas possibilidades futuras campo da Ciência da Informação.

2. Recuperação da Informação

Apresentam-se neste capítulo os fundamentos conceituais e interdisciplinares relacionados ao campo da Recuperação da Informação, enquanto área específica da Ciência da Informação.

Ao discutir conceitos, teorias, modelos ou metodologias relacionadas à visualização de interfaces para recuperação da informação, não podemos simplesmente ignorar ou contornar os conceitos e teorias básicas da Recuperação da Informação. Em outras palavras não podemos descrever de forma explícita os modelos gráficos para recuperação da informação sem buscar um referencial teórico que aponte os princípios da Recuperação da informação.

A recuperação da informação é uma extensa área de pesquisa com relativa maturidade teórica. A criação de seu conceito é atribuída a Calvin Mooers, que em 1950 definiu o problema da documentação como “uma busca de informação em um conjunto de dados armazenados, efetuados a partir da especificação de um tema”.

Nos sistemas analíticos de recuperação da informação devemos distinguir entre sistemas de recuperação de dados e recuperação de informação (RIJSBERGEN, 1975). Esta diferenciação está motivada principalmente pela técnica empregada na comparação entre a expressão de busca e espaço documental, denominadas como equiparação exata e equiparação parcial, respectivamente.

A equiparação exata é um método determinista de comparação entre expressão de busca e espaço documental no qual resultam somente

documentos que expressam completamente os critérios expressados na busca. Em outras palavras, só se recuperam os documentos representados por todos os termos utilizados na expressão de busca.

A equiparação parcial é um método que possibilita tanto a obtenção de resultados parcialmente válidos ou pertinentes como a ordenação destes resultados de acordo com o grau de relevância para a expressão de busca utilizada. Para isso, Herrero-Solana e Hassan-Montero (2006) apontam as necessidades que modelos de busca devem proporcionar:

- Um método para transformar documentos textuais em representações numéricas e computáveis;
- Um método que, sobre esta representação dos dados, realize uma ponderação automática dos termos que compõe o documento;
- Um método de equiparação que permita calcular o grau de relevância presumida de cada documento do *corpus* para uma expressão de busca com base em seus termos.

Herrero-Solana e Hassan-Montero apontam, entre diferentes modelos de recuperação de informação, o **modelo espaço vetorial** (ou simplesmente **modelo vetorial**).

2.1 Modelo Vetorial

O modelo vetorial sugere um espaço no qual é possível recuperar resultados que respondam parcialmente a uma busca. A representação dos dados se obtém mediante a vetorização do conjunto documental: o documento, representado por um vetor (V), onde cada elemento representa o peso do respectivo termo de indexação (t), representa uma dimensão ou eixo gerando uma matriz de espaço multidimensional com colunas representando os termos e as linhas do documento.

No modelo vetorial, introduzido pela primeira vez por Salton (1989), um documento é definido por n funcionalidades e atributos independentes. Estes recursos podem ser usados para a descrição de assuntos do documento. Em

muitos casos, estes atributos são extraídos de títulos, resumos, palavras-chave, ou do próprio texto completo (ZHANG, 2008, p.22).

$$d_i = (t_{i1}, t_{i2}, t_{i3}, t_{in})$$

Na equação acima, d_i é um documento, enquanto que t_i são as características que descrevem o documento. Seu valor ou peso reflete a importância do atributo t_{in} para o documento d_i . O valor de t_{in} é geralmente normalizado para variar entre zero e um, e n é o número de características ou dimensionalidade do espaço vetorial.

Ferneda (2003, p.28) apresenta dois exemplos de representação gráfica dos documentos DOC_1 e DOC_2 ; No primeiro é tem-se o DOC_1 com termos de indexação t_1 e t_3 , com pesos 0.3 e 0.5, respectivamente:

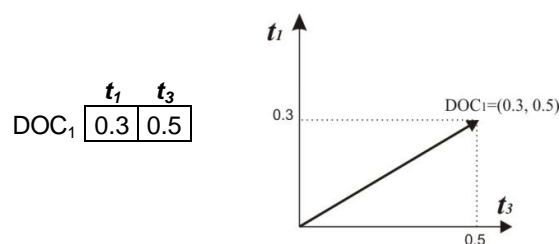


Figura 5: Representação vetorial de um documento com dois termos de indexação

Fonte: Ferneda (2003, p. 28)

No segundo exemplo, a Figura 6 representa graficamente um DOC_2 com termos de indexação (t_1 , t_2 e t_3) de pesos 0.5, 0.4 e 0.3, respectivamente, em um espaço tridimensional:

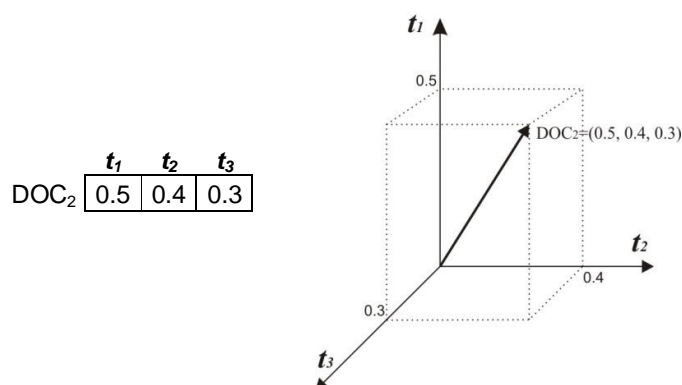


Figura 6: Representação vetorial de um documento com três termos de indexação.

Fonte: Ferneda, 2003, p. 28

O autor ainda ressalta que estas representações só foram permitidas devido ao uso de apenas 3 termos:

Obviamente, um sistema real contém um grande número de termos de indexação e documentos. Um *corpus* contendo um número indefinido de termos e documentos e termos de indexação pode ser representado através de uma matriz onde cada linha representa um documento e cada coluna representa a associação de um determinado termo aos diversos documentos. (FERNEDA, 2003, p.29)

A representação de grandes espaços vetoriais consiste de uma matriz *documentos-termos* na qual as linhas e colunas são documentos e atributos respectivamente.

	t_1	t_2	t_3	...	t_i
DOC ₁	$w_{1,1}$	$w_{2,1}$	$w_{3,1}$...	$w_{i,1}$
DOC ₂	$w_{1,2}$	$w_{2,2}$	$w_{3,2}$...	$w_{i,2}$
.
.
DOC _n	$w_{1,n}$	$w_{2,n}$	$w_{3,n}$...	$w_{i,n}$

Figura 7: Representação de um conjunto documental com a vetorização.

Fonte: Ferneda (2003, p.30)

A Figura 7 ilustra a representação de um grande espaço vetorial onde $w_{i,n}$ é o peso do documento n para o termo i . Desta forma, cada documento possui um vetor coligado e organizado por pares de elementos (termo1, peso1; termo2, peso2..., termo n , peso n).

O valor de cada elemento do vetor representa a importância do respectivo termo na descrição do documento. Estes pesos podem ser atribuídos manualmente, o que necessitaria de pessoal especializado trabalhando durante certo tempo. No entanto, Salton e McGill (1983, p.204-207) propõem um forma automática de calcular os pesos dos vetores que representam os documentos partindo-se da premissa que os termos de um documento representam o seu conteúdo (LUHN, 1958).

Inicialmente define-se a frequência de um termo (“*term frequency*” - *tf*) como sendo o número de vezes que um determinado termo *t* aparece no texto de um documento *d*.

$$tf_{t,d} = freq_{t,d}$$

Essa medida (*tf*) não faz distinção entre termos que ocorrem em todos o documentos do *corpus* e termos que ocorrem somente em alguns documentos. Sabe-se intuitivamente que um termo que aparece em todos os documentos terá provavelmente pouca utilidade em identificar a relevância dos documentos. Portanto, para um cálculo preciso do peso de um determinado termo de indexação é preciso uma estatística global que caracterize o termo em relação a todo o *corpus*. Esta medida, chamada “*inverse document frequency*” (*idf*), mostra como o termo é distribuído pelo *corpus*, e é calculada da seguinte forma:

$$idf_t = \frac{N}{n_t}$$

onde *N* é o número de documentos no *corpus* e *n_t* é o número de documentos que contém o termo *t*.

Quanto menor o número documentos que contém um determinado termo, maior o *idf* desse termo. Se todos os documentos do *corpus* contiverem um determinado termo, o *idf* desse termo será igual a um (1).

Finalmente, o peso de um termo *t* em relação a um documento *d* (*w_{t,d}*) é definido através da multiplicação da medida *tf* pela medida *idf*. Essa nova medida é conhecida como *tf*idf* e possui a seguinte fórmula:

$$w_{i,d} = tf_{i,d} \times idf_i$$

A medida $tf*idf$ é utilizada para atribuir peso a cada elemento dos vetores que representam os documento do conjunto documental. Os melhores termos de indexação (os que apresentarão maior peso) são aqueles que ocorrem com uma grande freqüência em poucos documentos.

Igualmente à representação documental, no modelo de espaço vetorial podemos, também definir como vetores as expressões de buscas feitas pelo usuário. Na equação abaixo q_n é o peso da característica n e seu valor será definido de acordo com a necessidade informacional do usuário. Fica evidente a semelhança das representações estruturais dos documentos e da expressões de busca.

$$q_i = (q_{i1}, q_{i2}, q_{i3}, q_{in})$$

Zhang (2008) nos adverte que em circunstâncias normais, o número de características individuais (n) em uma matriz *documentos-terms* é extremamente volumoso devido à quantidade de termos utilizada na indexação dos documentos. Quando o número de documentos em uma coleção aumenta, o número de atributos também aumenta. Entretanto, a relação entre o número de documentos e o número de características não é simplesmente linear. Quando a quantidade de documentos atinge um certo nível, a quantidade de atributos tende a se estabilizar (ZHANG , 2008, p. 23).

Listamos a seguir diversos pontos fortes do modelo de espaço vetorial:

1. A estrutura baseada em vetores é adequada para a representação de objetos com vários atributos. O espaço vetorial é um caminho natural para representação documentos devido sua grande variedade de características.
 2. A idéia de pesos complementa a indexação após a extração de termos relevantes que caracterizam o documento. Os pesos permitem quantificar quais termos representam melhor o objeto ou a totalidade em que se encontram
-

-
3. Da mesma forma, os pesos podem também ser ligados aos termos de busca, tornando-a mais acurada e flexível.
 4. Baseado no modelo vetorial, uma variedade de cálculos de similaridade podem ser desenvolvidos como medidas baseadas em distância, baseadas em ângulos etc. uma vasta gama de aplicações de similaridade no espaço vetorial permite usuários fazerem escolhas mais refinadas. Através de diferentes aplicações de comparação da expressão de busca e documentos.
 5. O modelo vetorial oferece uma atmosfera ideal para diversas implementações de métodos e técnicas sofisticadas de processamento como: *self-organization maps* (SOM), redes associativas *Pathfinder*, modelos de escalonamento multimnsional (MDS).

Apresentamos a seguir alguns problemas e desafios do modelo vetorial:

1. Uma das dificuldades encontradas no espaço vetorial é a alta dimensionalidade inerente, que faz com que sua eficiência dependa de coleções relativamente menores.
 2. Teoricamente, atributos e características (neste caso, termos) podem ser extraídos de um documento. Estas características são usadas para descrever o assunto dominante do objeto. Entretanto, estes termos selecionados são usados na construção de uma matriz para uma futura recuperação, neste período, o contextos semânticos dos atributos do documento podem ser perdidos. Desde que possuam diversos significados e seu exato significado só pode ser julgado no próprio contexto. Esta é a dificuldade em fazer atribuições corretas de termos em uma matriz *documentos-termos*, o que se não sanado pode resultar em problemas de ambigüidade.
 3. O modelo de espaço vetorial pressupõe que os descritores dos documentos são independentes. O problema é que esta hipótese pode simplificar demais a relação entre o uso de um termo e seu contexto.
-

Uma vez representado numericamente o conjunto documental e ponderado de forma automática cada um dos termos de cada documento, a comparação parcial se obtêm mediante funções de similaridade.

As funções de similaridade são usadas pra indicar semelhanças entre dois objetos através de um valor numérico. Um alto número representa uma grande similaridade, e vice e versa.

As funções de similaridade, conforme indica Moya-Anegón (1994), são aportes do campo da matemática que têm sido aplicados no campo da recuperação da informação. Existem várias funções de similaridade, porém as que têm apresentado melhores resultados são as que se baseiam em produto escalar: a de co-seno (também chamada de Salton), a de Dice e a de Jaccard. Todas estas funções de similaridade podem ser facilmente utilizáveis no modelo vetorial.

Os valores da similaridade entre uma expressão de busca e cada um dos documentos do *corpus* são utilizados no ordenamento dos documentos resultantes. Portanto, no modelo vetorial o resultado de uma busca é um conjunto de documentos ordenados pelo grau de similaridade entre cada documento do *corpus* e a expressão de busca.

Como vemos, a maior vantagem na utilização de técnicas de equiparação parcial, e, portanto, o modelo vetorial, é que possibilitam a ordenação dos resultados em função do grau de relevância para as necessidades informacionais do usuário.

2.2 Browsing

Em contraste com buscas formais (estratégias analíticas), percebemos também buscas informais baseadas em estratégias heurísticas. Essas buscas informais e interativas são conhecidas pelo termo *Browsing*. Geralmente *browsing* tem caráter informal e oportuno, e sua busca depende austeramente do ambiente organizacional.

Em uma estratégia de busca por *browsing*, em oposição ao *querying*, o usuário explora ou inspeciona o conjunto documental, sem a necessidade de

expressar previamente quais são suas necessidades informacionais. É uma resposta natural dos recursos humanos físicos, emotivos e cognitivos. Solucionando diversos problemas de buscas de informações do mesmo modo que humanos encontram e buscam objetos físicos. Isto se torna efetivo devido a ambientes, particularmente os criados por humanos, serem geralmente organizados sistematicamente e altamente redundantes - especialmente ambientes informacionais que são ajustados de acordo com algum princípio organizacional.

O *Browsing* se torna particularmente efetivo para problemas informacionais com dificuldade em definição de um tema ou temas interdisciplinares e/ou quando o objetivo da busca da informação é obter uma visão global sobre o tópico. Pode ainda ser útil para manter-nos atualizados frente ao desenvolvimento de um determinado campo.

Esta é uma estratégia usada em numerosas situações cotidianas (ao explorar as estantes de uma biblioteca ou livraria), porém em circunstâncias do meio digital, o melhor exemplo de busca por browsing é a atividade de navegação hipertextual, onde o usuário explora visual e “espacialmente” o conjunto hiperdocumental com o objetivo de encontrar ou localizar informação de seu interesse.

O termo é usado em uma variedade de áreas incluindo: zoologia, arquitetura, design, marketing, e Ciência da Informação (CHANG & RICE, 1993; MARCHIONINI, 1995). A literatura da Ciência da Informação e da Computação usam o termo Browsing simplesmente para significar navegação, *Scanning*, e *Scrolling*, e diversas interfaces incluindo “browsers” que permitem usuários moverem ou entrarem em telas, janelas registros e bases de dados. Na literatura da Biblioteconomia, *Browsing* ainda é associado a fichas catalográficas e livros em estantes. Em seu artigo, Bates (1989) trouxe a discussão da importância do Browsing em ambiente com estratégias de busca online, enquanto que Cove e Walsh (1985) propuseram a segmentação do browsing em três grandes categorias:

1. *Browsing* de busca (*Search browsing*), quando se tem claro o produto desejado a ser pesquisado.
-

-
2. *Browsing* de propósito Geral (*General purpose browsing*), quando se pesquisa em uma base para buscar fontes que contenham produtos de interesse.
 3. *Browsing* imprevisto (*Serendipity browsing*), quando a pesquisa é totalmente ao acaso, desestruturada, e não direcionada.

Browsing é usado para diversos propósitos e é manifestado em diferentes meios. Em contraste com estratégias analíticas, no qual a definição tem total peso para o sucesso e eficiência da busca, estratégias baseadas no browsing podem ser aplicadas a objetivos mais gerais ou informais e dependem muito mais da informação encontrada durante a busca.

Fazemos Browsing para obter uma visão geral de um espaço conceitual ou físico. Para esquadrihar uma cena ou um documento, nós podemos identificar características e usá-las para formar impressões do documento e fazer analogias com espaços ou cenas conhecidas. Por exemplo, ao esquadrihar o título da página, tabela de conteúdos, sumário, índices, e referências de um livro, nós ganhamos um senso do escopo do conteúdo, sua cobertura e a perspectiva organizacional do autor e assim podemos rapidamente decidir se devemos investir tempo lendo a obra. Isto é importante ao notar que nos casos dos livros, esses atributos utilizados primeiramente são pré-estabelecidos por diversos padrões. Hildreth (1982) aponta que esses recursos dos livros encorajam e potencializam o Browsing. Os ambientes eletrônicos ainda estão apenas começando a construir um suporte sistemático para o Browsing. Nós podemos fazer browsing em aplicações de softwares gastando alguns minutos ao atravessar menus, executar comandos gerais ou manipular objeto, ou ainda examinar o escopo da ajuda e das instruções.

Podemos listar algumas razões para realizar o browsing:

- Obter uma visão Geral
 - Monitorar um processo
 - Dividir/compartilhar uma carga cognitiva
 - Tornar mais claro uma necessidade informacional
-

- Desenvolver uma estratégia formal
- Descobrir/aprender

Discutimos inicialmente, ao longo do capítulo, o modelo tradicional de recuperação da informação, mais especificamente o modelo vetorial, suas características e funções de apoio como a inversão de frequência de documentos e as funções de similaridade.

Estes conceitos, algoritmos e teorias essenciais de recuperação da informação são pré-requisitos para o desenvolvimento de interfaces visuais para recuperação da informação. Estas funções são vasto espectro de soluções para um mesmo problema, cabe a pesquisadores escolhe-los e desenvolve-los de acordo com sua necessidade.

Ainda neste capítulo, verificamos características do browsing. Uma das áreas de investigação mais promissoras baseadas neste modelo são aquelas representadas por estudos em Visualização da Informação ou Visualização Científica, distinção terminológica condicionada à natureza dos dados representados (POLANCO e ZARTL, 2007). Estes estudos têm tido como enfoque direto nas representações de Interfaces visuais para a Recuperação da Informação.

3.

Visualização da Informação

O avanço das tecnologias de Informação e comunicação permitem disponibilizar um grande volume de informações de diversas origens e formatos. Acessando grandes bancos de dados ou realizando buscas na Web obtemos facilmente um volume gigantesco de informações. A sobrecarga de informações é uma das principais preocupações na representação de resultados obtidos através de sistemas de recuperação da informação. Uma abordagem para contornar as dificuldades para selecionar as informações relevantes dentre os resultados dentre o resultado da pesquisa é utilizar técnicas de visualização de informações (GERSHON & EICK, 1995; 1997).

Visualização da Informação (VI) é um notório campo de pesquisa com uma recente história, e tem crescido rapidamente devido a sua extensibilidade em campos de pesquisa interdisciplinares. Trabalhos em visualização da informação são agora encontrados em vários domínios, como bibliotecas digitais e interação homem-computador.

A visualização, genericamente, é o uso de imagens para representação de informação significativa (STASKO et al., 1997). Através de vários enfoques a visualização busca um objetivo claro: transformar o dado em algo com mais significado, ou seja, uma representação útil de forma que o observador humano possa ter um melhor entendimento. Atualmente, essa transformação é realizada com o auxílio do computador através de recursos gráficos, porém, nem sempre visualizações gráficas necessitaram de recursos computacionais. Diversas representações visuais são comuns no nosso cotidiano nos informam e nos orientam pelo uso da semiótica. Card (1999) define visualização de informação como o uso de representação visual, interativa e suportada por

computador, de dados abstratos para ampliar a cognição. O objetivo de representar o dado abstrato visualmente consiste em auxiliar os indivíduos a enxergarem um fenômeno no dado, usando a percepção para diminuir a sobrecarga cognitiva. Processo identificado como “cristalização do conhecimento”. Isso ocorre quando um indivíduo coleta dados para um propósito específico, os analisa através de estruturas de representação e então empacotam todas essas formas de representação para tomar uma decisão. Mas para que a representação visual do dado seja efetiva, é importante que o mapeamento preserve o dado, Card (1999) sugere um modelo de referência para mapear o dado em formas visuais que aplica uma série de transformações no dado bruto, o transformando em uma visualização.

A Figura 8 apresenta o modelo de geração de estruturas de visualização da informação. A concepção dessas estruturas é iniciada pela organização dos dados brutos em uma tabela de dados, a partir da qual se constrói uma estrutura visual que representa as informações presentes na tabela, como: gráficos, setores, diagramas, esquemas e mapas. Para isso, é realizada uma transformação dos atributos (tabela de dados) para formas gráficas espaciais representativas, para a obtenção de uma estrutura visual, uma imagem, que ative o sistema perceptivo do usuário. Este pode manipular a estrutura visual de várias maneiras (transformações de visões), ou seja, criam-se as visões que permitem ao usuário observar as estruturas visuais sob algum enfoque em específico e tomar alguma decisão ou realizar alguma ação (tarefa).

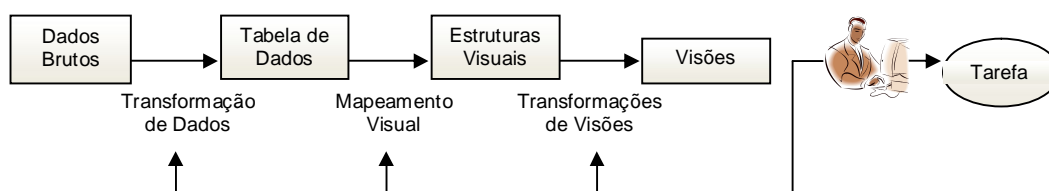


Figura 8: Modelo de Geração de Estrutura de Visualização da Informação

Fonte: adaptado de Card, 1999, p.17

Através de transformações de dados que contêm valores derivados ou estruturas, o Dado Bruto é transformado numa relação ou conjunto de relações

em forma de uma tabela de dados (mais estruturadas e mais facilmente mapeadas para formas visuais), chamada de entidade.

A concepção de estruturas de Visualização da Informação é iniciada pela organização dos dados brutos em uma tabela de dados, chamada de entidade, a partir da qual se constrói uma estrutura visual a fim de representar as informações presentes na entidade. A partir daí, as tabelas são mapeadas para representações visuais que otimizam e aumentam o conjunto espacial com propriedades gráficas a fim de codificar a informação.

Para demonstrar a expansão do campo científico e apresentar progresso na área, descrevemos a seguir algumas das técnicas de visualização da informação que representaram avanços significativos para este estudo.

3.1 Modelos de Interfaces Visuais

O crescente desenvolvimento de técnicas de visualização tem-se baseado, em geral, em implementações experimentais avaliadas de forma empírica e isolada. Observando a literatura podemos intuir que muitas das aplicações e conceitos utilizados atualmente advêm de técnicas que visam apresentar documentos estruturados, isto é, arquivos seqüenciais mas com organização hierárquica. Estas técnicas representaram avanços significativos na área de visualização da informação. Representações em redes, mapas e dispersão são tratadas mais profundamente no capítulo 4.

Nas representações hierárquicas, os elementos se apresentam em diferentes níveis, ramificações ou agrupamentos, que descendem de um nó (nodo) raiz – é uma representação comum de dados que tem em sua própria natureza uma visualização hierárquica, como por exemplo, uma visualização de estruturas complexas de diretórios de arquivos do computador. Conjuntos de dados desta estrutura foram amplamente utilizados em demonstrações de técnicas de visualização nos trabalhos pioneiros realizados no XEROX PARC.

Podemos perceber, ainda, que muitos dos conceitos utilizados atualmente surgiram de técnicas visavam apresentar documentos estruturados,

isto é, arquivos com organização hierárquica, como a visualização em olho de peixe (*Fisheye View*) de Furnas (1986) e a *Bifocal Display*.

Na técnica Bifocal Display (FURNAS, 1986), itens informacionais são colocados em três áreas diferentes, sendo a central a que apresenta a informação em destaque, enquanto as outras informações do contexto geral são figuradas nas laterais da região focal (Figura 9b). A parte central com o foco, por ocupar uma área privilegiada, acaba sendo exibida de forma distorcida, por ter tamanho maior que suas laterais. O mesmo conceito foi utilizado pela equipe da XEROX PARC na *Perspective Wall* (Figura 9b), onde o espaço de informação é colocado em uma "parede", com os "muros" laterais contendo a informação não destacada no momento.

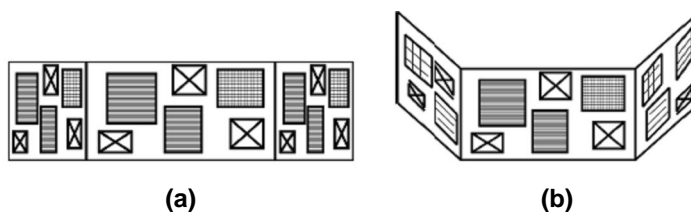


Figura 9: Representação na técnica Bifocal Display (a) e Perspective Wall (b).

Fonte: Freitas et. al., 2001, p.151

Abaixo, na figura 10, podemos verificar uma aplicação do *Perspective Wall* desenvolvido pela *Inxight*.

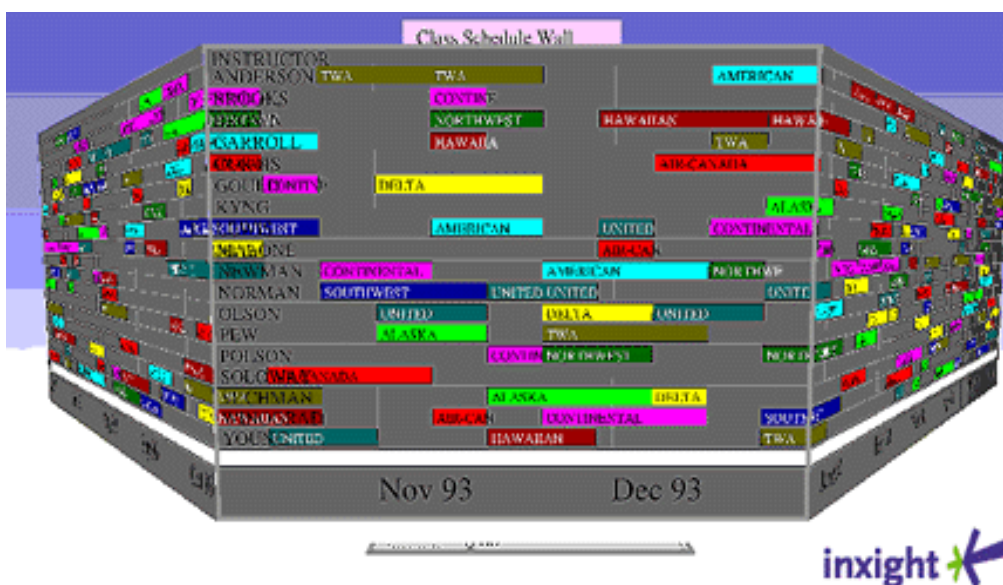


Figura 10: Aplicação ação do Perspective Wall

Fonte: <http://web.tiscali.it/soniasilva/vistimedata.html>

A técnica *Cone tree* se utiliza de uma representação de informações hierárquica na qual o nodo raiz de uma árvore, é localizado no ápice e suas ramificações são apresentadas logo abaixo (Figura 11). O objetivo desta representação é apresentar uma estrutura na qual uma hierarquia inteira ou um grande percentual dela seja visível sem necessidade de *scrolling*, e permita a supressão ou exibição de nodos durante a navegação. Também desenvolvidas pela XeroxPARC, são uma versão tridimensional com efeito de sombra das Hyperbolic Tree. As Hyperbolic Trees tinham como característica principal o crescimento exponencial de seus elementos ou nodos.

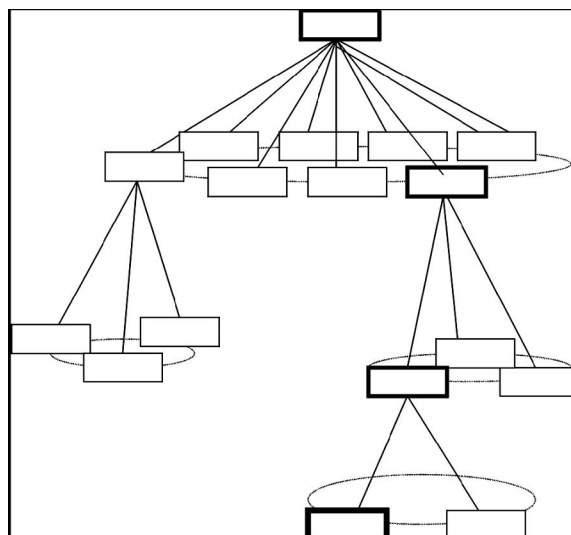


Figura 11: Esquema básico da técnica cone tree, com as ramificações dispostas na base de um cone, com nodo principal no topo.

Fonte: Freitas et. al., 2001, p.152

Com recursos de rotação, animação e *zoom*, a representação oferece acesso rápido às informações com boa orientação para visualização. *Cam trees* (Figura 12) e *Reconfigurable Disc Trees* são variações de *cone trees* com representação horizontal de hierarquia, no primeiro caso, e uso de discos no lugar de cones para reduzir a oclusão de nodos, no segundo.

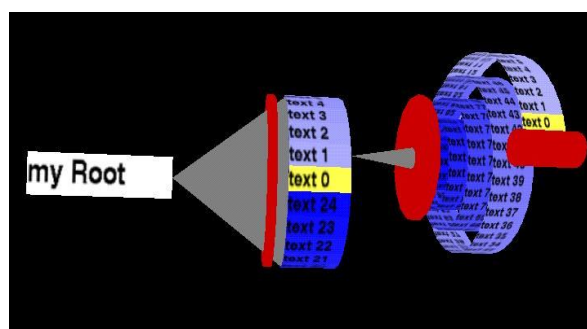


Figura 12: Interface tridimensional da Cam Tree (com desenvolvimento horizontal)

Uma abordagem diferente da utilização de espaços geométricos pode ser vista no *Tree maps* (Johnson, Shneiderman, 1991). Como podemos ver na figura 13, neste tipo de representação a estrutura hierárquica é representada em subdivisões sucessivas no espaço do *display*. Cada uma dessas

subdivisões representa um diretório, e que por sua vez, também possui outras subdivisões. Essa técnica, também conhecida como *space-filling*, foi o pioneiro de diversas outras abordagens similares (*Cushion TreeMaps*, *Information Slices* e a interface do *Sunburst*).

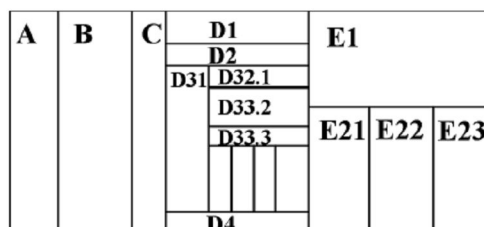


Figura 13: Esquema básico da técnica treemaps.

Fonte: Freitas et. al., 2001, p.153

Abaixo, apresentamos uma *tree-map* desenvolvida por Ben Shneiderman, do *Massachusetts Institute of Technology* (MIT). Este modelo foi baseado na Classificação Decimal de Dewey (manual de classificação usado por bibliotecas).

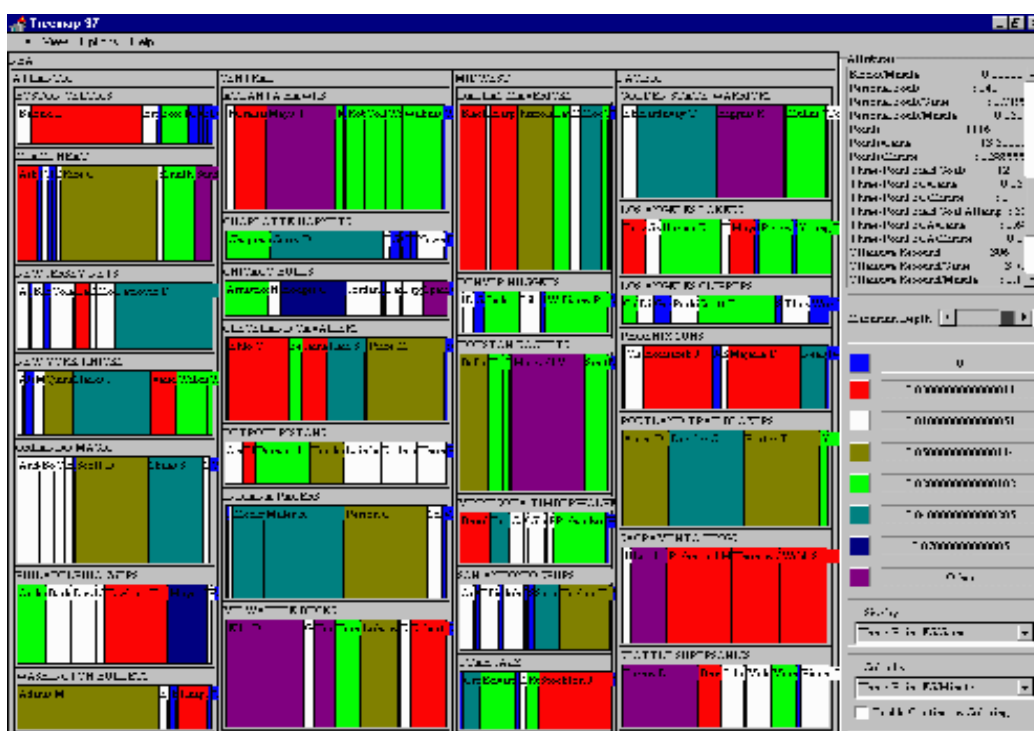


Figura 14 : Modelo Treemap apresentada pelo MIT.

Fonte: <http://www.cs.umd.edu/hcil/treemap-history/>

Como já afirmado antes, uma das técnicas que representou avanço considerável em termos de apresentação e navegação foi a *Hyperbolic Tree* (Figura 15). Esta representação apresenta hierarquias através de uma visualização radial disposta em um plano hiperbólico. Apresenta ainda, aspectos de construção conhecidos como espinha de peixe através da indicação de um nodo de interesse, que é exibido no centro da interface enquanto o contexto é mantido no restante do diagrama. Também empregada e comercializada pela Inxight Co. A *hyperbolic tree* possui uma versão de demonstração chamada Magnifind, que pode ser encontrada na página da Inxight (<http://www.inxight.com/>).

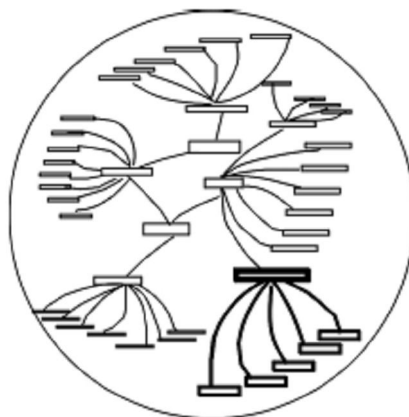


Figura 15: Esquema básico da técnica hyperbolic tree.

Fonte: Freitas et. al., 2001, p.154

Os itens mais distantes do foco central é menos detalhado que os nodos mais próximos. Os nós são representados por retângulos dispostos de maneira radial com foco no centro da estrutura.

Mantendo as subdivisões visíveis, a interface ganha a redução de sensação de perda de contexto por parte do usuário.

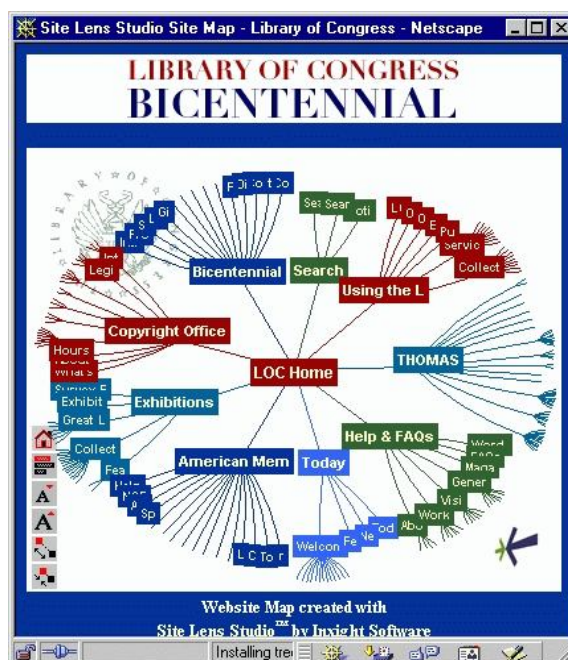


Figura 16: Modelo *Hiperbolic tree* apresentada pela Inxight

Fonte: http://www.cs.umd.edu/hcil/pubs/presentations/EyesHavelt-Oct2004_files/frame.htm

Uma das áreas de investigação mais promissoras baseadas nos estudos de Visualização da Informação É o Browsing Gráfico, mais especificamente os estudos com enfoque em representações de Interfaces Visuais para a Recuperação da Informação ou VIRIs (*Visual Information Retrieval Interfaces*).

Uma definição de VI aplicada às VIRIs é apresentada por Eick (2001), ao afirmar que a VI é “uma área de investigação enfocada na criação de interfaces visualmente ricas para ajudar o usuário a compreender e navegar através de espaços de informações complexos”.

Com base nessa definição podemos distinguir claramente duas funções de qualquer VIRI:

- Ajudar o usuário a compreender, proporcionando-o uma ‘grande imagem geral’ de todo o conjunto documental.
- Ajudar o usuário a navegar, possibilitando-o a localizar e recuperar documentos que satisfaçam suas necessidades informacionais através da exploração visual e a interação com a interface.

Como já colocado no capítulo anterior, o browsing como estratégia de acesso a informação pode, também, apresentar alguns inconvenientes (entre outros, a possibilidade de ocasionar ao usuário distração ao seu objetivo inicial (Marchionini 1995)). Herrero-Solana (2000) afirma, por este motivo, que os sistemas baseados no browsing não se colocam como substitutos de técnicas baseados no querying, mas sim técnicas complementares.

O autor Xia Lin (1997) apresentada algumas vantagens com a recuperação da informação através do browsing gráfico:

- A possibilidade de representar uma grande quantidade de informação em um espaço (visual) limitado
- A capacidade de revelar relações semânticas entre documentos e termos
- A facilidade de exploração visual e inferência perceptiva da interface

Marcos-Mora (2004) aponta o browsing gráfico como um meio fácil e cômodo devido sua naturalidade e a forma similar de buscar objetos no mundo

físico, e ainda propiciando a recuperação de informações de interesse de maneira fortuita.

4.

Produção Automatizada de Interfaces Visuais para Recuperação da Informação

Neste capítulo são tratados processos de criação de interfaces visuais para recuperação da informação e fatores relevantes no seu desenvolvimento. São descritas fases a serem levadas em consideração na escolha e implementação de modelos visuais. São também apresentadas e descritas três etapas fundamentais que compõe esquemas metodológicos propostos para a produção automatizada de interfaces visuais para a recuperação da informação.

4.1 Estabelecimento de modelos visuais para recuperação da informação

Interfaces visuais para recuperação da informação têm como objetivo apresentar ao usuário visões reduzidas e simplificadas de espaços de informação complexos (HERRERO-SOLANA E HASSAN-MONTERO, 2006). Construir essas ferramentas visuais é um complicado processo que sofre diversas interferências de variáveis devido à diversidade de possíveis estruturas visuais, objetos a serem visualizados, métodos de organização visual, abordagens de apresentações e controles de busca. Uma serie de fases devem ser consideradas ao estabelecer modelos visuais para recuperação da informação. Zhang (2008) divide em oito passos as etapas a serem consideradas onde podemos analisar e discutir em detalhes.

Passo 1: Determinação de um paradigma de visualização para recuperação da informação

Zhang (2008) descreve três possíveis paradigmas para a visualização de interfaces para recuperação da informação baseadas nos modelos *Browsing* e *Querying*. Apesar dos dois modelos possuírem características distintas, podemos sintetizá-los em ambientes visuais para RI aproveitando as vantagens de cada um deles (COX, 1992).

O primeiro paradigma apresentado por Zhang (2008) é o *QB (Querying and Browsing)*. Neste modelo é necessário que o usuário expresse uma questão de busca antes de visualizar alguma interface gráfica. O sistema então filtra uma área de interesse determinada pela questão e apresenta a interface já direcionada às informações específicas ao usuário.

O segundo paradigma apresentado é o *BQ (Browsing and Querying)*. Este modelo sugere que o usuário estipule sua expressão de busca já visualizando a interface. Feita a questão, os resultados relevantes são realçados no espaço visual revelando o contexto dos dados recuperados.

O terceiro e último paradigma descrito é o *BO (Browsing)*. Ou seja, é a aplicação pura do *Browsing*, onde não são necessárias expressões de busca. O usuário explora o espaço visual sem integrar os componentes do *Querying*. Vale destacar que o "Query Searching não é categorizado como um paradigma devido ele ser um paradigma tradicional da Recuperação da Informação onde não é necessário o uso de um espaço visual" (ZHANG, 2008, p.16).

Os paradigmas apresentados têm efeito direto na fonte e recuperação de dados brutos para visualização. O conjunto de dados é considerado a nascente para modelos *BQ* e *BO* e a recuperação de objetos nestes modelos é considerada estável. Para padrões *QB* um sistema de recuperação deve prover resultados para que o sistema seja usado como fonte de dados e a recuperação de seus dados brutos torna-se dinâmica devido a necessidade de constantes mudanças e reconstruções de configurações visuais da interface (COX, 1992).

Passo 2: Identificação das características dos dados

Zhang (2008) se refere à identificação destas características de dados como a tipologia dos objetos dispostos no espaço visual. Em uma coleção de dados há uma variedade de itens/peculiaridades no qual podem definir os objetos a serem apresentados na interface. Por exemplo: documentos, palavras-chave, periódicos, ou autores em uma base bibliográfica. A identificação destas características se torna altamente significativa para o conjunto de dados, usuários e a posterior recuperação da informação.

Passo 3: Extração de atributos

Um objeto pode ser descrito através de um grupo de atributos. Estes atributos podem não somente definir as propriedades de um elemento, mas também determinar sua posição em um espaço visual. Os atributos podem ser homogêneos (um grupo de palavras-chave, por exemplo), e também, heterogêneos (como data de publicação, autor, título, etc.). Eles devem, ainda, ser coerentes com todo conjunto semântico do ambiente visual para recuperação da informação e mensuráveis, permitindo expressá-los em formas quantitativas para posterior cálculo.

O método de aplicação de atributos no ambiente visual varia em diferentes situações. Lin (1997) nos lembra que estes atributos podem, não somente, ser aplicados de forma direta aos espaços visuais para recuperação, mas também de forma indireta, na qual são convertidas em formas significativas e aplicadas indiretamente.

Passo 4: Definição estrutural de um espaço visual

A definição estrutural de um espaço visual refere-se à determinação da dimensionalidade de um espaço visual e a definição de eixos para a coordenação do sistema. A dimensionalidade de um espaço visual pode ser em uma dimensão, bi-dimensional ou tri-dimensional. A definição estrutural permite reduzir e visualizar, ainda que com distorções, um espaço multidimensional às dimensões sensíveis a nossa percepção, tornando evidentes relações inicialmente ocultas (MOYA-ANEGÓN e HERRERO SOLANA, 1999).

Passo 5: Definição de um espaço visual semântico

A definição de um espaço visual semântico é vital e essencial devido a seu papel em definir a estrutura em que os elementos serão projetados: agregam-se informações, decidem-se padrões, estruturas internas são demonstradas e as interações observadas. A definição do espaço visual irá determinar a área onde todos os elementos e objetos serão projetados. Isto pode variar de simples formas a interfaces complexas com linhas, círculos, arvores, etc. (ZHANG, 2008, p.18). Robertson (1991) apresenta usos de metáforas para facilitar a compreensão de usuários não habituados com modelos visuais como, por exemplo, apresentações com metáforas de paisagens, sistemas solares, salas, etc. “Escolhendo representações apropriadas podemos solucionar uma difícil apreciação de dados, beneficiando análises subseqüentes” (ROBERTSON, 1991, p.69).

Passo 6: Projeção dos objetos

A projeção de dados deve ser considerada a parte central de todo o procedimento. Isto determina a posição final de cada objeto no espaço visual e conseqüentemente a configuração final da representação, pois a projeção dada pelos algoritmos influencia diretamente no espaço visual semântico e sua estrutura. As projeções dos objetos têm sua complexidade ligada diretamente ao tipo de modelo da interface visual para recuperação da informação (ZHANG, 2008).

Como já sabemos, um objeto pode ter várias facetas na qual definem um espaço informacional multidimensional. Quando o elemento é projetado dentro de um espaço visual, somente as facetas significantes são apresentadas na projeção devido a dimensionalidade limitada da interface. Neste sentido, devido a esta redução, a projeção deve ser considerada o reflexo de uma parte das relações entre os dados em detrimento de várias outras possíveis apresentações (LIN, 1997).

Passo 7: Desenvolvimento de recursos interativos para recuperação da informação

Os recursos interativos para recuperação da informação devem facilitar o trabalho das representações visuais, ou seja, ajudar na exploração e descoberta de conhecimentos. Com estas ferramentas interativas, usuários podem pesquisar as informações do conteúdo de um determinado objeto, um contexto local de uma área de interesse, e uma visão global do conjunto de dados. No *Querying*, os recursos interativos deveriam ser integrados no ambiente visual da recuperação da informação para satisfazer a necessidade de busca de informações específicas.

Novos modelos de recuperação de informação podem ser desenvolvidos com base em espaços semânticos e definições estruturais de interfaces visuais (ZHANG, 2008).

Passo 8: Avaliação

O último passo é a avaliação das interfaces visuais para recuperação da informação. A avaliação irá examinar se: os objetos, seus atributos extraídos, o sistema de coordenação, seu design e os recursos interativos estão coerentes e continuados no ambiente visual; se os dados estão dispostos adequadamente, clara e precisamente; se as apresentações são significativas e interpretáveis; e, por fim, se os recursos interativos estão bem integrados com a representação (ZHANG, 2008).

4.2 Desenvolvimento de Interfaces Visuais para Recuperação da Informação

Interfaces visuais para recuperação tentam, a priori, demonstrar tanto o conteúdo quanto a estrutura do espaço informacional através do mapeamento dos principais conceitos de um objeto (LIN, 1997). Diversos esquemas de produção destas interfaces envolvem três etapas ou fases que compõe diversos esquemas metodológicos propostos na literatura científica (CHUNG, *ET AL.* 2003; POLANCO E ZARTL 2002; TURETKEN E SHARDA 2003; BÖRNER *ET AL.* 2003): Análise e transformação de dados; Aplicação de

algoritmos de classificação e distribuição visual; e Aplicação de técnicas de transformação visual. No estágio Análise e transformação de dados, recursos essenciais de uma coleção são analisados e tabulados a fim de extrair conhecimentos, inicialmente herméticos, dos objetos e do espaço informacional. O estágio de Aplicação de algoritmos de classificação e distribuição visual se compromete a gerar uma metáfora visual que reflita visualmente a estrutura de dados gerada na etapa anterior, representando dados multidimensionais em estruturas de duas ou três dimensões. Na última etapa, Aplicação de técnicas de transformação visual, a estrutura visual é apresentada ao usuário e técnicas de interação e de redução de sobrecarga visual são aplicadas.

Descrevemos a seguir os processos utilizados nas três etapas de criação de interfaces visuais para recuperação da informação:

4.2.1 Análise e Transformação de Dados

O primeiro passo na visualização de espaços complexos de informação é a indexação automática de um conjunto de documentos eletrônicos, muito similar aos sistemas baseados no modelo vetorial de Salton (1989).

É importante lembrar que os sistemas baseados no Browsing têm, além do objetivo de possibilitar uma classificação automática do conjunto de documentos (como o Querying, que compara o conjunto com a busca realizada), como também de descobrir relações subjacentes entre os elementos deste conjunto.

Herrero-Solana e Hassan-Montero (2006) ressaltam ainda que quando documentos eletrônicos contêm informações complementares, como hipertextos ou etiquetas de marcação, podem também ser vetorizadas para posterior aproveitamento.

Tratado o conjunto documental, é possível uma série de análises com o objetivo de descobrir relações entre documentos, assim como entre seus termos.

Entendendo estas diferentes técnicas como métodos de extração de conhecimentos imprevistos, podemos claramente envolvê-las em um

importante campo de pesquisa nas áreas de Indexação, Recuperação da Informação e Banco de Dados: a Mineração de Dados (*Data Mining*). *Data Mining* pode ser definido como um processo de descobrimento de conhecimento (a priori desconhecido) em um conjunto de dados complexos por meio de informações potencialmente úteis, “ocultas” em padrões e relações estruturais implícitas (Kopanakis e Theodoulidis 2003).

Podemos ainda diferenciar a Mineração de Dados em três distintos tipos: mineração de conteúdo; de estrutura e de uso. Baeza-Yates (2004) os identifica abaixo na figura 17, onde tanto usuários como softwares estão envolvidos na geração ou extração de dados.

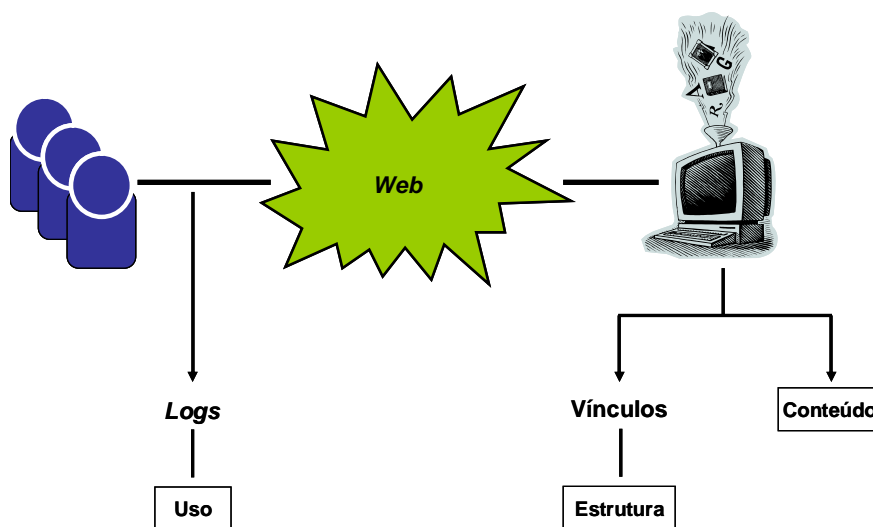


Figura 17: Tipos de minerações na WEB

Fonte: Adaptado de BAEZA-YATES 2004

Mineração de Conteúdo

Nasukawa e Nagano (2001) definem a mineração de conteúdo (por eles tratada como mineração de textos) como o “encontro, no texto, de padrões e regras úteis que indicam tendências e características significativas sobre assuntos específicos”. Podemos identificar relações semânticas entre textos que compartilhem mesmos termos. O que conhecemos como co-ocorrência.

Herrero-Solana e Hassan-Montero (2006) levantam a possibilidade de calcular a semelhança entre dois documentos comparando os vetores de cada

objeto através de funções de similaridade (explicadas no capítulo 2). “Igualmente é possível calcular a similaridade entre termos, tendo em conta que a ocorrência de dois termos em um mesmo documento implica uma relação semântica entre estes”. (HERRERO-SOLANA E HASSAN-MONTERO, 2006).

Diversos estudos buscam formas de assegurar a consistência da indexação automática de textos em linguagem natural. É necessário levar a cabo um controle terminológico antes de contabilizar freqüências de co-ocorrência. Gil Leiva (1996) destaca algumas técnicas de controle para indexação automática:

- Eliminação de termos mediante uma lista de palavras vazias previamente definidas.
- Uso de procedimentos de *stemming* para reduzir o conjunto de termos, identificando raízes terminológicas, eliminando duplicatas, variáveis de gênero, número, etc.
- Eliminação de termos com freqüências demasiadamente altas ou excessivamente baixas.

Mineração de Estrutura

Na mineração de estrutura é preciso lembrar-se de sua alta complexidade e de sua constante evolução do conjunto documental. A Estrutura de dados digitais pode ser usada por buscadores a fim de hierarquizar páginas mais citadas utilizando algoritmos de *Pagerank* (BRIN; PAGE, 1998) como o usado no buscador *Google*. A mineração de estrutura serve ainda para encontrar grupos que se apontam e representam determinada comunidade, conhecido em alguns casos como *Clusterings*.

A Mineração de estrutura não se refere a estrutura própria de cada documento de forma isolada. Ela trata da estrutura do espaço documental definido pelos vínculos entre seus documentos.

Quando um documento cita outro, cria-se um vínculo expresso em uma relação estrutural explícita entre os dois documentos. Porém através desta

relação é ainda possível descobrir estruturas implícitas e subjacentes de relações entre os documentos.

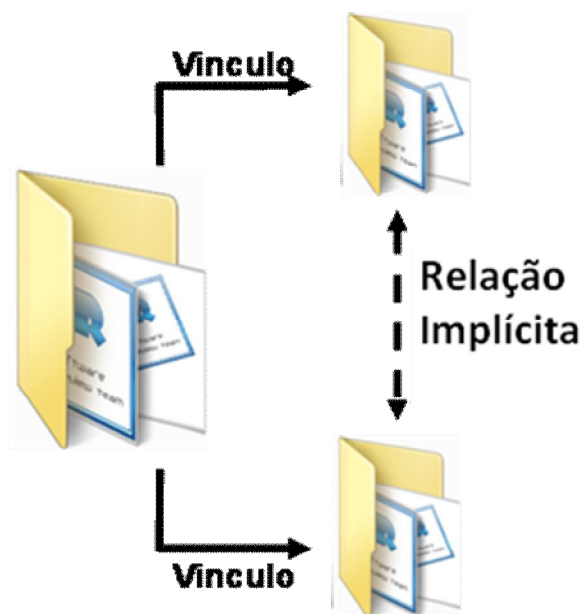


Figura 18: Esquema de co-citação entre documentos

Herrero-Solana e Hassan-Montero (2006) citam o caso da análise de co-citação, termo que se refere à citação através de hipertextos entre páginas *WEB*. A idéia é que quando determinado documento aponta dois documentos (ou mais) em sua estrutura ele também está apontando uma relação semântica intrínseca e subjetiva entre estes.

Igualmente, os autores sugerem a análise de citações em comum feitas por dois documentos distintos. É o que a Bibliometria trata como Acoplamento Bibliográfico (*Bibliographic Coupling*). O processo evidenciaria estruturas de similaridade subjacentes entre os documentos "citantes".

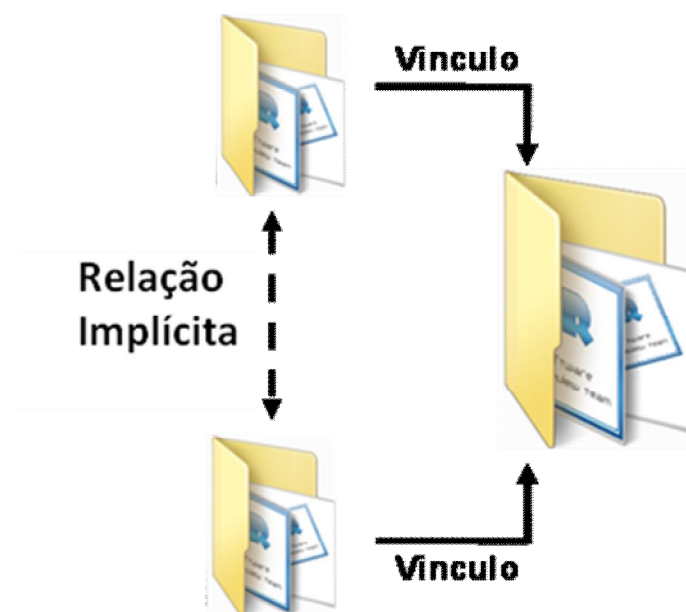


Figura 19: Esquema de Acoplamento Bibliográfico entre documentos

Mineração de Uso

Chegando à mineração de uso, apontamos como fundamental neste método, a análise de *logs*. Analisar arquivos de acesso ou *logs* de um servidor foi, primeiramente, de grande interesse ao setor comercial, que consideraram onde deveriam aplicar maiores investimentos de acordo com os setores mais acessados em sua página *WEB*.

Quando o espaço documental está acessível, a informação registrada no servidor pode nos oferecer informações relevantes acerca da relação semântica entre os elementos do banco de dados. Por exemplo, quando um mesmo usuário "co-visita" ou "co-acessa" dois documentos diferentes, ele estabelece uma relação entre os dois documentos. A intensidade da relação pode ser medida de acordo com a frequência deste padrão de uso. Hassan-Montero (2006) sugere ainda a análise da seqüência em que estes documentos são visitados.

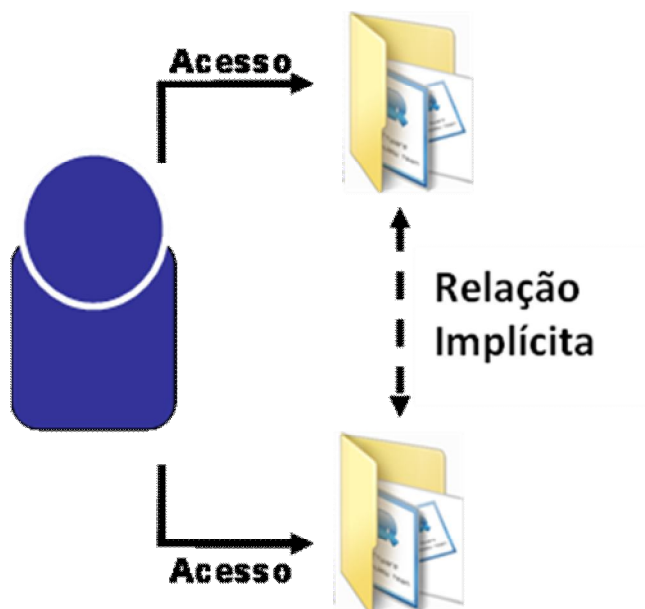


Figura 20: Esquema de Co-acesso entre documentos

Estes diferentes tipos de análises oferecem-nos uma ferramenta capaz de revelar relações semânticas de similaridade entre os documentos e termos, com base em um espaço documental. A informação oferecida deve ser vista como complementar, e sua utilização conjunta aumentaria a capacidade para revelar informações “ocultas”.

Portanto, apesar de normalmente serem utilizadas individualmente, sua utilização conjunta aumentaria sua capacidade de expor dados abstratos. Neste sentido Chen (1998) propõe um modelo genérico para estruturar e visualizar espaços de informações hipertextuais, denominado GSA (*Generalised Similarity Analysis*). O modelo oferece um esquema metodológico para a extração de relações semânticas com base nos três tipos de medidas de similaridade: relações hipertextuais (estrutura), similaridade de conteúdo e padrões de uso. Todas unificadas em uma única função, que o autor denomina de meta-similaridade.

4.2.2 Aplicações de Algoritmos de Classificação e Distribuição Visual

Em sua tese de doutorado, Herrero-Solana (2000) propôs uma classificação de diferentes interfaces visuais para recuperação da informação em função da estrutura explícita dos dados a serem representados:

hierárquicos, em rede, de busca, em linhas temporais e multidimensionais. A utilização de metáforas visuais fiéis à estrutura de dados das interfaces não resultam necessariamente em aplicações de técnicas de classificação e distribuição visual.

É imprescindível basear estas técnicas sobre estruturas de dados multidimensionais. Essas técnicas produzem de forma automática uma abstração gráfica, resultado de uma redução de um espaço de n -dimensões a um número de dimensões compreensíveis e perceptíveis ao olho humano (1D, 2D e 3D).

Nesta seção, abordaremos os diferentes tipos de técnicas e algoritmos em função da metáfora visual estudada em questão e não sua estrutura de dados, pois entendemos que sempre serão estruturas dimensionais que necessitam ser reduzidas ou simplificadas.

Seguindo o esquema de Lin (1997), podemos diferenciar entre os seguintes tipos de metáforas visuais: hierárquicas, de redes, de dispersão e mapas.

Representações Hierárquicas

Nas representações hierárquicas, os elementos se apresentam em diferentes níveis, ramificações ou agrupamentos, que descendem de um nó (nodo) raiz – é uma representação comum de dados que tem em sua própria natureza uma visualização hierárquica, como por exemplo, uma visualização de estruturas complexas de diretórios de arquivos de sistemas operacionais gráficos. Essa estrutura é representada na Teoria de Grafos como uma árvore. No caso da visualização de dados multidimensionais, onde não estão definidas de forma explícita as relações hierárquicas entre elas, as representações hierárquicas se utilizam de aplicações de técnicas de agrupamento.

Este é o caso de técnicas estatísticas de *clustering* ou análise de conglomerados, que através de um processo interativo são agrupados diferentes elementos em função de sua similaridade, formando assim grupos em diferentes níveis hierárquicos.

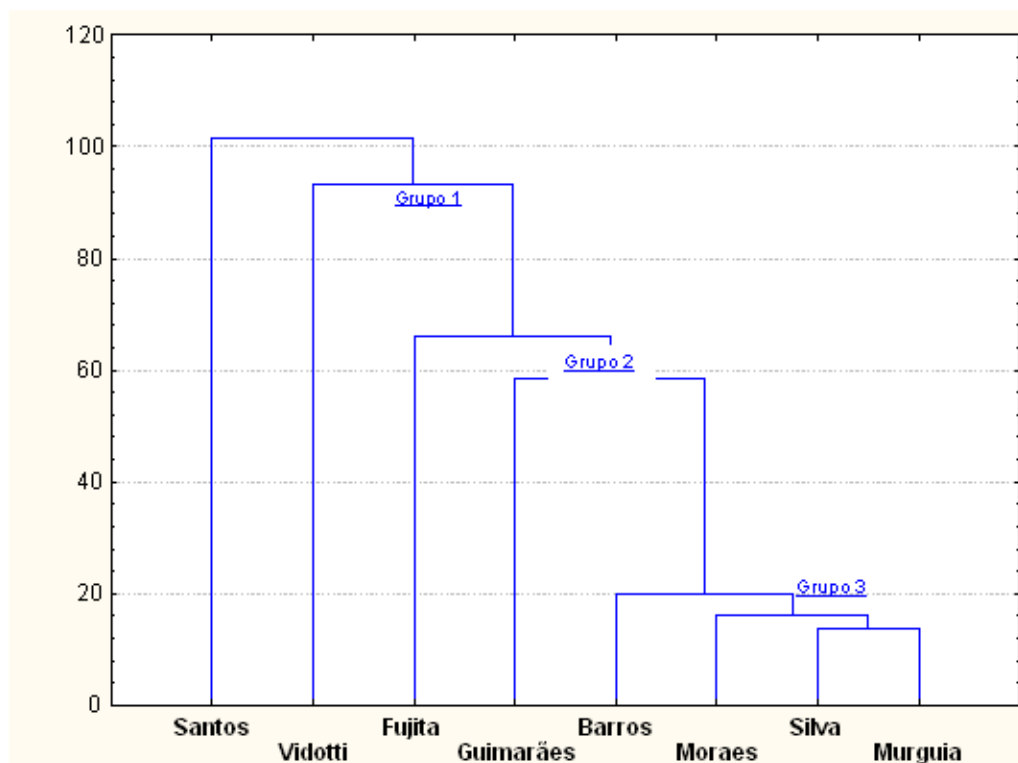


Figura 21: Aplicação de *Clusterings*

Fonte: XAVIER et. al, 2006

Entre as técnicas de clustering é possível diferenciar em torno de 150 tipos diferentes de regras de aglomeração (HERRERO-SOLANA, 2000). Dentre as mais comuns encontramos: encadeamento simples (*single link*); encadeamento completo (*complete link*), encadeamento médio; e o método de Ward, também conhecido como método da soma dos quadrados (HERRERO-SOLANA, 2000).

Os resultados gráficos de aplicações de técnicas de Clustering geralmente são apresentados em forma de dendogramas (uma representação arbórea hierárquica), onde é possível visualizar diferentes grupos e subgrupos dentro da interface. Para seu uso na produção de interfaces visuais para recuperação da informação, os *clusterings* se colocam como metáforas hierárquicas alternativas, como as mostradas no capítulo anterior (hiperbólicas (LAMPING et al, 1995) e Treemaps (JOHNSON, 1991)).

Representações em Rede

Representações em rede são aquelas em que diferentes elementos (documentos, termos, etc.) são apresentados em forma de nós ou vértices, de maneira que a estrutura, algumas vezes semântica, se encontra definida pelas relações ou arcos de proximidade de cada nó em forma de teia, não possuindo uma relação de hierarquia.

O problema ao representar dados multidimensionais desta forma é que o número de nós e vértices pode chegar a tal tamanho que seria incompreensível o grafo resultante.

É necessário, portanto, o emprego de técnicas de redução de relações a fim de que a interface final seja compreensível e útil para a visualização, sem perder ou distorcer a realidade estrutural da rede. Uma destas técnicas de redução é o método de escalonamento de rede Pathfinder (SCHVANEVELDT,1990), apresentando redes denominadas Pathfinder ou PFNETs.

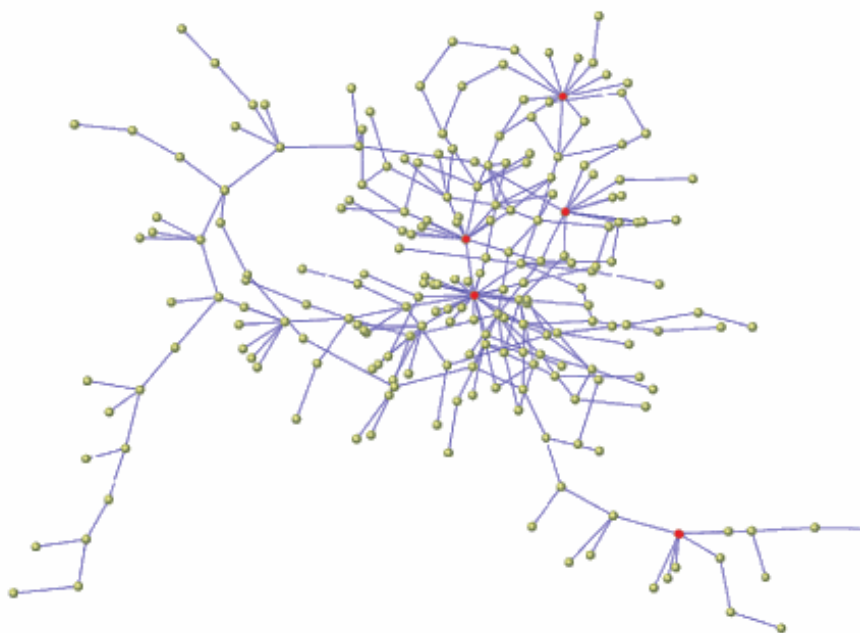


Figura 22: Representação em *Pathfinder*

Fonte: SCIMAGO, 2009

Uma vez reduzido o número de relacionamentos para a representação visual da rede, é necessário aplicar algum algoritmo de posicionamento dos nós. Mesmo que técnicas de redução dimensional pareçam a princípio úteis, há uma desvantagem a ser considerada na aplicação nos PFNETs: o posicionamento final dos nós provocaria cruzamento nos vértices de relacionamento, obscurecendo toda a visualização do gráfico (FOWLER et al, 1992).

Representações de Dispersão

Uma forma alternativa de representar visualmente estruturas de dados multidimensionais é o agrupamento de pontos e sua dispersão. Estes pontos estariam distribuídos em um espaço visual bi ou tridimensional, e distanciados um dos outros através das dissimilaridades e distâncias especificadas em uma matriz de similaridade.

Para poder apresentar essas distâncias em dimensões compreensíveis pelo ser humano, são necessárias técnicas de redução de dimensões. Este é o caso da técnica estatística multivariante denominada Escalonamento Multidimensional ou MDS (Multidimensional Scaling). Este método consiste de um processo interativo no qual vão se recolocando e re-posicionando diferentes elementos (documentos, termos, palavras-chave, autores, etc.) para analisar a distância entre eles.

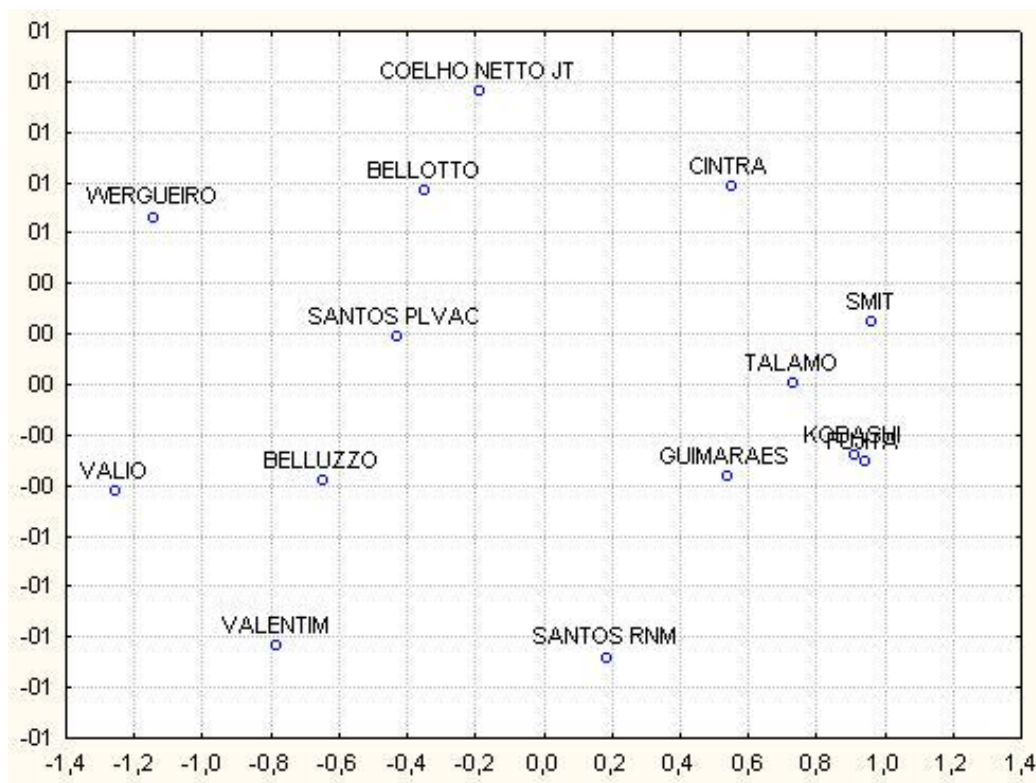


Figura 23: Aplicação de uma representação MDS

Fonte: XAVIER, 2006

Inevitavelmente, existe uma perda de informação de acordo com o número de dimensões do espaço multidimensional original e o número de dimensões ao qual se pretende reduzir.

Mapas

As representações visuais baseadas em mapas se fundamentam na idéia de utilizar da metáfora de mapas geográficos para a visualização de espaços de informação. De forma geral, o objetivo da utilização de qualquer tipo de metáfora visual em um desenho de interface é fazer visível para o usuário a estrutura e relação de um determinado conjunto de dados. Portanto, a idéia de utilizar este tipo de metáfora para visualização de espaços de informação complexos e abstratos faz todo sentido, já que oferece uma visão diferente do conjunto e que na maioria dos casos enriquecerá a imagem mental prévia que o usuário possua.

Dentre todas as técnicas possíveis para gerar este tipo de mapa, vale ressaltar o modelo de mapas auto-organizáveis ou SOM (*Self-Organizing Map*) (KOHONEN, 1989). O SOM são aplicações de Redes Neurais Artificiais (RNA) para organização e classificação automática da informação.

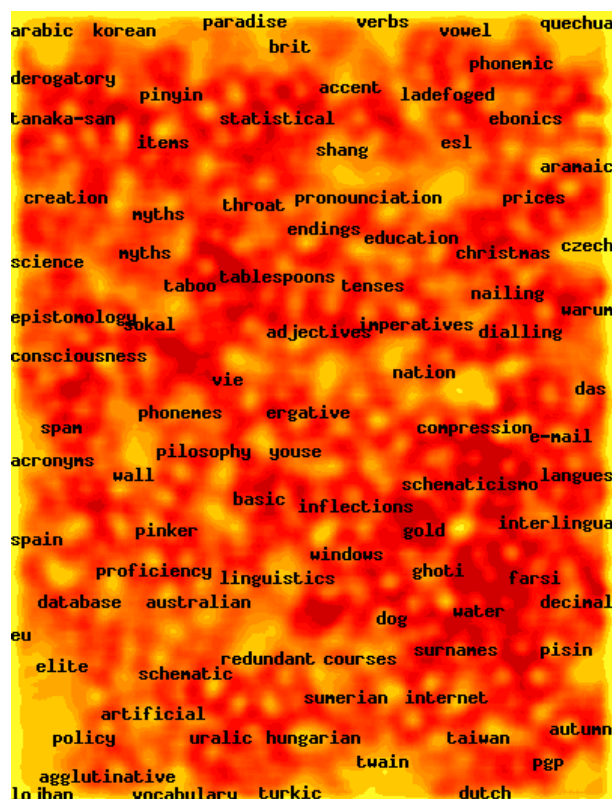


Figura 24: Análise de co-ocorrência baseado no WebSOM

Fonte: <http://websom.hut.fi/websom/>

O funcionamento deste modelo foca em estabelecer uma ligação entre a informação de entrada e um espaço de saída de duas dimensões. Desta maneira, os dados de entrada com características comuns formarão zonas próximas no mapa. Este modelo de rede neural representa-se de maneira bidimensional. Quando ocorre a entrada de um novo vetor de dados à rede, este deve se relacionar com seu vetor “concorrente” de forma que somente um vetor de saída seja ativado. Este vetor é denominado vencedor (*winner-take-all unit*) e determina um ponto no mapa bidimensional.

A rede neural classifica a informação de entrada, já que o vetor vencedor a representa. Além disso, entradas semelhantes ativam sempre o mesmo vetor. Portanto podemos utilizar esta rede neural para estabelecer relações previamente desconhecidas em um conjunto de dados. Trata-se de um processo de auto-organização onde a classificação ou organização estão condicionadas à própria natureza dos dados introduzidos.

Os diferentes métodos descritos, classificados em função de seu tipo de representação visual, podem ser considerados complementares, já que oferecem diferentes soluções para um mesmo problema (HERRERO-SOLANA, 2000). Apesar de funcionamento e procedimentos diferentes, todos compartilham o objetivo de reduzir espaços multidimensionais às formas “visualizáveis” em duas e/ou três dimensões.

Em qualquer redução, simplificação ou resumo de conjuntos complexos de dados, perde-se informações da realidade estrutural dos dados originais. É precisamente nas informações preservadas mais fielmente que podemos caracterizar e comparar os métodos apresentados.

Kaski (1997) compara o MDS e o SOM afirmando que enquanto o primeiro se orienta preservando sua estrutura através das distancias entre os elementos, os mapas auto-organizados tencionam sua estrutura resguardando sua topologia ou relações de proximidade.

Por outro lado, as técnicas de escalonamento Pathfinder geram representações em forma de grafos, onde o que se preserva e enfatiza são as relações locais entre elementos, representada mediante vínculos ou arcos entre os nós (CHEN, 1998).

As informações preservadas mediante as técnicas de Clustering são aquelas representadas por relações de elementos entre grupos. Apesar de se mostrarem úteis como técnicas de classificação e visualização da informação, este modelo não parece ser utilizado na mesma freqüência do que as outras aplicações. Herrero-Solana e Hassan-Montero (2006) acreditam que isso ocorra em parte pela impossibilidade de rotular ou etiquetar automaticamente cada um dos grupos ou subgrupos criados, onde somente o usuário poderia identificar a classe ou grupos através da exploração de seus elementos. Por

isto, o Clustering, no contexto das interfaces visuais para recuperação da informação, deveria ser considerado mais útil como complemento aos outros modelos (MDS, SOM, Pathfinder), do que como solução por si só.

A Figura 25 mostra os modelos de aplicação e distribuição visual apresentadas acima e quais estruturas são preservadas por cada uma:

Técnica	Elementos preservados pela estrutura
MDS	Distância entre os elementos
<i>Clustering</i>	Relações de agrupamento
<i>Pathfinder</i>	Relações locais entre os elementos
SOM	Relações de proximidade

Figura 25: Estruturas preservadas por cada técnica

Fonte: HERRERO-SOLANA E HASSAN-MONTERO, 2006

A preservação de um determinado elemento estrutural (e portanto a perda de outra informação), pode condensar a expressividade da interface. Ao mesmo tempo a escolha de uma ou outra técnica condicionará a efetividade da interface, a facilidade de sua interpretação e sua compreensão por parte do usuário. Bozydowski (2003) lembra que o grau de expressividade de cada técnica está fortemente influenciado pela natureza dos dados. Por isso, não podemos afirmar a superioridade de nenhuma delas em todos os contextos.

Entretanto, a efetividade e facilidade de interpretação da metáfora visual pode ser determinada independente da natureza dos dados. A natureza de grafo do *Pathfinder* resulta maior familiaridade com o que encontramos hoje na *Web*, e por isso possa ter um nível maior de usabilidade do que os modelos MDS ou SOM. Estes modelos podem parecer, a principio, tão incomuns para seus utilizadores que a ordem dos elementos SOM pode ser uma vantagem sobre o aparente caos na dispersão dos elementos MDS.

A comparação entre estes métodos pode se realizar com base em outras variáveis igualmente importantes, como: escalabilidade, carga computacional possibilidade de interpretação de dimensões, distribuição visual (estática ou dinâmica), e escala (global ou local) (BÖRNER *et al.* 2003).

4.2.3 Aplicações de Técnicas de Transformação Visual

Apesar das técnicas de classificação e distribuição visual descritas até o momento buscarem possibilitar a visualização do espaço documental de um conjunto de dados, nem sempre estas interfaces são fáceis de explorar visualmente devido ao seu tamanho e complexidade, provocando o que Herrero-Solana e Hassan-Montero (2006) chamam de ‘sobrecarga visual’.

Devido a essa sobrecarga, é necessário prover as interfaces de mecanismos dinâmicos e interativos que facilitem a visualização e exploração do usuário.

Estas técnicas são denominadas transformações visuais interativas (HERRERO-SOLANA E HASSAN-MONTERO, 2006; GUTWIN E FEDAK, 2004; LEUNG E APPERLEY, 1994). Suas investigações são anteriores aos estudos da VI, surgindo a partir da necessidade de apresentar muitos dados em uma tela de forma simultânea.

Gutwin e Fedak (2004) classificam estes mecanismos em quatro grupos: *panning*, *zooming*, visões detalhadas e *focus+context*. Herrero-Solana e Hassan-Montero (2006) consideram os três primeiros grupos como técnicas não pautadas a distorção enquanto que o grupo *focus+context* englobaria estruturas norteadas pela distorção visual.

Técnicas não Pautadas na Distorção Visual

As técnicas de *panning* consistem em oferecer ao usuário um mecanismo para “rolar a tela”, enquanto que o *zooming* oferece a opção de ampliar ou ver em detalhe uma parte da interface. Storey *Et al.* (1999) lembra que a utilização individual destes dois modelos pode trazer problemas de orientação do usuário, já que particionam o conjunto, impedindo a visão global de toda interface (HERRERO-SOLANA E HASSAN-MONTERO, 2006).

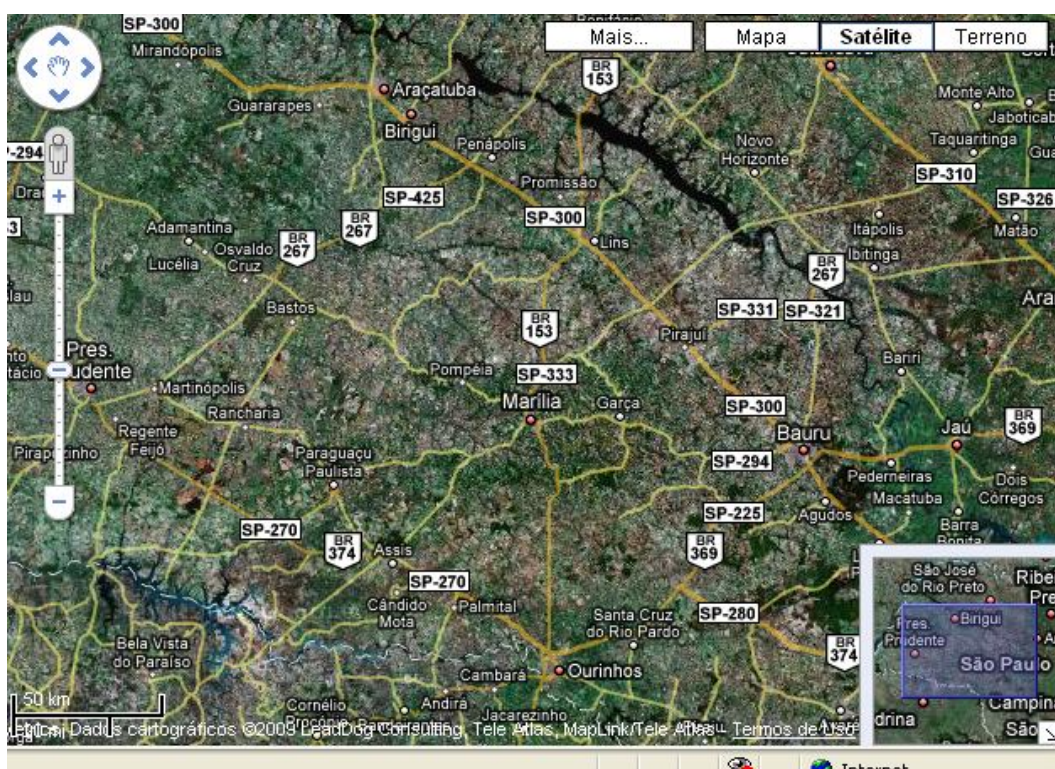


Figura 26: Aplicação da técnica *detail+overview* no Google Earth

Fonte: <http://earth.google.com/intl/pt/>

Storey *Et al.* (1999) aponta como possível solução a este problema o emprego de técnicas de visões detalhadas (*detail+overview*). Este modelo oferece simultaneamente mais de um ponto de vista da mesma representação. Ao mesmo tempo é apresentada a visão global para que o usuário se oriente e uma visão em detalhes para exploração. Através do *detail+overview*, é possível no nível detalhado, usar as técnicas de *zooming* e *panning* ajudando na redução de sobrecarga visual

Estruturas Norteadas pela Distorção Visual

Polanco e Zartl (2002) apontam as estruturas norteadas pela distorção visual, mais especificamente o *Focus+Context*, como não somente uma alternativa às visões detalhadas, mas, em alguns contextos, um complemento a ele.

A idéia consiste em oferecer ao usuário, através de uma única apresentação, tanto uma visão detalhada da representação, como uma visão global. As premissas são as seguintes (POLANCO E ZARTL 2002):

- O usuário necessita visualizar tanto a informação global (contexto) como a informação em detalhe (o foco de sua visualização) de forma simultânea
- A informação que necessita visualizar em detalhe é diferente da que necessita para contextualizá-la
- Estes dois tipos de informação devem ser combinados em uma única apresentação

De acordo com essas premissas, o funcionamento do *Focus+Context* consiste em distorcer a representação visual, tornando mais visível o foco e minimizando o contexto. Para isso Herrero-Solana e Hassan-Montero (2006) apontam os seguintes procedimentos:

- Eleição da informação visual; Ocultação das partes da estrutura visual de menor interesse no momento (até que necessitem ser visualizadas).
- Distorção da informação visual: alteração das posições, formas e tamanhos dos elementos visuais.

Dentre os diversos modelos de distorção visual bifocal, criados através do *Focus+Context*, o efeito visual mais aplicado com sucesso é o olho de peixe (*fisheye*).

Furnas (1986) definiu uma função chamada de grau de interesse (*Degree of Interest*) para formalizar a visão em “olho de peixe”. Para cada ponto da estrutura a ser visualizado é calculado um valor da função de grau de interesse. É escolhido um limite e os pontos onde o valor da função seja inferior a este limiar são eliminados.

Existe uma variedade de aplicações baseadas no conceito olho de peixe que são diferentes tanto em seu conceito de aplicação, como também em sua forma e método. Noik (1993 *apud* HERRERO-SOLANA E HASSAN-

MONTERO, 2006), afirma que a proposta original de Furnas estava focada principalmente na utilização de filtragem e eleição, enquanto que muitas outras aplicações posteriores estavam baseadas em procedimentos exclusivamente de distorção visual.

Herrero-Solana e Hassan-Montero (2006) apresentam ainda alguns modelos aplicados com o olho de peixe como os menus de interface gráficas de usuários (BEDERSON, 2000), documentos textuais de estruturas hierárquicas (HAYAMA *et al.* 2003) (HORNBAEK & FRØKJÆR, 2003) e hierarquias complexas (LAMPING *et al.* 1995).

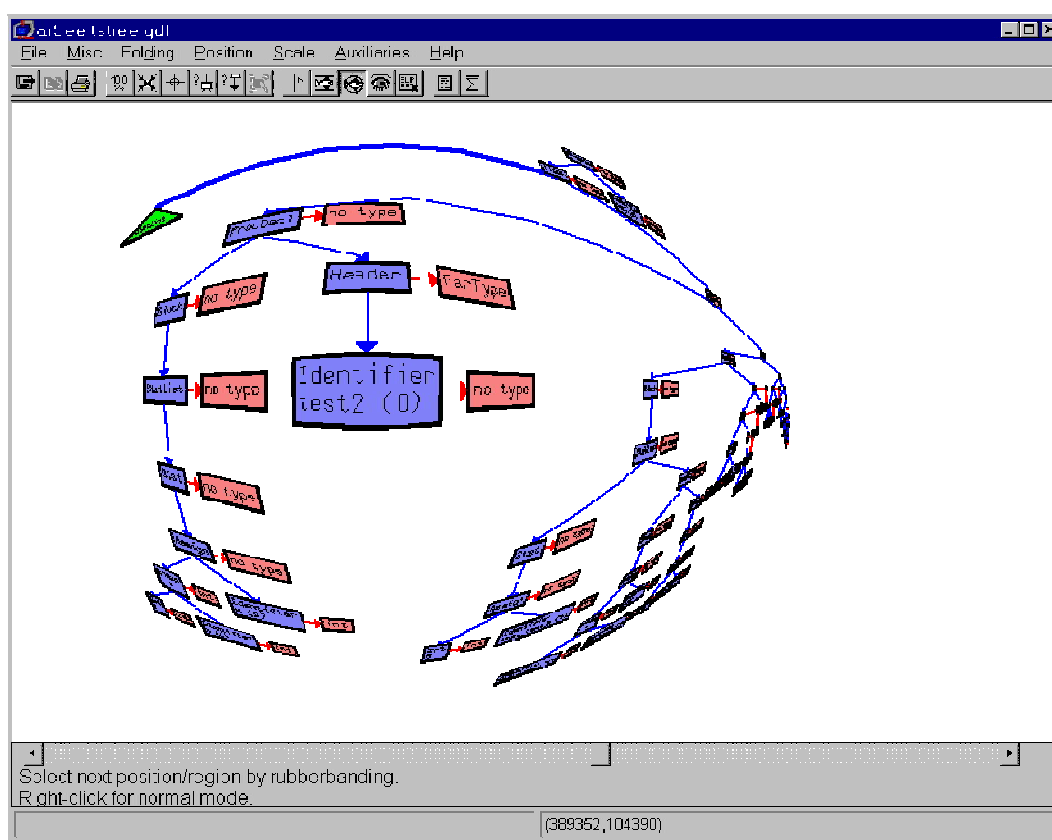


Figura 27: Aplicação da visão olho de peixe (Fisheye View) com a ferramenta “aiSee”

Fonte: <http://www.absint.com/aiSee/download/>

Polanco e Zartl (2002) lembram que um aspecto fundamental na aplicação destas técnicas é a necessidade de preservação do mapa visual adquirido pelo usuário antes da distorção da interface. O usuário deve ser

capaz de comparar mentalmente a representação distorcida com a visualizada antes da transformação visual.

Podemos destacar ainda algumas aplicações onde a técnica olho de peixe tenha sido aplicada às interfaces visuais para recuperação da informação de forma mais consistente, são elas: (FOWLER *et al.* 1992; KOIKE 1993; RENNISON 1994; CHEN 1998; TURETKEN & SHARDA 2003 e YANG, *et al.* 2003).

Podemos perceber que técnicas de *zooming* e *panning* são úteis em alguns contextos (ao não proporcionar uma visão global da representação), porém, encontram certa dificuldade ao proporcionar transformações em representações visuais complexas. Além disto, conforme Gutwin e Fedak (2004) e Herrero-Solana e Hassan-Montero (2006), as técnicas de *detail+overview* e *Focus+Context*, apresentam uma maior efetividade aplicada nas interfaces.

No entanto, a *detail+overview*, apesar de amplamente utilizada, apresenta alguns problemas como a discrepância visual entre a representação detalhada e a visão global (STOREY *et al.* 1999), dificultando a relação da interface visualizada em detalhes com seu contexto. Yang (*et al.* 2003) também considera a possibilidade de que a abstração de ambas representações tornem as transformações visuais ainda mais herméticas para usuários aumentando ainda mais sua sobrecarga cognitiva. É por estes motivos que Herrero-Solana e Hassan-Montero (2006) defendem o uso do *zooming* e *panning* em determinados contextos, diminuindo a sobrecarga que a *detail+overview* ocasiona.

Afirmar que dentre as técnicas aqui estudadas a *Focus+Context* se mostra mais adequada às interfaces visuais para recuperação da informação ainda não é possível. Mesmo que em tese a técnica pareça resolver diversos problemas de sobrecarga cognitiva, Storey (*et al.* 1999) lembra que há uma enorme carência de estudos de análises e avaliações empíricas que determinem a validade de cada um destes modelos.

4.3 Análise dos métodos

Ao longo deste capítulo, descrevemos, analisamos e comparamos possíveis técnicas e métodos empregados em diferentes etapas de construção e produção de interfaces visuais para recuperação da informação, porém ainda nos cabe questionar quais possíveis combinações entre técnicas, nas diferentes etapas, oferecem melhores resultados.

Observando a etapa de análise e transformação de dados, concluímos que as diferentes técnicas de mineração de dados apresentada, poderiam ser consideradas complementares e deveriam ser usadas de forma conjunta através de análises de meta-similaridade.

Com as etapas de ‘classificação e distribuição visual’ e ‘transformação visual’, é necessária a verificação de compatibilidade de cada método. A fase de transformação visual é necessária quando a representação produzida na etapa anterior (classificação e distribuição visual) se torna complexa demais e de difícil visualização. É o caso do MDS, que normalmente apresenta em sua representação visual os elementos demasiadamente aglomerados e sobrepostos, necessitando o emprego de técnicas interativas de transformação visual.

	MDS	SOM	Pathfinder
<i>Focus+Context</i>	Combinação adequada, porém com ressalvas	Combinação adequada	Combinação adequada, porém com ressalvas
<i>Overview+Detail</i>	Combinação inadequada	Combinação adequada, porém com ressalvas	Combinação adequada
<i>Zoom+Panning</i>	Combinação adequada	Combinação adequada	Combinação adequada

Figura 28: Combinações entre técnicas de desenvolvimento de interfaces visuais para recuperação da informação

Nas redes *Pathfinder* a sobreposição de elementos é menor, fazendo-se necessária a aplicação de distorção visual somente se o número de nós representados for demasiadamente grande.

Já as representações em mapas, mais especificamente o SOM, parecem aproveitar bem o espaço da interface, ficando livres de grandes distorções em suas representações.

Herrero-Solana e Hassan-Montero (2006) lembram que deve-se considerar que a escolha da técnica de distribuição visual interferirá diretamente na aplicação de técnicas de transformação visual. As técnicas de distorção são igualmente adequadas e compatíveis com todos os modelos de distribuição visual.

As técnicas de *zooming* e *panning*, por ter fácil aplicação e não causar distorção na interface podem ser aplicadas em qualquer representação visual.

Entretanto, o *detail+overview* não deveria ser considerado em representações MDS. É comum que estas representações não possuam estruturas globais claras e reconhecíveis, fazendo com que a técnica não contribua para orientação dos usuários. O *detail+overview* pode ser muito melhor aplicado em representações *Pathfinder*, por possuírem uma estrutura contextual melhor reconhecível através de seus arcos e nós. A combinação do *detail+overview* com representações em mapas não é necessariamente inadequada, seu aproveitamento está relacionado diretamente com a representação gerada pelo SOM.

No *Focus+Context*, a aplicação parece funcionar muito bem em estruturas de mapas, porém, se compromete ao ser usado no MDS.

O efeito olho de peixe aplicado a interfaces multidimensionais pode acabar “enganando” o usuário ao distorcer uma interface que prioriza a distancia entre cada um de seus elementos. Acontece o mesmo com as PFNETs, porém, como a prioridade no *Pathfinder* são as relações locais entre os elementos, o problema é minimizado. É possível verificar algumas aplicações de Fisheye em mapas SOM no trabalho de Yang (*et. al.* 2003).

5. Considerações Finais

O desenvolvimento desta dissertação fundamentou-se em um referencial teórico no campo da Ciência da Informação e seu caráter interdisciplinar ligado diretamente à tecnologia da informação através dos aspectos da Recuperação e Visualização da Informação, mais especificamente, o uso de interfaces visuais no processo de recuperação da Informação.

Neste sentido, buscamos auxiliar outros investigadores através de princípios e juízo para eleição de uma ou outra combinação de técnicas no desenvolvimento de propostas específicas de interfaces visuais para recuperação da informação. Destacando, ainda, o processo de transformação visual, onde encontramos menos estudos e aplicações em ferramentas visuais para a RI.

É claro que devido a enorme quantidade de possíveis combinações de técnicas, algoritmos e análises na geração de interfaces visuais, nem todas combinações possíveis foram exploradas e investigadas.

A linha de estudo partiu dos dois paradigmas da recuperação da informação apresentados em nossa introdução: o *query searching*, ou simplesmente *querying*, e o *browsing*. É amplamente aceito que ambos modelos tem seus pontos fortes e fraquezas, e não são excludentes um do outro, mas sim complementares.

A visualização da informação é um ascendente campo de pesquisa cujo objetivo é capitalizar a habilidade de percepção humana para compreensão de informações abstratas. A visualização da informação para recuperação da informação deve determinar e distribuir os elementos no espaço visual.

Entendendo que a maior contribuição deste texto seja dada nas apresentações e análises das três etapas de desenvolvimento e produção de interfaces visuais para recuperação da informação, o autor acredita que futuras pesquisas na área devem focar de forma especial a terceira etapa: técnicas de transformação visual. Este processo parece ser até o momento, o menos estudado e aplicado na produção de interfaces.

Da mesma forma que as técnicas de classificação e distribuição, as de transformação visual têm por objetivo simplificar espaços complexos de informação, gerando a interface final. Esta simplificação não somente possibilita a compreensão de grandes volumes de informação que não seriam compreendidos por usuários de outra forma, mas também facilita a visualização e interação entre usuário e interface.

Neste sentido, colocamos como necessário estudos sobre novas técnicas de transformação visual interativas, ou ainda, maior investigação sobre os já existentes.

Devemos ainda ressaltar que a pouca divulgação e aceitação destes tipos de interfaces para usuários não se encontram desta forma por sua capacidade como ferramenta de recuperação da informação. Sua aplicação sofre dificuldades ainda pela dificuldade de utilização e aprendizagem de seu usuário final.

A redução de carga visual da interface, e conseqüentemente da carga cognitiva do usuário em sua interação com a ferramenta deve ser tratado como o maior desafio a se superar na pesquisa científica sobre interfaces visuais para recuperação da informação.

A disponibilidade de “capacidade computacional” e o amadurecimento de técnicas combinados à crescente demanda de usuários fazem com que o papel das interfaces visuais tenha grandes chances de desenvolvimentos na área de recuperação da informação. Interfaces visuais para recuperação da informação detêm diversas promessas e desafios representando, talvez o futuro da recuperação da informação.

Referências

BAEZA-YATES, R. Excavando la web. **El profesional de la información**, v.13, n.1, p 4-10, 2004.

BATES, M. J. The design of browsing and berrypicking techniques for the on-line Search interface. **On-line Review**, v. 13, n.5, p.407-424, Oct. 1989.

BEDERSON, B.B. Fisheye menus. In Mark Ackerman & K. Edwards, (Eds.), **Proceedings of ACM Conference on User Interface Software and Technology (UIST 2000)**. p. 217-226. New York, NY: ACM Press, 2000

BELKIN, N. J.; CROFT, W. B. Retrieval techniques. **Annual Review of Information Science and Technology**, v.22, p.109-189, 1987.

BÖRNER, K.; CHEN, C.; BOYACK, K. Dynamics of Scholarly Communication: Visualizing knowledge domains. **Annual Review Of Information Science And Technology**, v. 37, n. 1, p.179-255, 2003.

BÖRNER, Katy; CHEN, Chaomei; BOYACK, Kevin W.. Dynamics of Scholarly Communication: Visualizing knowledge domains. **Annual Review Of Information Science And Technology**, v. 37, n. 1, p.179-255, 2003.

BRIN, S., PAGE, L. The anatomy of a large-scale hypertextual Web search engine. **Computer Networks and ISDN Systems**, v.30 n.7, p.107-117, 1998.

BUZYDLOWSKI, J.W. **A comparison of self-organizing maps and pathfinder networks for the mapping of co-cited authors**. Tese de Doutorado, Universidade de Drexel, Philadelphia, USA. 2003.

CARD, S.K.; MACKINLAY, J.D. E SHNEIDERMAN, B. Information Visualization. In: Card, S.K.; Mackinlay, J.D. e Shneiderman, B. (eds.) **Readings in Information Visualization - Using Visualization to Think**. San Francisco, Morgan Kaufmann Publ., p. 1-34, 1999

CHANG, S., RICE, R. E. Browsing: A Multidimensional Framework. **Annual Review of Information Science and Technology (ARIST)**. v. 28 p. 231-76, 1993.

CHEN, C. Generalised similarity analysis and pathfinder network scaling. **Interacting with Computers**. v.10, n.2, p.107-128, 1998

CHUNG, W., CHEN, H., & NUNAMAKER, J.F. Business intelligence explorer: a knowledge map framework for discovering business intelligence on the Web. In: **Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)** - Track1 - Volume 1, p 10 - 12. Washinton, DC: IEEE. 2003.

COOPER, W. S. Getting beyond Boole. **Information Processing and Managemen.** v.24, n.3, p.243-248, 1988.

COVE, J. F., WALSH, B. C. Online Text Retrieval via Browsing. **Information Processing & Management.** V. 24. N. 1. p. 31-37, 1988.

COX, K. **INFORMATION RETRIEVAL BY BROWSING.** Disponível em: <<http://web.simmons.edu/~chen/nit/NIT%2792/069-cox.htm>>. Acesso em: 17 maio 2009.

EICK, S.G. Visualizing on-line activity. **Communication of the ACM**, v.44, n.8, p 45-52. 2001.

FERNEDA, E. **Recuperação de Informação: estudo sobre a contribuição da Ciência da Computação para a Ciência da Informação.** São Paulo, 2003. 147p. Tese (doutorado em Ciência da Informação). Escola de Comunicação e Artes, Universidade de São Paulo.

FOWLER, R.H., WILSON, B.A., & FOWLER, W.A.L. **Information Navigator: an information system using associative networks for display and retrieval.** Edinburg, TX: University of Texas-Pan American, Department of Computer Science. 1992. Disponível em: http://www.cs.panam.edu/research/information_navigator.pdf Acesso em: 14 abr 2008.

FREITAS, C. M. D.; CHUBACHI, O. M.; LUZZARDI, P. R. G.; CAVA, R. A. Introdução à Visualização de Informações. **Revista Rita**, v. 8, n. 2, p. 143-158. 2001. Disponível em <<http://infovis.ucpel.tche.br/luzzardi/Rita.pdf>>. Acesso em março de 2009.

FURNAS, G.W. Generalized fisheye views. In Marilyn Mantei & Peter Orbeton (Eds.). **Proceedings of the SIGCHI conference on human factors in computing systems**, p 16-23. New York, NY: ACM Press. 1986.

GERSHON, N.; EICK, S. Visualization's New Tack. **IEEE Spectrum.** V. 32 N. 11. p. 38-56, 1995.

GIL LEIVA, I., RODRÍGUEZ MUÑOZ, J.V. Tendencias en los sistemas de indización automática: estudio evolutivo. **Revista Española de Documentación Científica**, v. 19, p. 31-51, 1996.

GUTWIN, Carl; FEDAK, Chris. Interacting with Big Interfaces on Small Screens: a Comparison of Fisheye, Zoom, and Panning Techniques. **Acm International Conference Proceeding Series**: Proceedings of the 2004 conference on Graphics interface, v. 62, p.145-152, 2004.

HASSAN-MONTERO, Y. **Visualización y Recuperación de Información**. In: Actas II Encontro de Ciências e Tecnologias da Documentação e Informação. Portugal. Acesso em: 27 de abr. 2006.

HAYAMA, T., KANAI, T., & KUNIFUJI, S. Personalized environment for skimming documents. **Lecture Notes in Computer Science**, v.2, p 771-778. 2003.

HERRERO-SOLANA, V. **Modelos de representación visual de la información bibliográfica: aproximaciones multivariantes y conexionistas**. Tese de Doutorado, Universidade de Granada, Granada, Spain. 2000.

HERRERO-SOLANA, V.; HASSAN MONTERO, Y. Metodologías para el desarrollo de Interfaces Visuales de Recuperación de Información: análisis y comparación. In: **Information Research**, 11(3), Abril 2006. Disponível em: <<http://informationr.net/ir/11-3/paper258.html>>. Acesso em: 07 set. 2008.

HILDRETH, C. R. Online Browsing Support Capabilities in Information Interaction. **Proceedings of the 45th ASIS Annual Meeting**. V. 19, 1982.

HORNBAEK, K.; FRØKJÆR, E. Reading patterns and usability in visualizations of electronic documents. **ACM Transactions on Computer-Human Interaction**. v.10, n.2, p.119-149, 2003

JOHNSON, Brian; SHNEIDERMAN, Ben. Tree-maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In: **IEEE VISUALIZATION**, 1991, San Diego. Proc. p. 284 - 291. 1991.

KASKI, S. *Data exploration using self-organizing maps*. **Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series**. n.82, 1997.

KOIKE, H. The Role of Another Spatial Dimension in Software Visualization. **ACM Transactions on Information Systems**. v.11, n.3, p.266-286, 1993.

KOHONEN, T.. **Self-organization and associative memory**. 3 ed. Berlim: Springer-verlag, 1989.

KOPANAKIS, I., THEODOULIDIS, B. Visual data mining modeling techniques for the visualization of mining outcomes. **Journal of Visual Languages and Computing**, v.14, n.6, p 543-589, 2003.

LAMPING, J., RAO, R., PIROLI, P. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In Irvin R. Katz, Robert Mack, Linn Marks, Mary Beth Rosson & Jakob Nielsen (Eds.). **Proceedings of the ACM SIGCHI conference on human factors in computing systems**. p 401-408. New York, NY: ACM Press. 1995

LEUNG, Y.K., APPERLEY, M.D. Review and taxonomy of distortion-oriented presentation techniques. **ACM Transactions on Computer-Human Interaction**, v.1 n.2, p 126-160. 1994.

LIN, X. Map displays for information retrieval. **Journal of the American Society for Information Science**, v.48, n.1, p 40-54. 1997.

LUHN, H. P. The automatic creation of literature abstracts. **IBM Journal of Research and Development**. v.2, n.2, p.159-165, 1958.

MARCHIONINI, G. **Information seeking in electronic environments**. New York, NY: Cambridge University Press. 1995.

MARCOS-MORA, M.C. **Interaction in information retrieval interfaces: concepts, metaphors and visualization**. Gijón: Ediciones TREA. 2004

MORSE, E. L.; LEWIS, M. Testing Visual Information Retrieval Methodologies case study: comparative analysis of textual, icon, graphical, and "spring" displays. **Journal of the American Society for Information Science and Technology**, v.53, n.1, p.28-40, 1997

MOYA-ANEGÓN, F. **Integrated library management systems: data structures and information retrieval**. Madrid: ANABAD. 1994.

MOYA-ANEGÓN, F.; HERRERO-SOLANA, V. Investigaciones en curso sobre interfaces gráficos en dos y tres dimensiones para el acceso a la información electrónica. **Cuadernos de Documentación Multimedia**. v.8, 1999.

NASUKAWA, T. NAGANO, T. Text analysis and knowledge mining system. **IBM Systems Journal**. V. 40, N. 4. p. 967-984, 2001.

ODDY, R. N. Information retrieval through man-machine dialogue. **Journal of Documentation**. v.33, n.1, p.1-14, 1977.

POLANCO, X., ZARTL, A. (2002). **Information visualization**. EICSTES Project. Deliverable 1.4. State of the art part c: WP9. acessado em 12 Abril, 2005 Disponível em: <http://eicstes.inist.fr/public/D1.4_Visualization_WP9.pdf>. Acesso em 07 set. 2008.

RENNISON, E. Galaxy of news: an approach to visualizing and understanding expansive news landscapes. In **UIST 94, ACM Symposium on User Interface Software and Technology**. New York, NY: ACM Press, p.3-12, 1994.

RIJSBERGEN, C.J. **Information retrieval**. London: Butterworths, 1975. Acessado em 21 jan 2009. Disponível em: <<http://www.dcs.gla.ac.uk/Keith/Preface.html>>

ROBERTSON, P. K. A Methodology for Choosing Data Representations. **IEEE Computer Graphics and Applications**. v.11, n.3, p.56-68, 1991.

SCHVANEVELDT, R. W. **Pathfinder associative networks: studies in knowledge organization**. Norwood, USA: Ablex Publishing Corp, 1990.

STASKO, J. DOMINGUE, J., BROWN, M.H., PRICE, B. A. **Software visualization: programming as a multimedia experience**. Massachusetts: MIT Press. 562 p. 1997.

STOREY, M. A.D. WONG, K., FRACCHIA, F.D. & MÜLLER, H.A. On integrating visualization techniques for effective software exploration. In **Proceedings of the 1997 IEEE Symposium on Information Visualization**. Washington: IEEE Computer Society. 1997.

TURETKEN, O. SHARDA, R. Development of a fisheye-based information search processing aid (FISPA) for managing information overload in the web environment. **Decision Support Systems**. v.37, n.3, p.1-20, 2004.

WILLETT P.; WOOD, F. E. The use of the instruct text retrieval program at the Department of Information Studies University of Sheffield. **Education for Information**. v.7, p.133-141, 1989.

XAVIER, R. F. **Interfaces Gráficas para Apresentação de Indicadores**. 2006. 63 f. Trabalho de Conclusão de Curso (Graduação) - Curso de Biblioteconomia, Unesp, Marília, 2006.

XAVIER, R. F.; CERVANTES, B. M. N.; EUCLIDES, M. L.; Herrero-Solana, V. Scientific Collaboration Networks in Information Science: aids for its construction based on the Brazilian experience with the Lattes Curriculum/National Council of Scientific and Technological Development (CNPq). In: Vicente P. Guerrero-Bote (Editor). (Org.). **Current Research in Information Sciences and Technologies: Multidisciplinary Approaches to Global Information Systems**. Badajoz, SPAIN: Open Institute of Knowledge (Instituto Abierto del Conocimiento), 2006, v. 2, p. 447-451

YANG, C.; CHEN, H.; HONG, K. Visualization of large category map for Internet browsing. **Decision Support System**, v. 35, n.1 89-102. 2003

ZHANG, J. **Visualization for Information Retrieval**. New York, NY: Springer. 2008.
