

Considerações sobre o uso de Visualização de Informação no auxílio à gestão de informação

Celmar Guimarães da Silva

Instituto de Computação – Universidade Estadual de Campinas (Unicamp)
Caixa Postal 6176 – 13084-971 – Campinas – SP – Brasil

celmar@ic.unicamp.br

Abstract. *One of the grand challenges in Computer Science research in Brazil for the years 2006 to 2016 is the challenge of management of information over massive volumes of data. This paper aims to present how the objectives of Information Visualization field are close to the objectives of this challenge, detailing the data visualization process and exemplifying analysis situations that gain with the use of Information Visualization techniques and concepts. This paper also presents some Information Visualization challenges as a way of pointing steps to promote the field development and the interdisciplinary use of its results, by which its advantages can be disseminated through distinct social, economical and cultural contexts.*

Resumo. *Um dos grandes desafios de pesquisa de Ciência da Computação brasileira para os anos 2006 a 2016 é o desafio da gestão de informação relativa a grandes volumes de dados. Este artigo visa apresentar como os objetivos da área de Visualização de Informação vão ao encontro dos objetivos desse desafio. Para tanto, o artigo detalha o processo de visualização de dados e exemplifica situações de análise que se beneficiam do uso de técnicas e conceitos de Visualização de Informação. O artigo também apresenta alguns desafios na área de Visualização de Informação, como forma de indicar passos a serem tomados para promover o desenvolvimento da área e a utilização interdisciplinar de seus resultados, pela qual seus benefícios podem ser disseminados em diferentes contextos sócio-econômico-culturais.*

1. Introdução

Sistemas computacionais de informação estão cada vez mais presentes em diversos aspectos da vida humana, coletando e armazenando volumes de dados heterogêneos que crescem exponencialmente e de maneira distribuída. Essas características em conjunto dificultam consideravelmente a obtenção de informação relevante nesses dados. Nesse sentido, pesquisadores brasileiros apontam como um grande desafio de pesquisas em Ciência da Computação no Brasil até 2016 a integração de diferentes áreas da Computação para desenvolver aplicações que beneficiem o contexto sócio-econômico-cultural do país, provendo soluções para tratar, recuperar e disseminar informação relevante a partir desses volumes de dados [Carvalho et al. 2006].

Esse objetivo é consoante com os objetivos da área de Visualização de Informação, que visa facilitar o processo de derivação e entendimento de informação a

partir da análise visual de conjuntos de dados. Diferentes técnicas de Visualização de Informação utilizam recursos computacionais para representar esses dados de maneira gráfica e interativa, procurando otimizar o uso das capacidades visuais humanas para compreender fenômenos que não possuem em si mesmos uma representação espacial própria [Card et al. 1999; Chen 2002].

Detalhando essa consonância de objetivos (Seção 2), este artigo procura exemplificar áreas que fazem uso de Visualização de Informação (ainda que de forma restrita), em diferentes cenários nos quais é necessário obter informações relevantes em grandes volumes de dados (Seção 3). Este artigo também apresenta alguns desafios atuais de Visualização de Informação, cuja solução terá impacto na forma como a área pode contribuir com atividades de gestão de informação (Seção 4).

2. Consonância de objetivos

Para apresentar como os objetivos de Visualização de Informação estão em harmonia com os objetivos do desafio de gestão de informação, é importante primeiramente detalhar o que se compreende por tratar, recuperar e disseminar informação relevante a partir de volumes de dados. A relevância de uma informação contida em um conjunto de dados está relacionada à tarefa de análise sendo executada, a qual normalmente envolve examinar características dos dados em questão. Recuperar informação relevante exige, assim, tratar esse conjunto de dados sob análise, visando sua organização em entidades, atributos e relacionamentos que sejam relevantes à tarefa de análise, e que possam ser consultados e comparados com relação às suas características. Recuperar informação relevante exige ainda que esses dados sejam apresentados de maneira clara e não ambígua às pessoas envolvidas na análise, de modo que elas possam interpretá-los corretamente e abstrair informação relevante a ser posteriormente compartilhada e disseminada.

Esse processo de tratamento, recuperação e disseminação de informação pode ser comparado com o processo pelo qual Visualização de Informação procura transformar em uma representação gráfica e interativa um conjunto de dados brutos. Este processo está representado no diagrama da Figura 1, obtido pela sumarização de alguns diagramas representativos do processo de Visualização de Informação [Card et al. 1999; Ware 2004; Spence 2001].

Nessa figura, um indivíduo possui em mente a tarefa de estudar uma situação particular e obter informações relevantes sobre ela. Para tanto, é necessário inicialmente coletar dados sobre essa situação, a partir dos quais se deseja obter as informações necessárias. A coleta desses dados pode ser feita por sistemas computacionais, por diferentes dispositivos como sensores e câmeras, ou manualmente pelo próprio indivíduo. O conjunto de dados brutos coletados pode ser heterogêneo, contendo datas, medidas, quantidades, códigos, descrições, coordenadas espaciais, imagens, vídeos, arquivos diversos, endereços de sites da Internet, entre outros. Diferentes entidades, atributos e relacionamentos podem estar representados nesse conjunto de dados.

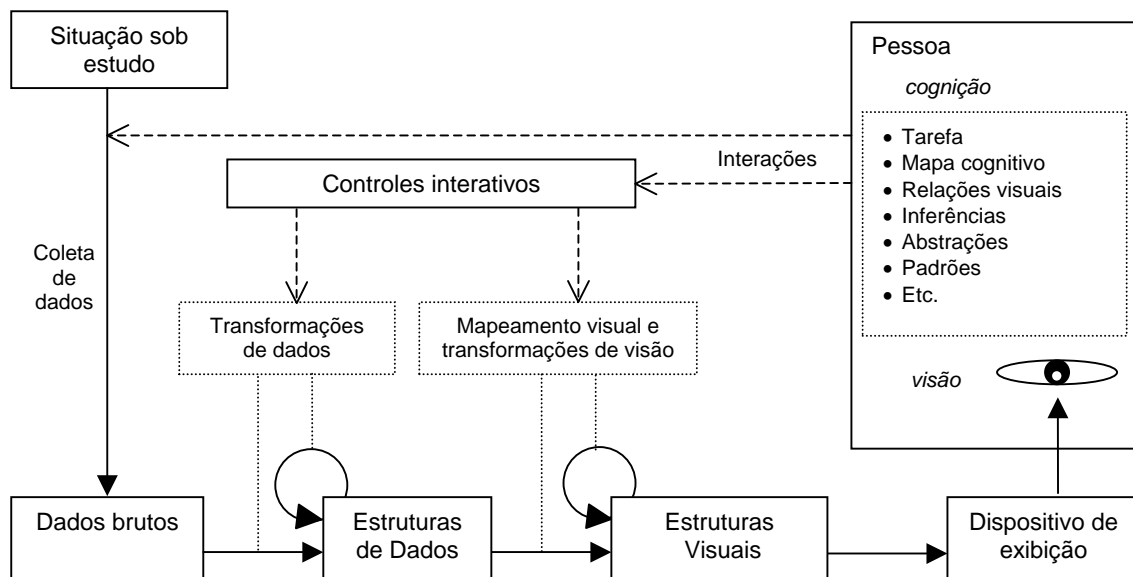


Figura 1. Diagrama representando o processo de Visualização de Informação.

Dada essa multiplicidade e heterogeneidade de dados, recuperar informação exige primeiramente a tarefa de organizar esses dados, de forma que alguma informação possa ser obtida pela consulta e comparação de entidades similares, de suas características e dos relacionamentos entre diferentes entidades. Esta tarefa de tratamento de dados envolve organizar esses dados em estruturas (ou tabelas), e é denominada *transformação de dados* [Card et al. 1999]. Quando os dados a serem tratados estão armazenados em um SGBD, essa transformação pode ser desnecessária, pois a organização do próprio banco de dados pode ser utilizada. Por outro lado, há casos em que essas transformações são necessárias, pois a situação estudada apresenta muitos dados dispersos e sem relacionamento aparente (como pode ocorrer na coleta manual de dados), exigindo organização para ser mais bem compreendido.

Esse primeiro nível de organização de dados pode não ser suficiente para que o indivíduo obtenha informações relevantes sobre a situação estudada. Nesses casos, ordenação, cálculo de médias e classificação de características são algumas operações que podem ser aplicadas sobre as estruturas de dados obtidas. Essas novas transformações de dados permitem classificar e agregar entidades que compartilham características semelhantes, possibilitando a sumarização e abstração do conjunto de dados coletado [Card et al. 1999]. Algoritmos de recuperação de informação, como os utilizados para cálculo de similaridades entre imagens [Torres et al. 2003], também podem ser usados para, a partir dos dados brutos, extrair dados que sejam mais relevantes com relação à tarefa de análise.

Mesmo após diferentes operações de classificação e de agregação de dados, o conjunto de dados disponível pode ser ainda suficientemente grande (em quantidade de dados e em número de dimensões) para se conseguir obter informação relevante. Isso acontece porque a representação de dados em formato textual exige que se faça um processamento visual controlado [Card et al. 1999] desses dados, um processo que é detalhado, serial, lento e de baixa capacidade de dados. Informações representadas

graficamente, por outro lado, tendem a ser processadas de maneira mais automática pela visão, em um processo mais superficial, paralelo (no sentido de obter várias informações simultaneamente), rápido e de capacidade elevada. Dessa forma, representar graficamente os dados a serem analisados é interessante do ponto de vista da obtenção de informação, pois faz com que não apenas mecanismos computacionais sejam usados para a análise de dados, mas também recursos da visão e da cognição humana.

Nesse sentido, um estudo de Larkin e Simon (1987) estendido por Card et al. (1999) aponta diferentes aspectos pelos quais visualizações podem possibilitar uma ampliação cognitiva (ou seja, uma maior facilidade no uso e aquisição de conhecimento): (1) aumentando os recursos de memória e de processamento disponíveis para usuários, através do uso direto dos recursos do sistema visual e da memória de trabalho externa e visual; (2) reduzindo a busca por informação, de diversas formas: agrupando ou relacionando visualmente informações, compactando-as, exibindo uma visão geral ou mesmo mostrando detalhes sob demanda; (3) usando representações visuais para melhorar a detecção de padrões; (4) habilitando operações de inferência perceptiva, como tornar óbvia a resposta de um problema através de uma representação visual; (5) usando mecanismos de atenção perceptiva para efetuar monitoramento de uma grande quantidade de eventos potenciais; e (6) codificando informação numa mídia manipulável.

Para que essa ampliação cognitiva possa ocorrer, é importante definir de maneira apropriada o modo como as estruturas de dados são transformadas em estruturas visuais, processo este chamado de *mapeamento visual*. Esse processo deve levar em conta quais as características dos dados a serem representados, das propriedades visuais que podem ser usadas para representar esses dados, da interface humano-computador envolvida no processo (no caso, a tela de exibição dos dados e os dispositivos de recebimento de interações do usuário), e da visão e cognição humana. Essas características são detalhadas a seguir.

Cada atributo armazenado nas estruturas de dados pode ser classificado de acordo com sua semântica e com suas dependências funcionais em relação a outros atributos. Diferentes autores classificam de forma semelhante a semântica de atributos (ou variáveis) e de seus dados [Spence 2001; Card et al. 1999; Ware 2004], embora com nomenclaturas ligeiramente diferentes. Generalizando as diferenças, eles consideram a existência de três categorias básicas de atributos:

- *Nominais* (ou categóricos): conjunto de elementos sem ordem específica, como uma lista de títulos de filmes ou de nomes de pessoas;
- *Ordinais*: conjunto de elementos que apresentam necessariamente uma relação de ordem entre si, como classificações em faixas etárias, ou uma lista de datas históricas (neste caso, um atributo ordinal temporal);
- *Quantitativos* (ou numéricos): conjunto de elementos que apresentam um escopo numérico, sendo possível aplicar operações aritméticas sobre eles. Por exemplo, distância entre cidades (atributo quantitativo espacial) ou quantidade de tempo que um usuário leva para ir de uma página Web para outra em um site (atributo quantitativo temporal).

Cada atributo também pode ser classificado de acordo com sua dependência funcional relativa aos demais atributos. De acordo com essa característica, os atributos de uma estrutura de dados podem ser divididos em atributos de entrada e atributos de saída, de tal forma que os primeiros determinam funcionalmente os últimos (de modo análogo a uma função matemática).

Ao definir o mapeamento visual dos dados, o que se procura fazer é estabelecer como cada atributo será representado, ou seja, quais propriedades gráficas e espaciais podem ser utilizadas para sua representação. Por exemplo, um gráfico bidimensional contendo círculos e triângulos de diferentes cores pode ser usado para representar ao menos quatro atributos: dois deles mapeados para os eixos X e Y do gráfico, outro para as formas das marcas usadas (triângulo ou retângulo), e outro para as cores dessas marcas, como exemplificado na Figura 2. Neste exemplo, ficam evidentes alguns elementos com os quais se pode trabalhar em um mapeamento visual: espaço, marcas e propriedades gráficas dessas marcas [Card et al. 1999]. Além destes, uma representação visual pode utilizar ainda o tempo como uma forma de mapear outros atributos dos dados analisados. Para tanto, o indivíduo que efetua a análise pode utilizar controles interativos que filtram quais dados desses atributos devem ser considerados. Ao modificar os valores considerados e desconsiderados por esses filtros, esse indivíduo pode observar como são afetados os demais atributos já representados.

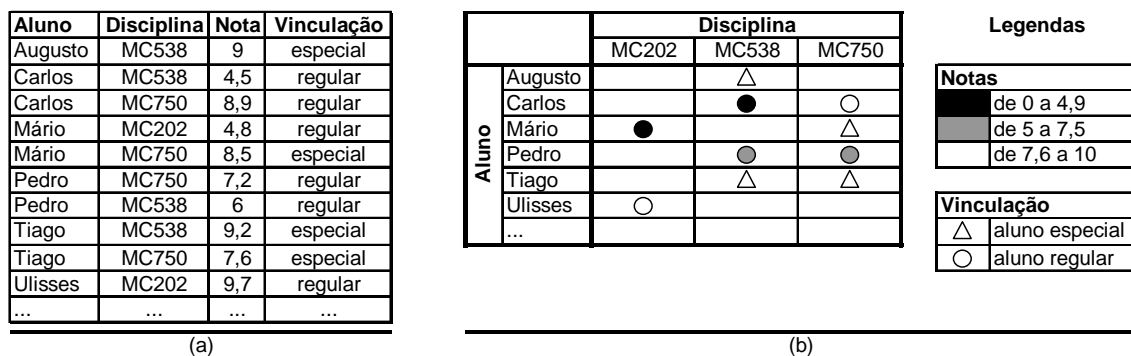
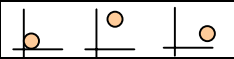





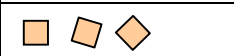


Figura 2. (a) Tabela representando notas de alunos regulares e de alunos especiais cursando disciplinas. (b) Representação visual dos atributos e dos dados de (a), evidenciando o uso de espaço, marcas e propriedades gráficas para representar dados.

Um dos fatores que dificulta a definição de um mapeamento visual é a quantidade de propriedades gráficas disponíveis, como cores (e suas diferentes componentes, como tonalidade, saturação e brilho), formas (como diferentes polígonos ou ícones), tamanho, textura e orientação, dentre outros (Tabela 1). Com diversas possibilidades de combinar essas e outras propriedades gráficas para definir o mapeamento visual, torna-se necessário escolher um mapeamento que expresse de maneira efetiva os dados. Os conceitos de expressividade e de efetividade de Mackinlay (1986) auxiliam neste sentido, indicando que: (1) estruturas visuais devem expressar todos os dados presentes na estrutura de dados, e somente eles; e (2) estruturas visuais devem ser efetivas, no sentido de permitirem rápida interpretação dos dados representados, e fácil distinção entre eles, levando à menor quantidade possível de erros de interpretação. A tabela de efetividade de propriedades retiniais de Card et al. (1999) permite comparar a efetividade de propriedades gráficas com relação a cada categoria semântica de atributo (nominal, ordinal ou quantitativa), auxiliando, assim, a escolher

quais categorias de atributos se deve representar por quais propriedades gráficas. Por outro lado, análises dos relacionamentos existentes entre eixos espaciais, marcas e suas propriedades gráficas também podem auxiliar na definição do mapeamento visual com relação à sua expressividade. Dada uma estrutura visual, a observação desses relacionamentos (que podem ser compreendidos como dependências funcionais) permite verificar se uma estrutura visual consegue expressar exatamente as mesmas dependências funcionais dos dados a serem representados [Silva 2006].

Tabela 1. Exemplos de propriedades espaciais e gráficas.
Adaptado de Silva (2006).

Propriedades	Exemplo	Propriedades	Exemplo
Posição espacial		Cor	
Tamanho		Textura	
Brilho		Forma	
Orientação			

Entre a representação visual dos dados e a visão humana encontra-se uma das partes que possibilitam a interação humano-computador desse processo de visualizar informação: os dispositivos de exibição de dados. Estes podem ser os mais diversos, variando em tamanho físico (por exemplo, telas de projeção e monitores tradicionais), em resolução (como monitores, PDAs e celulares), em dimensionalidade (bidimensional na maioria dos casos, mas pode ser tridimensional com o uso de capacetes de visão tridimensional ou de mecanismos de projeções tridimensionais), em quantidade de cores disponíveis, e mesmo em presença ou ausência de interatividade (por exemplo, representação em papel ou em monitor). Essa variedade de características dificulta a tarefa de definir um mapeamento visual, pois características distintas de dois dispositivos podem exigir mapeamentos visuais distintos, os quais explorem cada um a efetividade das propriedades gráficas, espaciais e temporais disponíveis em seu respectivo dispositivo de exibição.

Não menos importante, outra parte da interação humano-computador envolvida nesse processo é a atuação humana na modificação das estruturas de dados e das estruturas visuais do processo. Controles dispostos na tela (como listas de seleção e *range-sliders*) podem permitir que o indivíduo que efetua a análise modifique o conjunto de atributos e de dados a serem considerados na representação visual, ou mesmo a forma como o mapeamento visual representa esses dados. Por exemplo, o usuário pode escolher elementos em uma lista, definir intervalos de tempo, selecionar dados, etc., e essas ações podem afetar quais dados serão exibidos na estrutura visual, funcionando dessa forma como *transformações de visão*. Com isso, o usuário pode modificar dinamicamente qual a fatia do conjunto de dados a ser exibida, reduzindo o excesso de dados e auxiliando seu processo de obtenção de informação. Diferentes técnicas como consultas dinâmicas [Shneiderman 1994], detalhes-sob-demanda, manipulação direta [Ahlberg e Shneiderman 1994], seleção direta, visão geral + detalhes, movimento de câmera e aproximação (*zoom*) podem ser utilizadas para possibilitar a interação do usuário com as estruturas do processo de visualização [Card et al. 1999]. Toda essa interação depende dos dispositivos pelos quais as ações do usuário são recebidas pelo computador; apesar da predominância do *mouse* e do

teclado, também se deve estudar a interação por meio de outros dispositivos, como telas sensíveis a toque, canetas (como as utilizadas em PDAs) e mesmo por comandos de voz.

Com relação a características da visão e cognição humana nesse processo, alguns aspectos já foram cobertos nesta seção, como processamento automático de determinadas propriedades gráficas e efetividade de propriedades gráficas para representar categorias semânticas de dados. No entanto, é importante ressaltar que aspectos de acessibilidade também devem ser estudados em Visualização de Informação, levando em conta deficiências de visão como miopia, hipermetropia, presbiopia, dificuldade para distinguir cores (daltonismo), ou mesmo ausência parcial ou total de visão. No sentido de superar essas dificuldades, pode-se procurar prover representações sonoras dos dados, as quais procurem compensar de alguma forma as deficiências para visualizar informação. Essas representações sonoras são providas por uma área de estudo denominada Sonificação de Informação [Hermann e Ritter 1999], que por algumas vezes complementa representações visuais, e por outras provê representações completamente sonoras para dados.

Pela noção geral do processo apresentado nesta seção, considera-se que tratar e recuperar informação são atividades que podem ser fortemente auxiliadas por conceitos e técnicas da área de Visualização de Informação. Integrando os recursos visuais e cognitivos humanos aos recursos gráficos e interativos computacionais, Visualização de Informação pode ser usada no processo de gestão de informação como uma forma de pessoas interagirem diretamente com um conjunto de dados representativo de uma situação, com o objetivo de compreender melhor esses dados e de obter visões esclarecedoras sobre a situação estudada.

3. Alguns usos de Visualização de Informação

Considerando, pelo exposto na Seção 2, que os objetivos de Visualização de Informação convergem para os objetivos mais gerais do desafio de gestão de informação apresentado por Carvalho et al. (2006), é conveniente apresentar exemplos de como Visualização de Informação é utilizada atualmente na tentativa de facilitar essa gestão em diferentes atividades e áreas de conhecimento.

Em Educação a Distância (EaD), Visualização de Informação é utilizada para auxiliar a compreensão de diferentes aspectos sociais, cognitivos e comportamentais de estudantes de cursos a distância. Nesse sentido, diferentes trabalhos utilizam estruturas visuais como gráficos de dispersão, gráficos de barras, grafos, espirais e diagramas de Venn, dentre outros, para representar dados armazenados em ambientes de EaD [Silva 2006]. A título de exemplificação, os sistemas GISMO e CourseVis [Mazza e Milani 2005; Mazza e Dimitrova 2005] usam histogramas e gráficos de dispersão para representar dados de cursos a distância, como: acessos de estudantes ao curso e a conteúdos disponíveis, participação em discussões, resoluções de tarefas e respectivas notas. A Figura 3 exibe um exemplo de gráfico de dispersão do GISMO que representa a quantidade de mensagens de cada estudante em um fórum de discussão, distinguindo por cores as mensagens postadas, as mensagens lidas e os tópicos de discussão iniciados pelo estudante ao longo do tempo. A sumarização de dados feita por essa figura indica que a grande maioria dos estudantes acessava as mensagens postadas no fórum de discussão, e que 10 alunos não postaram nenhuma mensagem nesse fórum.

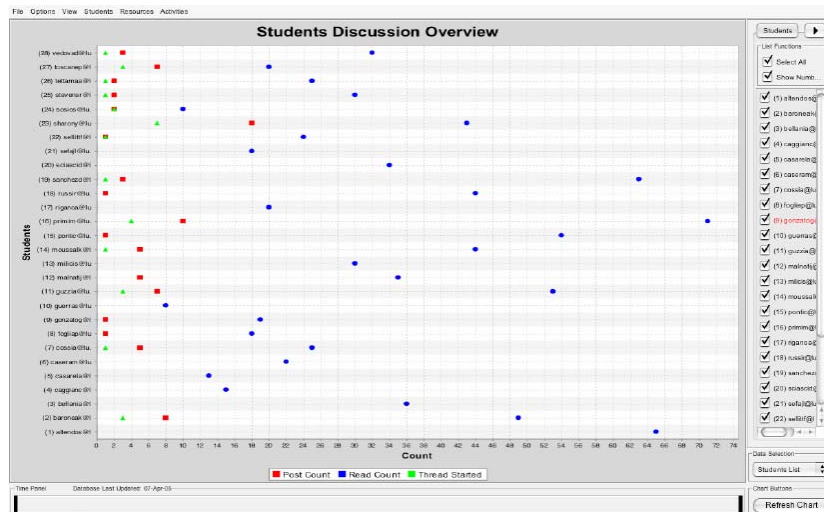


Figura 3. Visualização de participação de estudantes em um fórum de discussão ao longo do tempo na ferramenta GISMO [Mazza e Milani 2005].

Outro exemplo de uso de Visualização de Informação em EaD é a ferramenta InterMap [Romani 2000], que utiliza grafos, gráficos de dispersão e grafos orientados a força (como os propostos por Eades (1984)) para representar graficamente as interações ocorridas entre participantes de cursos em algumas ferramentas de comunicação do ambiente de EaD TelEduc [Rocha et al. 2002]. No exemplo da Figura 4, os vértices de um grafo representam participantes de um curso oferecido via TelEduc, e as arestas representam o envio de mensagens entre esses participantes na ferramenta de correio eletrônico desse ambiente em um período determinado. Essa figura traz à tona a informação de que quase todas as trocas de mensagens entre participantes do curso nesse período estão centradas no professor (vértice no centro do grafo), um indício de que colaborações entre estudantes podem não estar ocorrendo no curso.

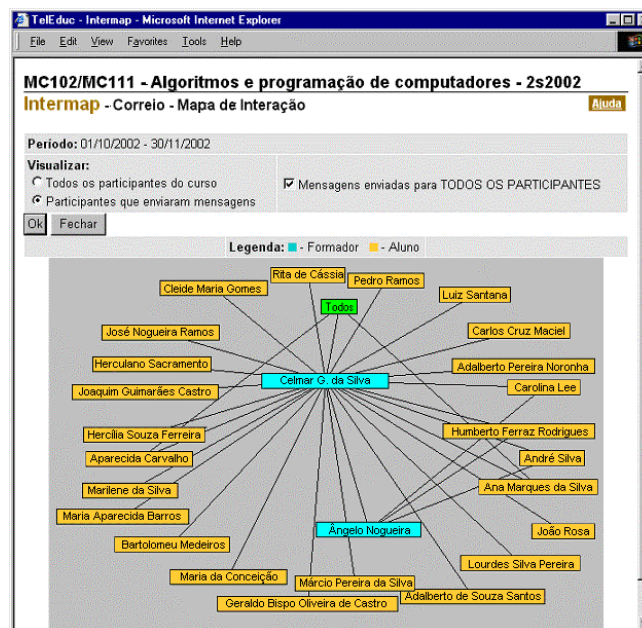


Figura 4. Visualização via ferramenta InterMap da interação entre participantes de um curso em um período determinado [Silva 2006].

Ainda em EaD, a pesquisa de Silva (2006) aponta uma solução mais genérica para que usuários de ambientes de EaD se beneficiem do uso de técnicas de Visualização de Informação para atender às suas demandas de informação sobre os dados armazenados por esses ambientes. Essa pesquisa procura possibilitar que usuários de ambientes de Educação a Distância, com diferentes demandas de análises de dados, definam diferentes combinações de dados a serem analisadas, em vez de utilizar apenas consultas predefinidas pelo sistema. Essa solução utiliza conceitos de modelo de relação universal de Bancos de Dados para possibilitar aos usuários escolherem atributos da base de dados a serem consultados, sem que esses usuários conheçam internamente a base de dados do ambiente consultado. Além disso, esses conceitos são usados para executar na base de dados do ambiente de EaD a consulta referente a essa escolha. Os resultados dessa consulta são convertidos automaticamente em representações gráficas interativas, escolhidas com o auxílio de conceitos de expressividade de estruturas visuais e efetividade de propriedades gráficas, e construídas com base em técnicas de consultas dinâmicas. Dessa forma, evita-se que o usuário precise se preocupar com a maneira de construir um gráfico adequado à análise desses dados.

Como exemplo dessa pesquisa, a Figura 5 mostra um usuário definindo uma consulta relativa a Fóruns de Discussão de um curso oferecido via TelEduc, selecionando os atributos “Fóruns de Discussão”, “Datas de mensagens dos fóruns de discussão” e “Quantidade de mensagens de fóruns de discussão”. Após o usuário encerrar essa seleção de atributos, o sistema efetua na base de dados do TelEduc uma consulta relacionada aos atributos escolhidos; em seguida, mostra o resultado dessa consulta como um gráfico de dispersão, tal como apresentado na Figura 6. Essa figura, um gráfico de dispersão, tem como eixos X e Y os dois primeiros atributos, e representa por cor o terceiro atributo. Resumindo visualmente os dados dos atributos escolhidos, a Figura 6 permite observar que a maior parte das mensagens de fóruns de discussão foram enviadas entre 30/03/2005 e 19/04/2005, e que os dois fóruns de discussão apresentados na parte inferior da figura tiveram uma participação pequena e concentrada em um único dia de curso. A Figura 7 apresenta o resultado de outra consulta aos dados do ambiente TelEduc, dessa vez relacionada à ferramenta Portfólio, que armazena itens (conteúdos) postados por participantes de um curso e comentários postados sobre esses itens. Essa figura, obtida por uma consulta definida pelos atributos “Usuários donos de itens de portfólio”, “Usuários que comentaram itens de portfólio” e “Quantidade de comentários de itens de portfólio”, permite observar que apenas três participantes (o professor, seu monitor e um aluno) estavam efetivamente postando comentários nos itens postados pelos demais colegas, como indicado pelas três linhas verticais observadas na figura. Uma tênue linha diagonal aponta ainda que alguns participantes comentavam seus próprios itens, o que provavelmente era feito para responder aos comentários que eles recebiam nesses itens.

Em Biologia Computacional, Visualização de Informação pode auxiliar o processo de montagem de genomas. Nesse sentido, a ferramenta Hawkeye [Schatz et al. 2007] procura auxiliar a identificar e corrigir erros de montagem de genomas, utilizando Visualização Científica e de Informação para facilitar a inspeção de dados de montagem de larga escala, enquanto minimiza o tempo necessário para detectar erros de montagem e para fazer julgamentos apurados sobre a qualidade da montagem. Utiliza para isso visões gerais sobre dados, filtragem dinâmica e clusterização automatizada, visando focar a atenção do usuário e destacar anomalias nos dados.

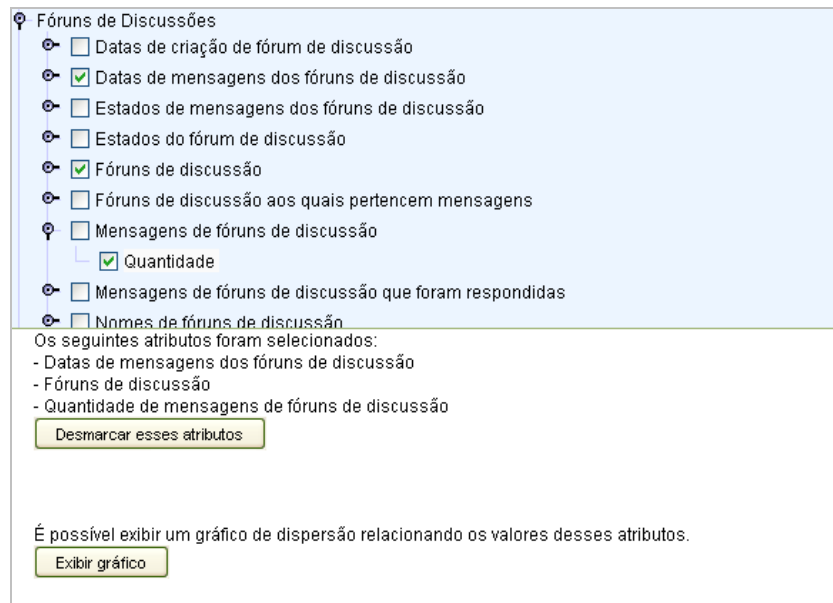


Figura 5. Exemplo de uso do protótipo de Silva (2006) para consultar dados de ambientes de EaD. Usuário selecionando atributos da base de dados do ambiente TelEduc para definir sua consulta.

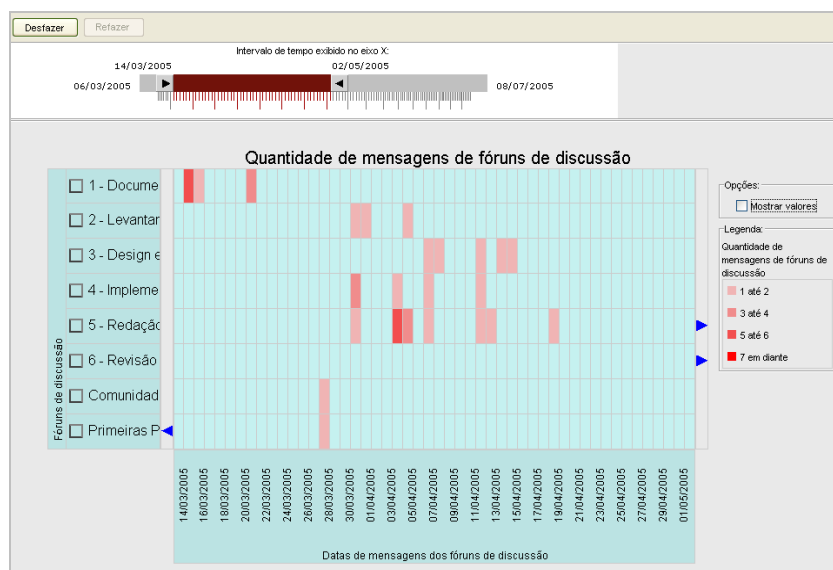


Figura 6. Representação visual e interativa feita pelo protótipo de Silva (2006) para o resultado da consulta definida na Figura 5.

Em Medicina, podem ser citados exemplos de uso de Visualização de Informação no estudo de dados neurais. Walter et al. (2004) utilizam um túnel tridimensional como estrutura visual para representar a atividade de neurônios ao longo do tempo, possibilitando que analistas detectem visualmente a sincronia de atividades entre neurônios e o acoplamento entre eles. Outro exemplo interessante é o uso da ferramenta LifeLines [Plaisant et al. 1996], que resume de maneira gráfica e interativa registros médicos sobre o histórico de um paciente, relatando a ocorrência de diferentes eventos ao longo do tempo, como: consultas, manifestações de doenças, tratamentos, hospitalizações, entre outros.

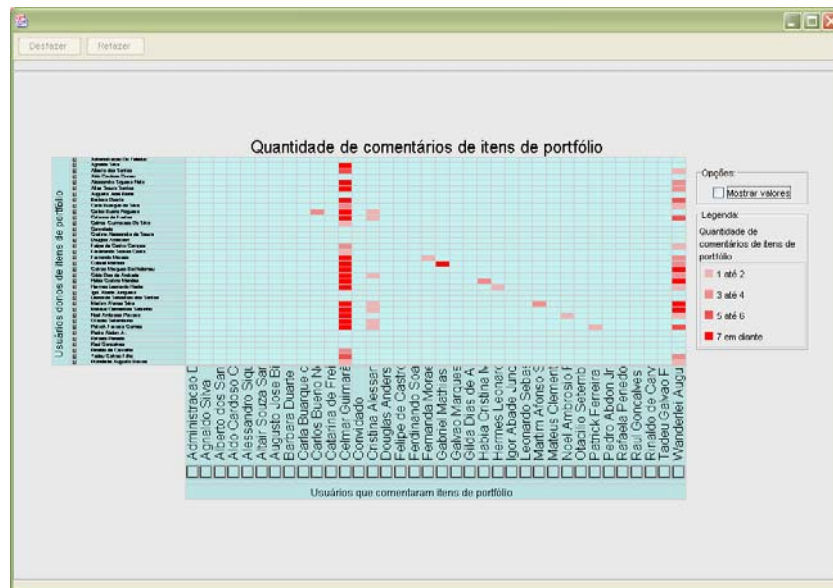


Figura 7. Representação visual feita pelo protótipo de Silva (2006) para resultado de consulta relativa à quantidade de comentários postados em itens de portfólios do ambiente TelEduc em um curso.

Visualização de Informação também pode ser de grande valor em atividades relacionadas à detecção de fraudes. Nesse sentido, Spence (2001) apresenta uma situação de uso da ferramenta Netmap [Westphal e Blaxton 1998 apud Spence 2001] para acelerar a detecção de fraudes pela identificação de padrões em relações entre entidades. A ferramenta utiliza grafos interativos em formato circular que representam a conectividade entre entidades; no caso estudado, essas entidades são pessoas que emprestam dinheiro, que compram ou vendem casas, e as casas em si. Spence reporta que, com o uso dessa ferramenta, um analista conseguiu em quatro semanas identificar uma fraude que, para ser identificada na vida real, demandou o trabalho de oito pessoas-ano analisando documentos.

Os exemplos apresentados tornam evidente que interdisciplinaridade é uma característica intrínseca da área de Visualização de Informação, uma vez que essa área pode auxiliar a atender a demandas por informação vindas de diferentes áreas de conhecimento. Por outro lado, esse atendimento a demandas por informação propicia que técnicas e conceitos de Visualização de Informação sejam desenvolvidos, avaliados e refinados. Como uma área jovem e em pleno desenvolvimento, muitos são ainda os desafios a serem vencidos, como apresentado pela próxima seção.

4. Alguns desafios da área

Visualização de Informação apresenta uma série de desafios a serem superados, seja na consolidação de suas técnicas e conceitos, seja na utilização de seus resultados por outras áreas. Efetuar pesquisas na direção desses desafios implica o aperfeiçoamento da área e, por consequência, traz melhorias ao processo de gestão de informação.

Nesse sentido, uma abordagem para apresentar os problemas da área é listar seus maiores problemas ainda não resolvidos. Chen (2005) aponta o que ele considera os dez maiores problemas não resolvidos de Visualização de Informação, sumarizados na lista a seguir:

1. Acelerar e incentivar estudos de usabilidade e avaliações empíricas na área, os quais consigam propor e validar sistemas de Visualização de Informação de acordo com metas específicas da área, como a de possibilitar que o usuário reconheça padrões e tendências nos dados visualizados;
2. Entender tarefas perceptivo-cognitivas, relacionadas a identificar e decodificar objetos visualizados, bem como identificar agrupamentos e tendências de pontos em uma representação visual.
3. Definir sistemas de Visualização de Informação que se adaptem ao nível de conhecimento prévio que o usuário possui para entender a informação visualizada.
4. Investir em diferentes aspectos relativos à pesquisa em Visualização de Informação, como: aprendizado de conhecimentos de Semiótica e de Comunicação Visual; solidificação dos fundamentos teóricos da área pela constante comparação com exemplos e novos sistemas; tornar óbvio a pessoas de outras áreas o potencial de Visualização de Informação; e tomar consciência de problemas de outras disciplinas que podem ser resolvidos com o auxílio de Visualização de Informação.
5. Definir medidas intrínsecas de qualidade, ou seja, medidas que consigam responder a perguntas como “quão fiel e eficientemente um sistema de Visualização de Informação representa os dados subjacentes”, ou “até que ponto esse sistema preserva as propriedades intrínsecas dos fenômenos subjacentes”.
6. Prover diferentes níveis de escalabilidade nos sistemas de Visualização de Informação, tanto nos softwares desenvolvidos quanto no hardware que os suporta.
7. Estudar a estética de uma representação visual, e seu impacto no processo de compreensão dos dados representados.
8. Acoplar a sistemas de Visualização de Informação mecanismos de detecção de tendências, com colaborações das comunidades de Mineração de Dados e de Inteligência Artificial.
9. Prover mecanismos que possibilitem a observação de causalidade, formulação de hipóteses (por exemplo, por inferências visuais) e avaliação de evidências existentes em um conjunto de dados, em especial desenvolvendo algoritmos que resolvam evidências conflitantes e removam ruídos de fundo existentes nos dados.
10. Prover meios de visualizar todo um domínio de conhecimento.

Cook e Thomas (2005) complementam o que Chen indica como problema de escalabilidade. Eles apontam que as tecnologias atuais não suportam a escala e a complexidade de crescentes desafios de análises de dados, dando origem a cinco problemas de escalabilidade:

- *Escalabilidade de informação.* Esta questão envolve a capacidade de extrair informação relevante de conjuntos massivos de dados, por meio de métodos para filtrar, reduzir e abstrair dados, e sua adaptação para diferentes audiências.

- *Escalabilidade visual.* Esta questão se relaciona à capacidade de efetivamente representar conjuntos de grande quantidade e/ou dimensionalidade de dados. Considera as metáforas utilizadas nessa representação, as técnicas usadas para interagir com essa representação, e as capacidades perceptivas humanas.
- *Escalabilidade de exibição.* Esta questão envolve o desenvolvimento de técnicas de visualização e interação consistentes e efetivas que independam da área de exibição disponível para representar os dados, considerando os diferentes tamanhos de telas de dispositivos como celulares, PDAs e monitores convencionais, por exemplo.
- *Escalabilidade humana.* Esta questão considera que, apesar de as capacidades humanas não serem escaláveis, o desenvolvimento de técnicas colaborativas de visualização pode permitir que a quantidade de pessoas envolvidas em um processo de análise visual seja escalável.
- *Escalabilidade de software.* Esta questão se relaciona à capacidade de se desenvolver softwares capazes de manipular conjuntos cada vez maiores de dados, e evitando custos de reimplementação e de manutenção de softwares de visualização monolíticos e não escaláveis.

O conjunto de problemas apresentados nesta seção não esgota os problemas de Visualização de Informação, mas indica caminhos pelos quais a área pode evoluir e, por consequência, melhorar seu suporte à gestão de informação.

5. Considerações finais

Este artigo procurou apresentar uma visão geral da área de Visualização de Informação, ressaltando como os seus objetivos estão em harmonia com os propostos para o grande desafio de gestão de informação proposto para a Computação no Brasil até 2016. Detalhando o processo de conversão de dados em estruturas visuais interativas, foi possível apresentar onde se encaixam nesse processo os desafios de tratar grandes volumes de dados e de recuperar informação, considerando a capacidade de análise visual humana e a possibilidade de interação em diferentes etapas desse processo graças a mecanismos interativos providos por sistemas computacionais.

O artigo ressalta, em alguns pontos, que a interdisciplinaridade é um fator importante de Visualização de Informação, uma vez que essa área auxilia outras áreas e delas se beneficia para seu auto-aprimoramento. Essa interdisciplinaridade pode ser praticada dentro da própria Computação, aproveitando resultados de áreas como Interfaces Humano-Computador, Mineração de Dados, Recuperação de Informação (*Information Retrieval*) e Bancos de Dados, dentre outras. Por outro lado, pode ser praticada também externamente à Computação, interagindo com outras áreas como Biologia e Psicologia, e estendendo suas aplicações para beneficiar diferentes atividades e áreas de pesquisa.

Considera-se, por fim, que lidar com o grande desafio de gestão de informação é uma oportunidade ímpar para se investir no potencial de Visualização de Informação como área de pesquisa, incentivando pesquisas nesta área e utilizando seus resultados de maneira interdisciplinar para trazer benefícios sócio-econômico-culturais para o país e para o mundo.

Agradecimentos

O autor agradece à Professora Doutora Heloísa Vieira da Rocha, ao Núcleo de Informática Aplicada à Educação (Nied/Unicamp), ao Instituto de Computação (IC/Unicamp) e ao Laboratório de Administração e Segurança de Sistemas (LAS/IC/Unicamp) pelo apoio recebido direta ou indiretamente na elaboração deste trabalho.

Referências

- Ahlberg, C.; Shneiderman, B. (1994). Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. *Proceedings of CHI'94, ACM Conference on Human Factors in Computing Systems*, New York, pp. 313-317.
- Card, S. K.; Mackinlay, J. D.; Shneiderman, B. (1999) *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufman Publishers.
- Carvalho, A. C. P. de L. F. et al. (2006). *Grandes Desafios da Pesquisa em Computação no Brasil – 2006 – 2016*. <http://www.sbc.org.br/index.php?language=1&subject=8&content=downloads&id=231> (20/Abril/2007).
- Chen, C. (2002). Editorial – Information Visualization. *Information Visualization 1*, pp. 1-4, Palgrave Macmillan.
- Chen, C. (2005). Top 10 Unsolved Information Visualization Problems. *IEEE Computer Graphics and Applications 25* (4), pp. 12-16.
- Eades, P. (1984). A heuristic for graph drawing. *Congressus Numerantium 42*, pp. 149-160. http://www.cs.usyd.edu.au/~peter/old_spring_paper.pdf (15/Novembro/2006).
- Hermann, T.; Ritter, H. (1999). Listen to your Data: Model-Based Sonification for Data Analysis. *Advances in intelligent computing and multimedia systems*, Germany, pp. 189-194.
- Larkin, J.; Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science, 11*(1), pp. 65-99. Apud Card et al. (1999), p. 15.
- Mackinlay, J. D. (1986). Automating the Design of Graphical Presentations of Relational Information. *ACM Transactions on Graphics, 5*(2), pp. 110-141.
- Mazza, R.; Dimitrova, V. (2005). Generation of Graphical Representations of Student Tracking Data in Course Management Systems. *9th IEEE International Conference on Information Visualisation*, London, pp. 253-258. http://www.istituti.usilu.net/mazzar/Web/Web_area/Pubblicazioni/IV05/IV05_mazza_dimitrova.pdf (16/Novembro/2006).
- Mazza, R.; Milani, C. (2005). Exploring Usage Analysis in Learning Systems: Gaining Insights From Visualisations. *Workshop on Usage analysis in learning systems. 12th International Conference on Artificial Intelligence in Education (AIED 2005)*. Amsterdam, The Netherlands, pp. 65-72. http://www.istituti.usilu.net/mazzar/Web/Web_area/Pubblicazioni/aied05/aied-ws2005.pdf (24/Maio/2007).
- Plaisant, C.; Milash, B.; Rose, A.; Widoff, S.; Shneiderman, B. (1996). LifeLines: Visualizing Personal Histories. *Proceedings of CHI'96*, pp. 221-227.

- Rocha, H.V. et al. (2002). Projeto TelEduc: Pesquisa e Desenvolvimento de Tecnologia para Educação a Distância. *IX Congresso Internacional de Educação a Distância da ABED (Associação Brasileira de Educação a Distância)*. http://teleduc.nied.unicamp.br/pagina/publicacoes/premio_abed2002.pdf (15/Novembro/2006).
- Romani, L. A. S. (2000). *InterMap: Ferramenta para Visualização da Interação em Ambientes de Educação a Distância na Web*. Dissertação de Mestrado. Instituto de Computação, Universidade Estadual de Campinas. http://teleduc.nied.unicamp.br/pagina/publicacoes/lromani_disser.pdf (15/Novembro/2006).
- Schatz, M. C.; Phillippy, A. M.; Shneiderman, B.; Salzberg, S. L. (2007). Hawkeye: an interactive visual analytics tool for genome assemblies. *Genome Biology* 2007, 8:R34. <http://genomebiology.com/2007/8/3/R34> (24/Maio/2007).
- Shneiderman, B. (1994). Dynamic Queries for Visual Information Seeking. *IEEE Software*, 11(6), pp. 70-77.
- Silva, C. G. (2006). *Exploração de bases de dados de ambientes de Educação a Distância por meio de ferramentas de consulta apoiadas por Visualização de Informação*. Tese de Doutorado. Instituto de Computação, Universidade Estadual de Campinas. <http://www.ic.unicamp.br/~celmar/tese> (20/Abril/2007).
- Spence, R. (2001). *Information Visualization*. Addison-Wesley.
- Thomas, J. J.; Cook, K. A. (Eds.) (2005). *Illuminating the Path – Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center. IEEE Press.
- Torres, R. S.; Silva, C. G.; Medeiros, C. M. B.; Rocha, H. V. (2003). Visual Structures for Image Browsing. *Proceedings of ACM CIKM'03*, Louisiana, USA, pp. 49-55.
- Walter, M.; Stuart, L.; Borisyuk, R. (2004). The representation of neural data using visualization. *Information Visualization* 3, pp. 245-256.
- Ware, C. (2004). *Information Visualization: Perception for Design*. Morgan-Kaufmann Publishers.