

Visualização em Sistemas Científicos de Informação

Luis Gustavo Neves

June 28, 2013

Abstract

Survey de aplicações de Visualização em Sistemas Científicos

1 Introdução

O projeto de sistemas científicos de informação com funcionalidades de visualização apresenta um desafio para o desenvolvedor, pois estes sistemas são ferramentas de trabalho que almejam serem utilizadas por várias pessoas ao longo de todas as etapas do processo de visualização [4]. Essa abrangência demanda o projeto de diversas funcionalidades de visualização que serão usadas individualmente ou em conjunto.

Cabe ao desenvolvedor projetar adequadamente cada uma dessas funcionalidades, mantendo em mente que serão os pesquisadores os executores da visualização. Até mesmo a possibilidade da visualização não ser feita em sua totalidade pela ferramenta sendo desenvolvida precisa ser considerada.

Como veremos em alguns exemplos, não ficar no caminho dos dados e fornecer meios para exportação dos mesmos é crucial. Isso não só ajuda ao pesquisador a desenvolver suas próprias soluções como também para fornecer concepções de novas formas de visualização que poderão ser incorporadas ao sistema em versões futuras.

Como a correta visualização é sensível a contexto e o desenvolvedor não é o ator final em sua execução, surgem questões em relação ao projeto destes sistemas: O pesquisador vai conseguir usar adequadamente essas funcionalidades? Este conjunto é eficiente ou na verdade uma fonte de distração? Ele vai usar a funcionalidade correta para cada etapa de seu trabalho? A disposição do conjunto favorece isso? São alguns dos questionamentos que precisam ser feitos no desenvolvimento destes sistemas.

A resposta a esses questionamentos demanda um conhecimento mais aprofundado do processo de trabalho dos pesquisadores aos quais a ferramenta se destina. Para projetar uma ferramenta capaz de transmitir a mensagem desejada, é preciso ter alguma noção de como é essa mensagem.

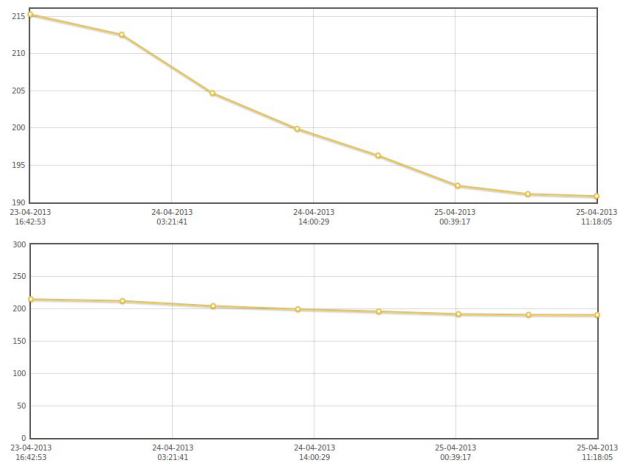


Figure 1: Limites do Eixo Y alterando a visualização de um gráfico.

Além disso, também é necessário um conhecimento dos princípios básicos de visualização de informação, que serão aplicados conjuntamente ao longo do processo de desenvolvimento.

Entretanto, a maior parte da literatura descreve isoladamente a visualização de uma mensagem em um contexto específico, para um papel específico. Observar o destaque de princípios de visualização em sistemas reais demonstra melhor a integração desses princípios.

O objetivo deste trabalho é fazer uma revisão de princípios de visualização expostos em literatura consagrada e destacar a aplicação desses princípios em sistemas científicos reais. Para melhor contextualização, será feita uma breve introdução de cada um desses sistemas.

2 Breve descrição dos Sistemas

2.1 Revestir

Esta aplicação permite o dimensionamento de colunas de revestimento e de produção de poços de petróleo para cargas estabelecidas pelo o projetista ao longo da interface do sistema. Essas cargas tem diversas naturezas e podem ocorrer em todas as etapas da instalação e produção do poço.

O sistema orienta o projetista na aplicação em seu projeto do conjunto de cargas obrigatórias conforme estabelecido por norma (ABNT ???), além de permitir a configurações de outras cargas específicas, de acordo com as particularidades do projeto.

O correto dimensionamento dessas colunas é fundamental para o correto funcionamento do poço e também para prevenção de acidentes. Notoriamente, o acidente no poço Macondo, no golfo do México foi decorrente de uma falha

no projeto de revestimento. Este acidente matou 11 pessoas, gerou um grande vazamento de óleo e a perda definitiva do poço. [2]

No contexto da visualização de informação, o sistema apresenta como pontos de interesse a criação dinâmica de um infográfico do projeto de poço e o uso de tabelas e gráficos associados. Além disso, abordaremos como o sistema foi feito para orientar o pesquisador ao longo do projeto.

2.2 Fouling

Sistema desenvolvido pela Petrobras para fazer o acompanhamento da eficiência de baterias de pré-aquecimento (BPA) das unidades de destilação atmosférica, craqueamento catalítico e coqueamento retardado. Permite avaliar graficamente os resultados obtidos nos cálculos do Fouling TR. (obter descrição mais atualizada).

O uso de gráficos, mapas georeferenciados e infográficos navegáveis, permitindo um acompanhamento virtual na unidade industrial (refinaria ou plataforma) será abordado neste trabalho.

2.3 CMIP

Sistema web para cadastro e consulta de microrganismos bem como a geração de relatórios, da Coleção de Microrganismos da Indústria do Petróleo (CMIP). Este projeto tem como objetivo preservar, manter e distribuir aos grupos de pesquisa do CENPES/Petrobras, culturas de microrganismos viáveis para utilização nos diversos campos da pesquisa biotecnológica aplicados a exploração, produção de petróleo e energias renováveis.

O tratamento e exibição de imagens armazenadas em um banco de dados será abordado como ponto de interesse na visualização de informações. Além disso, serão discutidas as alternativas que poderão ser disponibilizadas para visualização das sequências de código genético de interesse que estão sendo mapeadas nesses microrganismos e que serão armazenadas no banco.

2.4 ARGUS

Ferramenta padronizada e customizada para a análise, diagnóstico e acompanhamento contínuo das unidades de geração de hidrogênio (UGH) e hidrotratamento (HDT)

Atualmente o ARGUS 1.0 contempla as informações de referência de projeto, partida e cálculos de desativação (HDT) e apresenta os valores atuais de variáveis pré-selecionadas da unidade, com defasagem máxima prevista de 30 segundos e um histórico destas variáveis, com períodos escolhidos pelo usuário (desde a partida até o dia anterior ao da consulta) através de médias diárias, que podem ser visualizadas como tabelas e gráficos dinâmicos. Todas as tabelas geradas podem ser exportadas no formato xls e os gráficos como png, além disso, existe também a possibilidade de criação de gráficos personalizados pelos usuários.

O acompanhamento gerado pelo ARGUS 1.0 trará diversas vantagens, tais como a otimização das unidades, com redução de custos operacionais, maior conhecimento da operação das mesmas e possibilidade de uma avaliação mais simples da desativação do catalisador, permitindo uma melhor previsão da época de troca do mesmo.

2.5 SAGAS

Solução Analítica do Gás.

O uso de gráficos em escalas diferentes, inclusive do mesmo conjunto de dados, para se obter a visualização adequada é ponto de destaque nesse sistema.

2.6 STRITA

Análise de testes de injetividade em Reservatórios Estratificados.

O uso de uma animação feita a partir de gráficos e como ela permite a anotação de pontos específicos para exibição estática, bem como será abordado neste trabalho.

3 Revisão de Literatura

3.1 Percepção

A comunicação almejada pelos sistemas científicos é feita pelo meio visual, canal de recepção de informação ao qual são dedicados 70% dos receptores de sentido do corpo humano.[3] O processamento dessa informação é feito em diversas etapas que sumarizam e abstraem a mensagem a ser entendida. Estima-se que apenas 10% da informação percebida pelos olhos seja transmitida da retina ao cérebro pelo nervo ótico e no cérebro ela é primeiramente processada rapidamente em níveis automáticos e inconscientes. [7]

A maioria dos princípios de visualização baseiam-se em manter a percepção da informação neste nível inconsciente, pois isso facilita a leitura e aumenta a chance de que a transmissão da mensagem seja concretizada corretamente. Níveis mais elevados de consciência são recursos limitados que podem não estar disponíveis no momento da leitura da informação. A resposta a essa indisponibilidade pode variar da mera recusa consciente à leitura ou no pior caso, à substituição inconsciente da mensagem por estereótipos e pré-concepções. [1]

3.2 Tabelas

São amplamente utilizadas em sistemas científicos. A partir da definição básica de tabela do padrão HTML, muitas adaptações são feitas, de recursos básicos de visualização, como alternância de cores nas linhas a implementações mais elaboradas, como o uso de fontes coloridas para destaque de erros e pontos de atenção nos dados.

Dentre os recursos disponíveis estão a possibilidade de reordenação das tabelas a partir de alguma coluna selecionada ou para tabelas maiores uma barra de rolagem mantendo o cabeçalho fixo, a edição de dados e o autocompletar de campos.

A possibilidade de edição dos dados por vezes torna a programação de tabelas uma tarefa complexa. Um pequeno conjunto de regras de validação mal especificadas pode rapidamente tornar a implementação impossível por incoerência ou exponencialmente complexa e inviável para tabelas com maior quantidade de dados.

3.3 Gráficos

Gráficos são uma ferramenta básica e muito importante de visualização, com grande variedade: gráficos de dispersão, de barras, histogramas, bolhas, de séries temporais, são alguns exemplos. Apesar de ser uma das técnicas mais básicas de visualização, gráficos derivados do eixo cartesiano ainda são frequentemente apresentados de forma inovadora, proporcionando uma apresentação diferente dos dados, seja pelo aprimoramento de algum elemento ou pela combinação de seu uso.

3.4 Matthias Shapiro - Once Upon a Stacked Time Series

Visualização de informação requer uma amplitude de habilidades e a capacidade de alternar entre essas habilidades rapidamente. As melhores são feitas por pessoas com formação multidisciplinar ou por grupos pequenos trabalhando muito próximos.

A habilidade de contar histórias é uma das mais ignoradas dessas habilidades. Nem todas as visualizações precisam de uma história, mas a maioria tem o objetivo de apresentar seus dados no contexto de alguma história.

A maioria das histórias de visualização partem de uma questão que orienta o público ao tópico e contexto no qual o dado é mais significativo.

3.5 Danyel Fisher - Animation for Visualization: Opportunities and Drawbacks

Animação ajuda na visualização ou apenas serve para confundir?

A criação de animações atualmente é bastante facilitada tanto pelas tecnologias da Web quanto pela vasta literatura conceitual disponível sobre como produzi-las.

Animações podem ser uma técnica poderosa quando usadas apropriadamente ou muito ruim quando mal utilizadas. Podem aumentar o apelo visual, mas podem dificultar a exploração visual.

Podemos entender a animação como uma série de mudanças visuais entre os quadros. É fácil entender o que aconteceu quando o número de mudanças é pequeno e a animação fica mais complexa conforme as mudanças aumentam.

Objetos se movendo na mesma direção e velocidade serão percebidos como um único grupo. Objetos com suas próprias trajetórias se destacarão visualmente como separados. É difícil acompanhar mais de quatro ou cinco objetos separadamente. Os olhos passam a seguir apenas alguns e tratam os outros como ruído (Cavanagh and Alvarez 2005).

A animação é bastante utilizada em visualização científica por partir de dados tridimensionais e ser muito difícil mostrar os processos ao longo do tempo de outra forma.

Em contraste, a visualização de dados normalmente lida com espaços de dados abstratos onde as dimensões não correspondem às do mundo real. Por isso, existe comparativamente muito menos exemplos de animações publicadas na comunidade de visualização de informação.

Em mais de 100 estudos de visualização e animação, o uso da animação não superou o uso de diagramas estáticos. Em um estudo de visualização de algoritmos, as animações se mostraram eficientes quando a animação permitia que os algoritmos fossem manipulados e refletir essas alterações. Passivamente assistir as animações não se mostrou mais eficiente que outros métodos.

A visualização GapMinder do Hans Rosling é um caso de sucesso, mas apresenta mais informação que os estudos de percepção demonstram ser possível processar. A apresentação guiada de Rosling parece ser crucial para o sucesso da visualização.

Estudos no Microsoft Research mostraram que a animação é uma forma mais lenta e menos precisa de fornecer informação. Para explorar os dados por conta própria, os usuários precisam reexibir a animação várias vezes para verificar algum fato. Os usuários claramente desejam ter a capacidade de controlar a animação no tempo e isso pode torná-la menos eficiente que uma série de imagens estáticas, pois estas últimas permitem que se vá diretamente para a parte que se quer ver.

No entanto, animações são emocionalmente poderosas e mais envolventes. No contexto de apresentação, os dados e a mensagem é conhecida pelo apresentador e o público assiste. A animação é apresentada diretamente, sem muitas voltas, ao contrário do que faria alguém explorando os dados. Nesse contexto, a animação é recomendada.

3.6 Eddie Jabbour, Julie Steele - Mapping Information: Redesigning the New York City Subway Map

O tamanho do mapa e como ele acomoda as informações que devem ser mostradas é um fator importante, mas não tão relevante quanto a forma como os dados são apresentados.

A geografia é o ponto de partida para o projeto de mapas e parte da experiência de quem vai usá-los. Ela precisa ser estilizada e simplificada, mas os relacionamentos dos pontos de referência devem ser preservados.

Além da geografia, as linhas devem denotar os caminhos a serem seguidos. Mais do que meramente serem usadas separadamente para facilitar a distinção das linhas, as cores podem ser organizadas em espectro para fornecer mais

semântica ao mapa. O uso do cores é importante para tornar o mapa mais vaculhável, além de apenas mais legível.

Mapas podem envolver questões emocionais. Algumas cores devem ser preservadas, como o azul representando água. Da mesma forma, representar elementos marcantes por suas formas tem mais apelo que rótulos em texto.

O texto também pode ser separado em tons, guardando o preto para as informações mais relevantes. Bem organizado, o mapa fornece camadas de informação sem comprometer sua clareza e funcionalidade.

3.7 Visualizando Séries Temporais

3.8 Anscombe's Quartet

4 Sagas

No sistema SAGAS, por exemplo, a tabela apresentada na parte superior da página pode ser visualizada através de três gráficos iterativos. O primeiro apresenta uma parte dos dados da tabela em uma interpolação de pontos sobreposta por um gráfico de dispersão, com ambos os eixos em escala logarítmica. No segundo, apenas o eixo y tem escala logarítmica e o terceiro apresenta os eixos em escala linear. A legenda unificada dos três gráficos apresenta os valores de forma iterativa, atualizados conforme o passar do mouse nos pontos plotados. O conjunto desses valores em cada ponto equivale a uma linha da tabela.

A implementação desses gráficos pode se tornar complicada. No caso do sistema SAGAS, programação adicional teve que ser realizada para exibir os eixos com escala não linear e para unificar a legenda. Além disso, para plotagem dos gráficos de dispersão foi necessário um algoritmo de amostragem (sampling) teve que ser aplicado. Para as curvas interpoladas, ao contrário foram fornecidas as opções de interpolação linear a cada ponto ou o uso de splines.

Um recurso bastante utilizado em nossos sistemas consiste em fornecer um pequeno gráfico com toda a curva que pode ser parcialmente selecionado. O gráfico maior apresenta apenas a seleção, que pode ser examinada em detalhe, enquanto o gráfico pequeno mantém a idéia da curva como um todo.

São o tipo de gráfico mais utilizado mas detalhes como diferenças de formato de hora e data, a internacionalização e a variabilidade intervalo de tempo a ser mostrado tornam complexa a implementação de programas computacionais dedicados a produção desses gráficos.

5 Fouling

No sistema Fouling, séries temporais são mostradas utilizando o pacote Flot. Essa biblioteca permite a produção com gráficos com bastante iteratividade, sendo possível redefinir os intervalos de tempo e apresentar várias séries diferentes no mesmo gráfico, na mesma ou em várias escalas.

As funcionalidades da biblioteca tiveram que ser extendidas para se adequar ao sistema Fouling. O algoritmo que controla a marcação nos eixos (tick marks) teve que ser refeito. Funcionalidades como a exibição de legendas e marcação de valores de acordo com o movimento do mouse, bem como o destaque de qual eixo estava controlando cada ponto selecionado também tiveram que ser programados.

[5]

Mapas distorcidos e infográficos são técnicas de visualização bastante popularizadas e a implementação em nossos sistemas é frequentemente solicitadas. Quando essa demanda é relevante, a experiência com o Fouling nos serviu de referência para recomendar que o formato SVG seja utilizado.

Entretanto, a criação manual de mapas e infográficos é trabalhosa e pode demandar mais tempo e recursos que o disponível para o projeto. Em alguns casos, a criação de mapas e diagramas automaticamente é possível e para isso utilizamos ferramentas como o graphviz ou futuramente a criação de SOMs com o pyMVA.

Um caso de uso mais elaborado de pós-processamento, ocorreu no sistema Fouling, onde o formato SVG foi escolhido para ser usado na visualização de mapas georeferenciados. Os arquivos com os mapas das unidades de refinaria foram desenhados no Inkscape e nos lugares adequados, foram colocadas palavras chaves predefinidas para indicar qual variável de medição deveria ser apresentada nessas posições.

Com o uso de expressões regulares, foi possível determinar rapidamente quais variáveis de medição deveriam ser usadas. A posição correta foi obtida automaticamente, meramente substituindo a palavra chave (já posicionada corretamente) pela variável de medição correspondente.

Também é possível adicionar funcionalidades ao arquivo SVG. No caso do sistema Fouling, foram adicionados links para exibição de gráficos nas variáveis e tooltips descrevendo seu significado.

No sistema Fouling, o software Inkscape foi utilizado para edição das figuras em SVG e uma extensão foi desenvolvida para ele. No futuro, pretendemos utilizar o software livre SVG Edit, que por ser feito em javascript nos permitirá manter todo o projeto no ambiente Web.

6 Tecnologias

6.1 SVG

Scalable Vector Graphics (SVG) [6] é um formato de imagem vetorial 2D baseado em XML com especificação aberta desenvolvida pelo World Wide Web Consortium (W3C). Esse formato foi apresenta diversas características que o tornam desejável para o contexto de visualização de informação. Dentre elas:

- É um formato de imagem vetorial, podendo se adequar a qualquer necessidade de tela e de resolução no momento da exibição.
- É um formato de texto, podendo ser processado programaticamente após sua produção.
- Per-

mite interatividade e suporta animações. - Pode ser exibido em web browsers. - Suporta navegação por hyperlinks. - É programável em linguagem JavaScript. - Possui ferramentas de edição poderosas e livres, como o software Inkscape ou o SVG Edit. - Pode ser totalmente processado no servidor Web e ainda prover interatividade no Browser.

As imagens em arquivos SVG podem ser geradas a partir de softwares de edição manual como o Inkscape ou a partir de bibliotecas de software como o GraphViz, Matplotlib ou Pygal. O ajuste e extensão de funcionalidades desses arquivos é bastante facilitado pelo fato do formato SVG ser meramente um formato de texto. As transformações podem ser realizadas a partir de substituições de string e do uso de expressões regulares.

Para exemplificar o processo, apresentamos a figura abaixo, gerada a partir da biblioteca Pygal e salva em formato SVG. Este arquivo foi lido e processado, para tradução do título e da legenda foi feita com substituição de string simples.

6.2 SOM

Self Organizing Maps (SOM - mapa auto organizado) também conhecidos como mapas ou redes de Kohonen são um tipo de rede neural artificial treinada usando aprendizado não supervisionado para produzir representações com menos dimensões (tipicamente duas dimensões) do espaço definido pelos exemplos de treino. Esses mapas são úteis para visualização de dados com muitas dimensões pois usam uma função de vizinhança que preserva as propriedades topológicas do espaço de entrada. Esse modelo foi descrito inicialmente como uma rede neural artificial pelo professor finlandês Teuvo Kohonen.

7 Futuros Trabalhos

References

- [1] R.F. Baumeister and J. Tierney. *Willpower: Rediscovering the Greatest Human Strength*. Penguin Group US, 2011.
- [2] Robbie Brown. <http://www.nytimes.com/2010/05/30/us/30rig.html?pagewanted=all>, May 2010.
- [3] S. Few. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. ANALYTICS Press, 2012.
- [4] Ben Fry. *Visualizing Data*. 1st edition, 2008.
- [5] Laursen Schnur. <http://www.flotcharts.org/>.
- [6] W3C. <http://www.w3.org/graphics/svg/>.
- [7] Matthew Ward, Georges Grinstein, and Daniel Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications*. A. K. Peters, Ltd., Natick, MA, USA, 2010.