

Avaliando o Uso de Múltiplas Visões Coordenadas com Tarefas de Análise de Dados

¹Michele Maciel, ¹Paulo Godinho, ¹Bianchi Meiguins, ²Aruanda Meiguins,
²Marcelo Garcia Meiguins

¹Programa de Pós-Graduação em Ciência da Computação
Universidade Federal do Pará (UFPA)
²Área de Ciências e Tecnologia
Centro Universitário do Pará (CESUPA)
Belém-Pará-Brasil

{chelepsmaciel, piagodinho}@gmail.com; bianchi.serique@terra.com.br, {aruanda,
marcelo}@redeinformatica.com.br

Abstract. The use of information visualization tools that support multiple coordinated views for data exploration and analysis has become more frequent. However, few usability studies have focused this group of tools. The goal of this paper is to perform usability experiments with tools supporting multiple coordinated views on the same screen, in separate screens and tools that support each technique individually. To ensure a fair evaluation, a complexity-based task taxonomy was developed. The tools were therefore used to answer questions of the same level of complexity. The analysis of the correctness rate had shown that the coordinated features are better explored for tasks of higher complexity. When comparing the two coordinated approaches, the correctness rate was better when the views were presented in the same screen.

Resumo. O uso de ferramentas de visualização de informação que suportam múltiplas visões coordenadas para análise e exploração de dados tem se tornado mais freqüente. Entretanto, existem poucos estudos de usabilidade voltados para esse grupo de ferramentas. O objetivo deste artigo é realizar experimentos de usabilidade com ferramentas que suportem múltiplas visões coordenadas na mesma tela, em telas separadas e ferramentas que suportem cada técnica individualmente. Para a avaliação foi desenvolvida uma taxonomia de tarefas baseada na complexidade das tarefas. As ferramentas foram usadas para responder questões do mesmo nível de complexidade. De maneira geral, os estudos mostraram que a abordagem coordenada é melhor utilizada quanto mais complexa for a tarefa. Dentre as duas abordagens coordenadas, as visões na mesma tela tiveram maior taxa de acerto e menor taxa de erro.

Palavras chaves: Visualização de Informação, Múltiplas Visões Coordenadas, Usabilidade.

1 Introdução

As ferramentas de visualização de informação são utilizadas no auxílio ao desenvolvimento de tarefas como exploração e análise de grandes volumes de dados. Entretanto ainda existem alguns desafios nesta área no que diz respeito ao entendimento do usuário, ao processo de desenvolvimento de suas ações para a solução de problemas e a usabilidade dessas ferramentas.

Segundo Chen [1], o problema de usabilidade na área de visualização de informação é um reflexo da significativa carência de estudos, pesquisas e análises mais aprofundadas das técnicas e ferramentas de visualização de informação. Contudo, novas abordagens para o uso de técnicas de visualização de informação vêm surgindo, como a exploração e análise de dados através de Múltiplas Visões Coordenadas, onde cada visão pode ser representada por uma técnica diferente e alterações ou configurações realizadas em uma das visões são refletidas nas demais visões. Enquanto isso, as dúvidas ainda persistem: Por que se deve usar? Quando se deve usar? Qual a melhor forma de usar essa abordagem? É mais difícil de utilizar? É mais produtiva? Qual o perfil de usuário que deve utilizá-la?

Este artigo visa contribuir para a discussão do impacto do uso de múltiplas visões coordenadas na exploração e análise de dados, e verificar como a coordenação e as múltiplas visões influenciam para a realização das tarefas dos usuários.

Para a análise foi utilizada uma ferramenta de múltiplas visões coordenadas denominada PRISMA [12], a qual utiliza as técnicas: coordenadas paralelas, dispersão de dados e treemap. A partir do PRISMA foram criadas 5 ferramentas menores: 3 com técnicas individuais, uma com 3 técnicas coordenadas na mesma tela e por fim uma ferramenta com 3 técnicas coordenadas em telas diferentes.

Para proporcionar uma análise justa, as tarefas desenvolvidas foram classificadas por complexidade, todas as ferramentas foram utilizadas para resolver tarefas de mesma complexidade. Assim, foi desenvolvida uma taxonomia de tarefas em níveis de complexidade baseada na quantidade de atributos envolvidos e no tipo de resposta requerida. Foram realizados testes com 10 usuários e três bases de dados.

O artigo apresenta uma visão geral sobre Visualização de Informação e Múltiplas Visões Coordenadas; algumas taxonomias; descrição dos testes de usabilidade e do resultado da análise; considerações e trabalhos futuros.

2 Visualização de Informação

As ferramentas de visualização de informação são ferramentas exploratórias que permitem ao usuário investigar, procurar informações e experimentar cenário e espaço visual para gerar hipóteses e desenvolver um entendimento melhor do conjunto de dados apresentados [2][1].

Para desenvolver sistemas de visualização é necessário entender melhor como o ser humano interage com a informação, quais tarefas ele irá desempenhar com o apoio dessa ferramenta. Segundo Carr [3], uma boa ferramenta de visualização deve possuir características definidas em função das tarefas executadas pelo usuário, entre elas destaca-se: Visão geral: o usuário precisa ganhar noção sobre todos os dados que

serão analisados; Zoom: permite focar em certo subconjunto dos dados para análise; Filtro: permite a redução do tamanho do conjunto de dados visualizados, eliminando itens baseados em seus atributos e Detalhes sob demanda: permite ao usuário enxergar detalhes relacionados a um único ou um conjunto de dados.

Existem sistemas de visualização de informação que oferecem apenas uma visão dos dados, entretanto, existem sistemas que apresentam mais de uma visão dos dados, através da utilização de várias técnicas, para fornecer aspectos diferentes dos dados com intuito de auxiliar no processo de investigação.

2.1 Múltiplas Visões Coordenadas

Sistemas de múltiplas visões utilizam duas ou mais visões distintas para auxiliar o processo de investigação de uma única entidade conceitual [4]

As principais características de coordenação em visualização de informação são flexibilidades em relação aos dados - utilização de mais de um conjunto de dados, visões – escolha de um tipo de representação visual para determinado conjunto de dados e coordenação - definem as características da coordenação entre cada par de visões [5].

Para a utilização de sistemas de visualização de informação com múltiplas visões coordenadas, são freqüentemente encontradas as seguintes recomendações [4]: usar múltiplas visões quando há diversidade de atributos, modelos, perfis de usuário, níveis de abstração ou gênero; quando as visões diferentes destacam correlações ou disparidades; quando os dados são complexos e deve-se usar múltiplas visões minimamente, justificando o uso de múltiplas visões versus custo de aprendizado do usuário e espaço de visualização.

As principais possibilidades de coordenação entre as visões podem ser representadas por ações de [7] Seleção, Filtro, Cor, Transparência, Tamanho, Ordenação Rótulo e Manipulação de Atributos.

Assim, é importante saber quando usar ou não as múltiplas visões coordenadas, sempre considerando as tarefas a serem realizadas, os níveis de complexidade, aplicação dessas tarefas e sua avaliação.

3 Algumas Taxonomias

Pillat [7] desenvolveu uma taxonomia de tarefas para técnicas de visualização de informações multidimensionais, com o intuito de identificar as tarefas que os usuários realizam enquanto utilizam uma técnica de visualização. A taxonomia é composta por sete tarefas: identificar, determinar, visualizar, comparar, inferir, configurar e localizar. Tais tarefas podem ser consideradas como objetivos ou necessidades dos usuários enquanto utilizam uma técnica de visualização de dados multidimensionais. Para avaliar a taxonomia proposta a autora realizou ensaios de interação com o usuário, para o qual foram estabelecidas quatro questões / tarefas a serem respondidas durante os testes.

Wiss e Carr [8] realizaram um estudo comparativo de três visualizações em ambientes tridimensionais para saber quais fatores são importantes para realização das

tarefas. Neste estudo foram utilizadas as visualizações Cam Tree, Information Landscape e Information Cube para o domínio de sistema de arquivos hierárquicos. As três tarefas utilizadas no estudo foram baseadas nas sete tarefas de visualização de informação descritas por Shneiderman: overview, zoom, filter, details-on-demand, relate, history e extract.

Laskowski e Plaisant [9] direcionam a avaliação de sistemas de visualização de informação para três níveis: nível de componente, nível de sistema e nível de ambiente de trabalho. O nível de componente inclui algoritmos individuais, representações visuais, técnicas interativas e designs de interface. O nível de sistema foca em interfaces que combinam e integram múltiplos componentes e precisam ser avaliados comparando-os com a tecnologia utilizada pelos usuários alvo. Essas avaliações normalmente ocorrem em experimentos controlados em laboratórios, usando cenários e pequenas tarefas. O nível de ambiente de trabalho refere-se à descrição das tarefas reais dos usuários em seu ambiente natural.

Henry e Fekete [10] apresentaram uma taxonomia de tarefas com o intuito de avaliar como a visualização de informação apóia a exploração de tabela visual de dados. Desenvolveram um experimento controlado para verificar como o layout da tabela influencia no entendimento do usuário e no seu processo de exploração.

Os trabalhos apresentados contribuíram como embasamento teórico para esta pesquisa, pois fomentaram a criação de uma taxonomia de tarefas de acordo com níveis de complexidade, a qual foi utilizada em testes com usuários para verificar o impacto das múltiplas visões na análise de dados.

4 Experimento

Para avaliar e projetar ferramentas de visualizações coordenadas com melhores níveis de usabilidade é necessário compreender as tarefas executadas pelas pessoas no uso destas técnicas, de modo a aplicar este entendimento no desenvolvimento de suas aplicações [11]. Diversas ferramentas ou técnicas de visualização de dados multidimensionais têm sido desenvolvidas nos últimos anos, mas poucos trabalhos abordam questões relacionadas à avaliação de sistemas ou aplicações que implementam estas técnicas [7]

Assim, foi criada uma taxonomia de tarefas baseada em grau de complexidade para verificar a influência das múltiplas visões coordenadas na análise de dados. A ferramenta PRISMA, que trabalha com múltiplas visões coordenadas, foi utilizada neste projeto. O PRISMA gerou cinco “novas” ferramentas, três ferramentas que possuem técnicas de visualização de informação individuais: dispersão de dados, treemap, e coordenadas paralelas. As duas outras ferramentas possuem visões coordenadas das técnicas de visualização já citadas, na mesma tela e em telas separadas.

A exploração dessas cinco ferramentas foi a estratégia utilizada para realizar uma análise comparativa (entre as Múltiplas Visões Coordenadas em uma tela, Múltiplas Visões Coordenadas em tela separadas e técnicas individuais) que pudesse fundamentar o objetivo da pesquisa que é verificar o impacto das Múltiplas Visões na análise de dados.

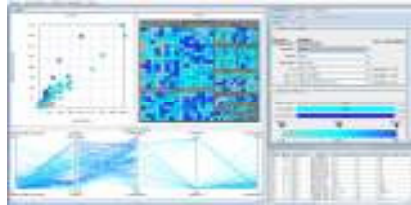


Fig. 1. Visões Coordenadas na mesma tela

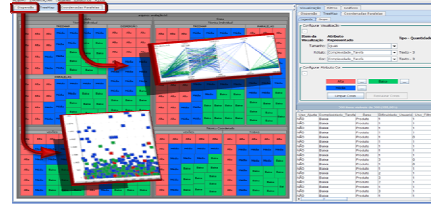


Fig. 2. Técnicas em telas separadas

A taxonomia proposta divide o grau de complexidade em alto, médio e baixo. Esta classificação se refere ao tipo de resposta e à quantidade de atributos envolvidos:

- Tarefa de complexidade baixa:
 - Requer uma resposta simples, como um “sim” ou “não”;
 - Não requer análise de vários atributos, apenas um ou dois, por exemplo.
- Tarefa de complexidade média:
 - Requer uma resposta simples, mas concreta, não apenas positiva ou negativa, por exemplo: Qual a origem dos carros mais vendidos?
 - Requer a análise de três ou quatro atributos;
- Tarefa de complexidade alta:
 - Requer uma resposta mais complexa, como uma análise que envolva comparações, por exemplo.
 - Requer a análise de vários atributos.

Foram utilizadas três bases de dados para os testes, uma sobre dados das copas do mundo de futebol, uma sobre venda de produtos eletrônicos e outra sobre carros. Para cada base foram geradas 15 tarefas, cada grupo de cinco tarefas pertencentes a cada um dos níveis de complexidade. As tarefas foram construídas baseadas na taxonomia proposta, por exemplo, o grupo de tarefas de complexidade baixa foi criado de maneira que exigisse uma resposta simples, e que para alcançar essa resposta, poucos atributos estariam envolvidos, além do que, a tarefa deveria ser algo relevante.

Os usuários (dez por base) possuíam boas habilidades de interação com o computador, porém sem conhecimento sobre técnicas de visualização de informação, também responderam ao questionário de interação e execução das tarefas e ao questionário de avaliação do treinamento. As bases se encontram disponíveis em <http://paginas.terra.com.br/informatica/bianchiserique/bases/copa.txt>, [produtos.txt](http://paginas.terra.com.br/informatica/bianchiserique/bases/produtos.txt) e [carros.txt](http://paginas.terra.com.br/informatica/bianchiserique/bases/carros.txt).

Segundo Preece [11], normalmente, em testes com usuários participam a partir de 5 (cinco) até 12 (doze) usuários. A autora relata também os testes de usabilidade de uma comunidade virtual chamada HutchWolrd, no qual participaram sete usuários.

No desenvolvimento da pesquisa, cada usuário deveria escolher uma entre as perguntas de um determinado nível de complexidade, sendo que tal tarefa poderia ser realizada em qualquer uma das ferramentas. E assim, a pesquisa prosseguia até o fim das 15 tarefas. Ou seja, qualquer tarefa poderia ser realizada através de qualquer ferramenta, desde que o usuário utilizasse cada uma das ferramentas para responder ao menos uma tarefa de cada nível de complexidade.

4.1 Análise

A análise dos resultados seguiu duas abordagens. A primeira foi realizada pela complexidade das tarefas classificadas pelos usuários e a segunda pela complexidade das tarefas classificadas pelos avaliadores.

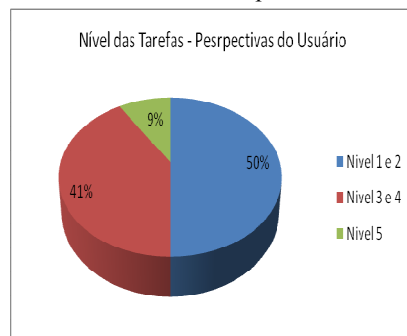


Fig. 3. Complexidade das tarefas – Perspectiva do usuário

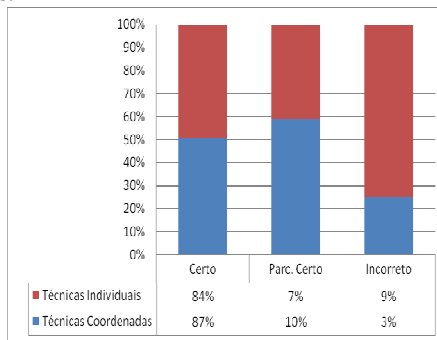


Fig. 4. Taxa de acertos por tarefas de complexidade-nível 1 e 2

Após completar um grupo de tarefas (por complexidade), era solicitado ao usuário classificar as tarefas realizadas com valores de 1 a 5, considerando a complexidade das tarefas. Das tarefas que o usuário classificou como dificuldade 1 e 2 somam 50% do total (Figura 3). Dessas tarefas, 41% foram resolvidas com técnicas que envolviam coordenação, com 87% das respostas certas, 10% parcialmente certas e 3% erradas (Figura4). Para tarefas de dificuldade 3 e 4, a taxa de acerto para tarefas coordenadas foi de 75% certas, 16% parcialmente certas e 9 % erradas (Figura 5). Para tarefas de dificuldade 5, a taxa de acerto pra tarefas coordenadas foi de 80% , 13% parcialmente certas e 7 % erradas (Figura 6).

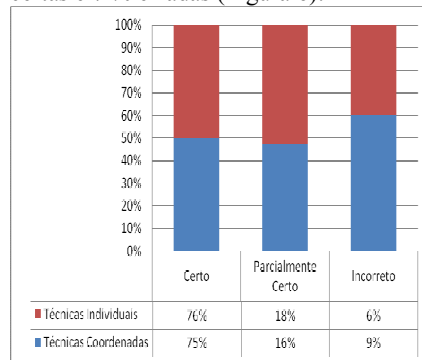


Fig. 5. Taxa de acertos por tarefas de complexidade-nível 3 e 4.

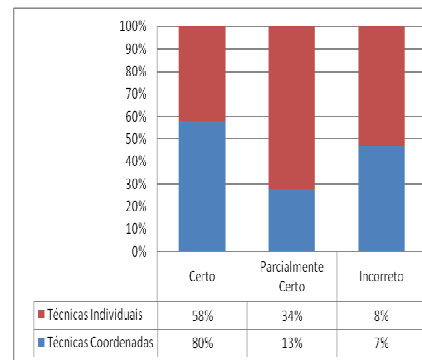


Fig. 6. Taxa de acertos por tarefas de complexidade-nível 5.

Para complexidade de tarefas definida pelos avaliadores, a taxa de acerto foi diminuindo conforme a complexidade das tarefas aumentava (figura 7). Para tarefas

de baixa complexidade a taxa de acerto foi similar tanto para técnicas individuais quanto para coordenadas (Figura 8).

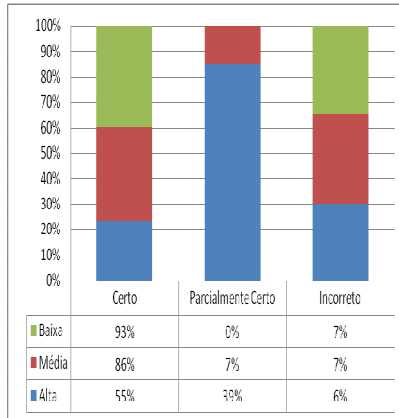


Fig. 7. Taxa de acertos por complexidade de tarefas – perspectiva dos avaliadores.

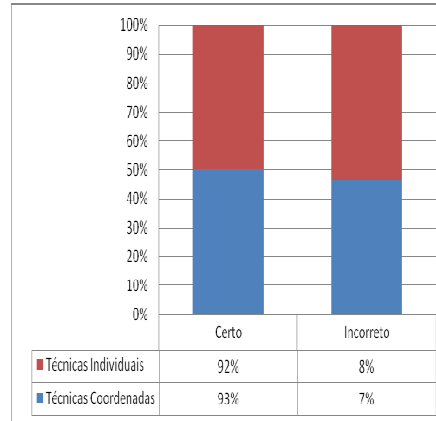


Fig. 8. Taxa de acertos por técnica em tarefas de complexidade baixa.

Para tarefas de complexidade média, as técnicas coordenadas obtiveram taxa de 88% contra 83% das técnicas individuais de respostas certas (Figura 9), e para tarefas de alta complexidade as técnicas coordenadas obtiveram 61% contra 58% das técnicas individuais (Figura 10).

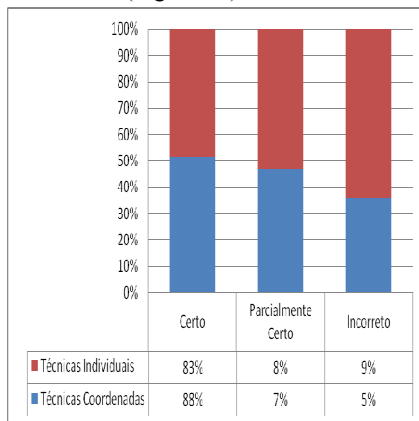


Fig. 9. Taxa de acertos por técnica em tarefas de complexidade média.

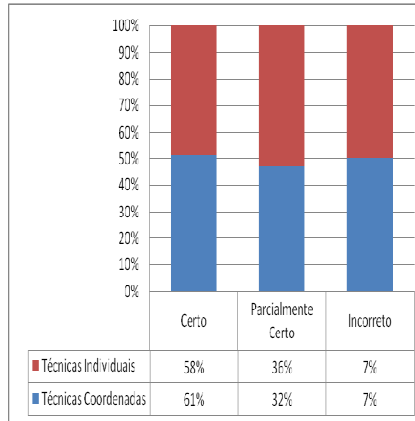


Fig. 10. Taxa de acertos por técnica em de tarefas de complexidade alta.

Pode-se se afirmar que, em média, quanto maior a complexidade das tarefas mais úteis são as técnicas coordenadas, tanto para tarefas definidas pelos usuários quanto pelos avaliadores. Para tarefas menos complexas os resultados são similares entre técnicas individuais e coordenadas. Em relação ao tempo para a solução das tarefas,

no intervalo entre 0.5 e 4 minutos estão incluídos 51% das tarefas de baixa complexidade. No intervalo entre 4 a 10 minutos estão 33% das tarefas de complexidade média e quanto às tarefas que levaram mais de 10 minutos para serem resolvidas, 73% delas eram de alta complexidade.

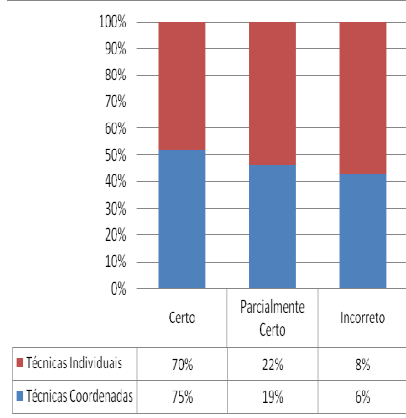


Fig. 11. Taxa de acertos por técnica em técnica de visualização.

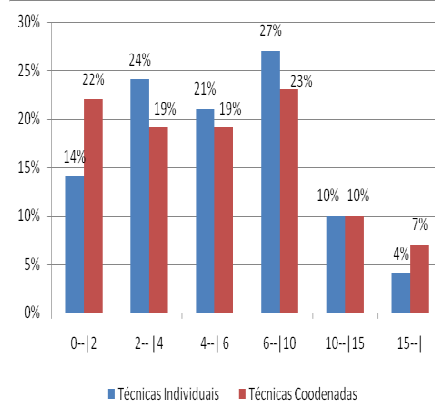


Fig. 12. Tempo de resolução das tarefas por tarefas de complexidade alta e média.

Houve uma evolução quanto ao uso das ferramentas e entendimento das técnicas utilizadas. É possível observar através da evolução da taxa de acerto de uma base para outra, a base de dados sobre copas do mundo foi a primeira a ser aplicada (Figura 13). Em relação as técnicas coordenadas na mesma tela e telas diferentes, destaca-se um percentual de acerto maior e de menor erro para técnicas coordenadas na mesma tela (Figura 14).

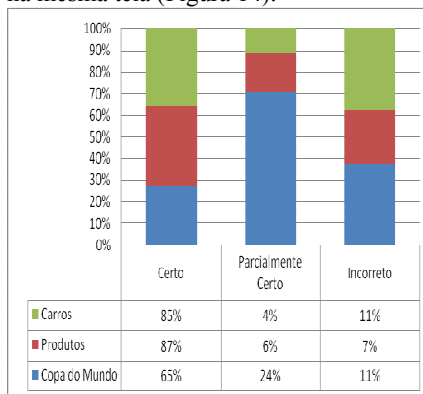


Fig. 13. Taxa de acertos por base de dados.

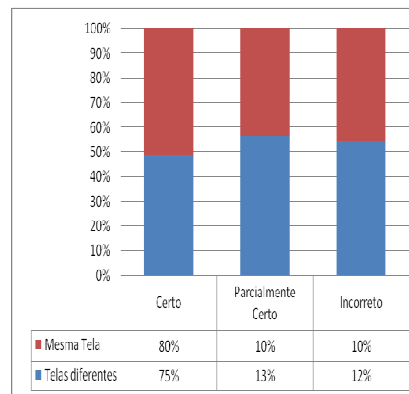


Fig. 14. Acertos por visão coordenada.

Em 26% das tarefas os usuários interagiram com a tabela de dados para comparar dados (20%), aprender sobre os dados (5%) e selecionar itens através do *brushing*

(2%). Os filtros foram utilizados em 83% das tarefas. Detalhes sob demanda foi usado por 59% dos usuários. Esse comportamento foi similar na utilização de qualquer uma das ferramentas. Os gráficos auxiliares foram utilizados com mais frequência nas técnicas individuais (6%) do que nas técnicas coordenadas (4%).

Segundo Carr [3] gerenciar o espaço de um monitor é uma tarefa complexa no campo da VI, a alternativa padrão é a utilização de janelas e algo relevante é o relacionamento estabelecido entre elas, já que visualização de informação é muito mais do que operações gráficas. Assim, a análise dos dados se torna mais prática através da análise comparativa com todos os elementos na mesma tela, ao alcance de um movimento dos olhos e caso o usuário sinta a necessidade de expansão dos dados, ele poderá fazer isso também.

De acordo com 90% dos usuários a experiência com a ferramenta de visualização foi considerada fácil, prática, rápida e lúdica. Foi perceptível o envolvimento do usuário na utilização da ferramenta. Entretanto, os usuários argumentaram que uma ferramenta desse tipo exige interpretação gráfica, o que talvez pudesse atrapalhar o desenvolvimento da tarefa. Isso também foi comprovado pelos avaliadores durante os testes, pois os usuários realizavam o procedimento correto na ferramenta durante a realização de uma tarefa, encontravam a resposta, mas não conseguiam “enxergá-la”. Mas apesar disso, os usuários sustentaram a premissa de que a forma gráfica era de suma importância pela quantidade de dados. Quanto ao perfil do usuário, foi possível observar que qualquer pessoa é capaz de utilizar bem uma ferramenta de visualização de informação, entretanto também é notável que as pessoas que possuíam certo conhecimento acerca do conteúdo da base de dados, realizavam as tarefas com maior habilidade por conseguirem interpretar os dados com maior facilidade.

5 Considerações Finais

O objetivo deste artigo é primeiramente contribuir com a discussão sobre a usabilidade de ferramentas de visualização, em particular com ferramentas que usam abordagem de múltiplas visões coordenadas. O artigo apresentou um estudo com a ferramenta PRISMA, uma ferramenta que usa múltiplas visões coordenadas, que foi subdividida em cinco ferramentas, três dessas usando técnicas individuais e duas com abordagem coordenada e múltiplas técnicas de visualização. Além disso, foram desenvolvidas tarefas baseadas em uma taxonomia de complexidade para desenvolvimento do estudo.

De maneira geral, os estudos mostraram que a abordagem coordenada é melhor utilizada quanto mais complexa for a tarefa, baseada na taxa de acerto. Dentre as duas abordagens coordenadas, as visões na mesma tela tiveram maior taxa de acerto e menor taxa de erro. Ou seja, relacionar informações representadas em técnicas diferentes e em telas separadas pode não ser uma tarefa fácil para usuário. Contudo, muitos usuários comentaram após os testes que o impacto inicial das múltiplas visões na mesma tela era grande, e que 37% prefeririam ter apenas duas visões na tela em vez três. Apesar de ser possível dimensionar a área das visões e deixar visível apenas uma ou duas visões, esse recurso não foi utilizado. Questionados entre as duas

abordagens coordenadas, 73% preferiram as técnicas de visualização coordenadas na mesma tela.

Outro ponto de discussão é quanto as respostas dos usuários, recomenda-se, principalmente para questões mais complexas, que exista um padrão de resposta para orientação dos usuários, isso facilita no desenvolvimento das respostas por parte dos usuários e no momento de avaliar as respostas fornecidas.

Como trabalhos futuros destacam-se:

- Reavaliar as questões desenvolvidas e criar padrões de respostas;
- Utilizar outras bases e ampliar o estudo;
- Utilizar outras técnicas e ferramentas de visualização;
- Utilizar outros perfis de usuário.

Referências

1. Chen, C. and Cserwinski, M. (2000) "Empirical Evaluation of Information Visualization: An introduction", *Int'l J. Human-Computer Studies*, vol.53, pag 631-635.
2. Spence, R. *Information Visualization*. Addison Wesley - ACM Press, 2001. 459 p.
3. Carr, D. A. (1999) "Guidelines for Designing Information Visualization Applications". *Proceedings of ECUE'99*. Stockholm, Sweden. December. 1999.
4. Baldonado, M. Q. W.; Woodruff, A.; Kuchinsky, A. Guidelines for using multiple views in information visualization. *Proceedings of the working conference on Advanced Visual Interfaces*, pp. 110 – 119. Palermo. Italy. 2000.
5. North, C. and Shneiderman, B. (2000) "Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata". *ADVANCED VISUAL INTERFACES International Working Conference* May 23-26.
6. Pillat, R. M.; Freitas, C. D. S. Coordinating Views in the InfoVis Toolkit. *Proceedings of Advanced Visual Interface*. pp. 496-499. Venezia, Italy. 2006.
7. Pillat, R. M., Valiati, E. R. and Freitas, C. M. D. (2005) "Experimental Study on Evaluation of Multidimensional Information Visualization Techniques". In: *CLIHC'05*, Cuernavaca - Mexico. p. 20 - 30.
8. Wiss U., Carr D.: An empirical study of task support in 3d information visualizations. In *Proceedings of the International Conference on Information Visualisation (IV) (1999)*, IEEE Computer Society, pp. 392—399.
9. Laskowski, S., Plaisant, C., Evaluation Methodologies for Visual Analytics section 6.1, in Thomas, J., Cook, K. (Eds.) *Illuminating the Path, the Research and Development Agenda for Visual Analytics*, IEEE Press (2005) 150-157.
10. Henry, Nathalie, Fekete, Jean-Daniel, Evaluating Visual Data Understanding. *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*; p. 1-5. Venice, Italy. 2006.
11. Preece, J., Rogers, Y. and Sharp, H. *Interaction Design: beyond human-computer interaction*. Wiley. 2005
12. PRISMA – A Multidimensional Information Visualization Tool Using Multiple Coordinated Views. *Proceedings of the 11th International Conference on Information Visualization*, pp. 23-32. Zurich, 2007.