

APPWM - Agrupamento Personalizado de Pontos em Web Maps usando um modelo multi-dimensional

Marcio Bigolin¹, Helena Grazziotin Ribeiro², Renata Galante¹

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

²Centro de Computação e Tecnologia da Informação Universidade de Caxias do Sul
(UCS) – Caxias do Sul – RS – Brazil

{mbigolin, galante}@inf.ufrgs.br, hgrib@ucs.br

Nível: mestrado

Programa: Programa de Pós-graduação em Computação - PPGC/UFRGS

Ano de ingresso: março/2012

Época esperada para conclusão: março/2014

Etapas já concluídas: proposta de mestrado (OUT/2012), créditos (DEZ/2012), seminário de andamento (MAI/2013), submissão de artigo (JUN2013).

Etapas futuras: submissão de artigos (NOV/2013), defesa da dissertação (MAR/2014).

Resumo. *Com a evolução dos Sistemas de Informação Geográficos (SIG) e de mapas na web, a quantidade de informação e a sua complexidade cresceram de forma descontrolada. Essa grande quantia de dados em um mapa pode tornar as interpretações confusas. O objetivo deste trabalho é, através de um modelo multi-dimensional, organizar e estruturar os dados para a criação de agrupamentos, visando uma melhor interpretação. O modelo proposto utiliza dimensões clássicas que podem ser replicadas a diversos problemas de espacialização de dados. Com essa estrutura, pretende-se que o usuário, sem muito conhecimento em processamento de imagem, possa gerar mapas e consultas interativas, com um maior poder de decisão, sem a necessidade de um grande processamento computacional.*

Palavras-chave: Modelagem Multi-dimensional, Agrupamento de dados, Sistemas de Informação Geográfico

Abstract: *The evolution of Geographic Information Systems (GIS) and maps on the web, the amount of information and its complexity increased uncontrollably. This large amount of data on a map can make confusing interpretations. The objective of this work is, through a multi-dimensional model, organize and structure the data for clustering, seeking a better interpretation. The proposed model uses classical dimensions that can be replicated to several problems of spatial data. With this structure, it is intended that the user without much knowledge in image processing, can generate maps and interactive queries with greater decision-making power, without requiring a large computational processing.*

Keywords: *Modeling Multi-Dimensional, Clustering Data, Geographic Information Systems*

1. Introdução

Uma das características dos Sistemas de Informação Geográficos (SIG) é a visualização de planos de informação, ou seja, a sobreposição de camadas de informações distintas, podendo assim, em tempo real, produzir mapas diferentes. Um exemplo de plano de informação muito comum é o de pontos. Com ele, é possível identificar a localização de eventos ou outros objetos geográficos (pontos telefônicos, casas, pontos comerciais, hidrantes, entre outros). Normalmente, o objetivo deste plano de informação é mostrar a distribuição dos dados no mapa [Tyner 2010]. Em muitos casos o número/quantidade de pontos é relevante para a compreensão da informação, bem como a análise da frequência dos objetos e/ou eventos.

Um dos grandes desafios que a comunidade de Informação Geográfica enfrenta é fornecer aos tomadores de decisão ferramentas avançadas, que sejam capazes (semântica e visualmente) de integrar aspectos quantitativos, qualitativos e cognitivos de um domínio de interesse. Quando uma grande quantidade de dados está disponível, a síntese de informação, bem como um resultado derivado pode resultar em uma atividade demorada e de custo elevado [De Chiara et al. 2011]. A análise Geovisual trabalha com problemas que envolvem o espaço geográfico e vários objetos, eventos, fenômenos e processos. As técnicas visuais de análise são essenciais para lidar com os conjuntos de dados atuais (que ampliam-se rapidamente em tamanho e complexidade). As abordagens que trabalham em um ambiente puramente analítico ou em um nível puramente visual não obtém sucesso, devido à dinâmica e complexidade dos processos subjacentes [Andrienko et al. 2011].

1.1. Problema de pesquisa

O grande volume de dados que ao ser plotado em um mapa pode gerar visualizações confusas e de difícil interpretação. Esse problema somado as diversas possibilidades de navegação que um mapa em ambiente *Web* proporcionam amplificam o problema. A Figura 1 ilustra o problema no qual pode-se ver o mapa com pontos de produtores de suínos dos municípios de abrangência do COREDE-SERRA no Rio Grande do Sul [FEE 2011].

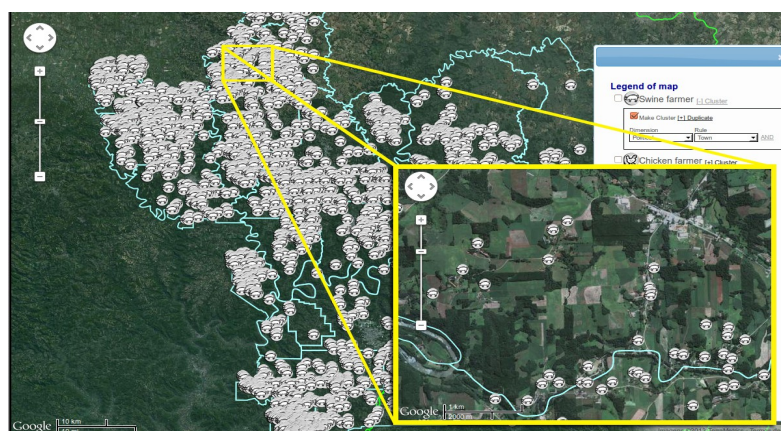


Figura 1. Produtores de suínose destaque com zoom aproximado

A parte em destaque na Figura 1 mostra um *zoom* de determinada área, onde apresenta-se uma grande quantidade de pontos. Isto demonstra que, dependendo do nível de detalhamento ou da região de abrangência analisada, torna-se quase impossível

visualizar e acessar a dispersão dos pontos, sendo, neste caso, necessário utilizar técnicas de agrupamento ou de frequência para melhorar o entendimento. Para obter uma navegação mais eficiente, com possibilidades de tomada de decisão em diferentes níveis de *zoom* e/ou atomicidade da informação, de forma a reduzir a carga cognitiva e proporcionar maior agilidade na interpretação de informações, está sendo proposto o desenvolvimento de um método suportado por uma ferramenta para atender essas necessidades.

1.2. Principais contribuições do trabalho

O trabalho visa propor um método de agrupamento que utiliza um modelo multi-dimensional, semântico, genérico e bem definido para a escolha dos critérios dos grupos de dados para melhorar a visualização de grande densidade de informação em um mapa na *web*. O modelo multi-dimensional facilitará ao usuário a escolha dos grupos e refinamento dos mesmos. Cabe ressaltar que o modelo estará implícito para o usuário, isto é, ele não precisará ter conhecimento em modelagem multi-dimensional. Paralelamente, o modelo também permitirá o estudo para melhorar o desempenho da aplicação. Experimentos preliminares mostram que a transferência de dados e o desenho de informação geográfica em mapas *web* foi mais rápido.

O modelo proposto poderá ser considerado genérico se mostrar-se capaz de ter aplicabilidade em outros domínios e aplicações. O modelo deve ser bem definido de modo que, com a aplicação em áreas distintas, não sejam feitas alterações substanciais. Experimentos envolverão aplicações reais da área de zoologia, engenharia civil e agricultura. Por fim, o modelo deve ser semântico por dar sentido aos agrupamentos gerados.

As agregações e a desagregação de múltiplos planos de informação em um mesmo *web* mapa, necessita de outras variáveis visuais possibilitando-a apresentar diversos tipos de informação ao mesmo tempo no mesmo mapa. A ferramenta visa suportar isso com o uso do modelo multi-dimensional, assim como com a geração de artifícios visuais para os ícones facilitando a interpretação.

2. Principais trabalhos relacionados

Existem diversas técnicas para o agrupamento de informações em mapas: símbolos ou círculos graduados, na qual os valores quantitativos são agrupados em classes, dentro de uma classe, todos os recursos são desenhados com o mesmo símbolo [Auer et al. 2011; MacEachren et al. 2011]; mapas coropléticos, os quais consistem em uma forma de representar grande quantidade de dados usando a abstração de cores [Goovaerts 2012]; *Dot Density Map* é uma solução simples, que utiliza o seguinte conceito: um (1) ponto representa um conjunto de informações [Young and Jensen 2012]. Em *web* mapas é comum a utilização de bibliotecas de grupo (*cluster*). Estas bibliotecas a partir de um considerado número de pontos em um quadrante no mapa geram um grupo. Esses grupos são gerados novamente conforme o usuário navega pelo mapa de forma interativa.

Para análise de informações geográficas usando técnicas OLAP pode se citar o *Spatial OLAP (SOLAP)* Bedard et al. (2007). SOLAP explora os dados através de uma navegação multi-dimensional em qualquer forma de visualização de dados, métricas calculadas e filtragem pelos atributos das dimensões possuindo estrutura dos dados que suporte as diferentes formas geométricas e múltiplas representações para diferentes

escalas. Silva, R. e Santos (2011) propõe o agrupamento de dados para densidade espacial utilizando o algoritmo *DBScan*. Gerando uma nova representação para cada agrupamento, diminuindo o número de objetos espaciais que precisam ser colocados no mapa. Se houver visualização de dados não espaciais, os mesmos também são agregados mantendo a sincronização entre o mapa e uma visualização tabular [Bedard et al. 2007; Silva and Santos 2011]. A ferramenta proposta utiliza um modelo multi-dimensional voltado para consulta e genérico portanto independente de algoritmos para criação de grupos.

Os requisitos elencados, tendo como base a necessidade verificada no protótipo (Figura 1), assim como de trabalhos como o de [Auer et al. 2011] para gerar um mapa na *web* com uma grande densidade de informação são: mostrar densidade aparente da informação (em uma primeira visão obter onde possui mais ou menos informação); densidade exata da informação (apresentar, quantos eventos estão agrupados); permitir visualizar granularidades diferentes (ver grupos conforme a escala do mapa); e manter a mesma abstração em escala diferentes (para diminuir a carga cognitiva, ou seja, manter a mesma forma de analisar o mapa independentemente do *zoom*, por exemplo, um círculo graduado que em um determinado nível de *zoom* represente uma quantidade e em outro nível um círculo de mesmo diâmetro represente quantidade diferente). Esses quatro requisitos elencados foram tabulados comparativamente na Tabela 1.

Tabela 1. Tabela resumo dos recursos disponíveis em cada técnica.

Requisitos	Círculos graduados	Mapas coropléticos	Dot Density Maps	Agrupamento
Mostrar densidade aparente de informação	Sim	Sim	Sim	Sim
Densidade exata da informação	Não	Não	Sim	Sim
Permitir visualizar atomicidades diferentes	Sim/Médio	Sim/Difícil	Sim/Fácil	Sim/Fácil
Manter mesma abstração em nível de zoom diferente	Não	Não	Pode	Sim

3. Andamento da proposta

Esta seção descreve a proposta de mestrado que é gerenciar o agrupamento de pontos em mapas web com a utilização de um modelo multi-dimensional. Uma visão geral da aplicação é ilustrada na Figura 2.

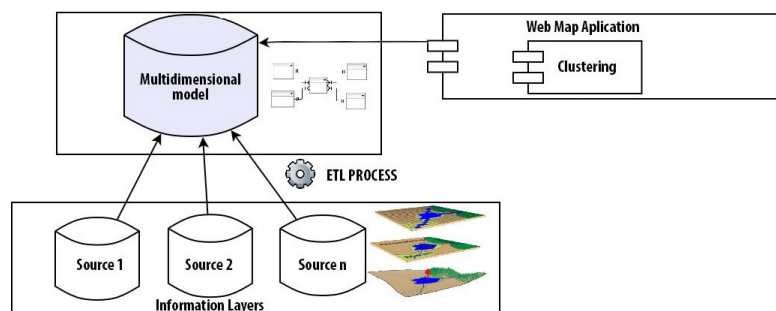


Figura 2. Visão geral do sistema

3.1. Modelo Multi-Dimensional

O modelo proposto trabalha com 3 dimensões bases a fim de otimizar a geração dos grupos: a dimensão “*dim_politica*” (auxilia na organização dos grupos por divisões políticas, como município, estado e países), a dimensão “*dim_fisica*” (auxilia na organização por meio de características físicas comuns, como bacia hidrográfica, região hidrográfica, formação vegetal entre outras) e por ultimo a dimensão “*dim_tempo*”

(essa dimensão armazena características temporais referente ao ponto, como estado ou validade da informação). A Figura 3 mostra em mais detalhes a relação destas dimensões com o fato Ponto. A dimensão física está subdividida em duas outras dimensões a fim de permitir a implementação das características de forma a manter o modelo em formato estrela [Kimball and Ross 2002] e simplificar as consultas. Outras dimensões físicas como tipo de solo e características climáticas podem ser adicionadas, conforme for verificado a necessidade.

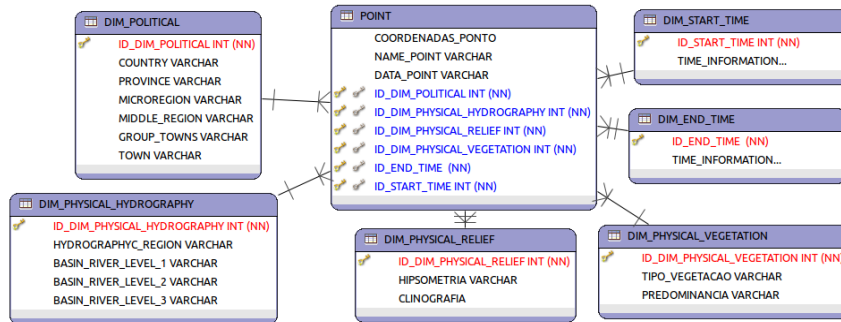


Figura 3 . Modelo Multi-dimensional

3.2. Ferramenta de mapa web

O desenvolvimento da ferramenta que servirá para validar o trabalho utiliza a API do GoogleMaps v 3.0. Os dados são armazenados em um SGBD PostgreSQL com o módulo Postgis, que fornece diversas funcionalidades geográficas além de possibilitar uma arquitetura integrada dos dados geográficos com os dados alfanuméricos do banco de dados [Ferreira et al. 2005].

Para a geração dos grupos, os dados geográficos associados são de fundamental importância para o processo de ETL (*Extraction Transform and Load*). Planos de informação clássicos são utilizados onde são feitas as consultas para verificar se o ponto pertence ou não a uma determinada região geográfica.

A Figura 4 (A) apresenta a ferramenta que permite realizar as consultas tendo como base o modelo multi-dimensional para construir os agrupamentos. As funcionalidades propostas e implementadas são: criar o agrupamento (permitir habilitar ou não o agrupamento sobre os planos de informação); duplicar o plano de informação (permitir que seja visualizado o mesmo plano de informação por agrupamentos diferentes); e criar restrições do tipo AND (criar grupos por mais de uma dimensão).

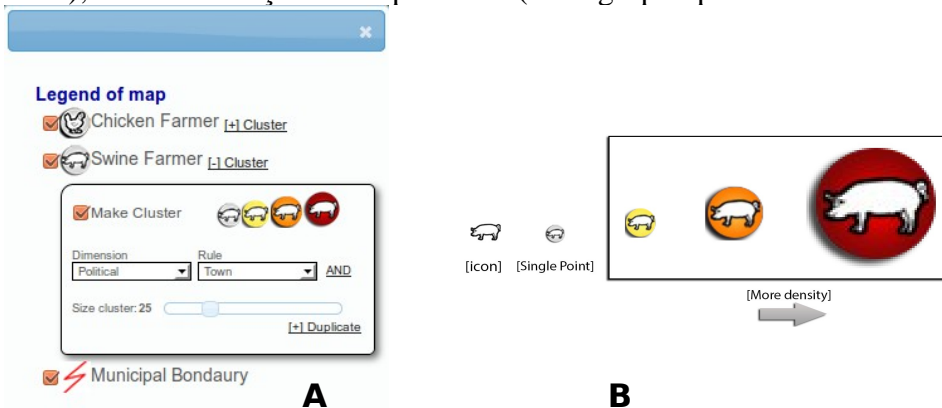


Figura 4. Ferramenta para a criação dos grupos

A solução para o plano de informação da Figura 1, é apresentada na Figura 5 (A) usando apenas a técnica de agrupamento. Neste caso, os grupos levaram apenas em consideração a distância entre os pontos já na Figura 5 (B) visualiza-se o grupo agregado por município.

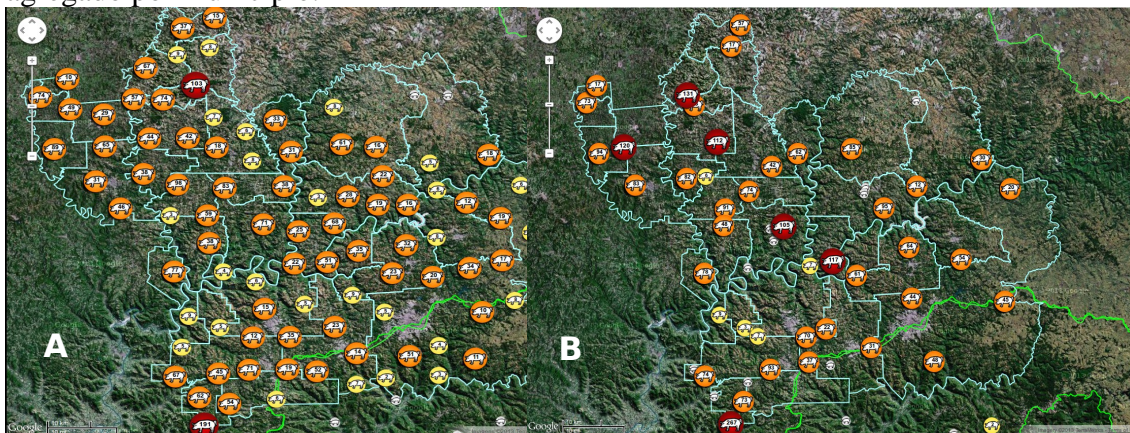


Figura 5. Mosaico de funcionalidades

4. Próximas Atividades

As próximas atividades incluem desenvolver uma amostragem com dois grupos distintos, para comprovar que o modelo utilizado é genérico. Um dos grupos será o GRID (Grupo de Gestão de Riscos e Desastres) da Universidade Federal do Rio Grande do Sul (UFRGS) e outro do ISAM (Instituto de Saneamento Ambiental) da UCS (Universidade de Caxias do Sul). Serão desenvolvidas pequenas aplicações com os dados provenientes destes grupos. Para o ISAM serão utilizados os dados de produtores rurais de região de abrangência do COREDE-SERRA. Para o GRID será utilizado os dados de desastres que ocorreram no estado do Rio Grande do Sul.

Para os dois estudos de caso, os experimentos incluem a análise de usabilidade com usuários reais das aplicações. Além disso, serão recolhidos depoimentos com vistas a analisar de forma qualitativa e quantitativa a real aplicação da proposta de utilização de uma ferramenta de web-mapa com a utilização de um modelo multi-dimensional como base para a tomada de decisão e geração de *cluster* em um mapa.

5. Considerações Finais

Este artigo apresentou um modelo de gerar grupos de pontos de forma automatizada e por dimensões. Por se tratar de informação geográfica, qualquer informação pontual, ou seja, qualquer domínio que utilizar um ponto georreferenciado poderá usar este método. O modelo multi-dimensional proposto apresenta-se como um modelo genérico para melhorar a visualização de um grande número de informação, independente da escala do mapa. É um modelo mais semântico, pois o mesmo visa dar ao usuário do mapa uma melhor visualização da informação controlada pelo contexto do usuário e suas escolhas, ou seja, permite visualização de dados mais direcionada à necessidade de exibição de cada aplicação e/ou especialista. Cabe destacar a necessidade da aplicações extras conforme o proposto visto a validar a aplicabilidade em outras áreas desse mesmo trabalho.

Referencias

Andrienko, G., Andrienko, N., Keim, D., MacEachren, A. M. and Wrobel, S. (aug

- 2011). Challenging problems of geospatial visual analytics. *Journal of Visual Languages & Computing*, v. 22, n. 4, p. 251–256.
- Auer, T., MacEachren, A. M., McCabe, C., Pezanowski, S. and Stryker, M. (mar 2011). HerbariaViz: A web-based client–server interface for mapping and exploring flora observation data. *Ecological Informatics*, v. 6, n. 2, p. 93–110.
- Bedard, Y., Rivest, S. and Proulx, M.-J. (2007). Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures, and Solutions from a Geomatics Engineering Perspective. *Data Warehouses and OLAP: Concepts, Architectures and Solutions*. p. 22.
- De Chiara, D., Del Fatto, V., Laurini, R., Sebillio, M. and Vitiello, G. (jun 2011). A chorem-based approach for visually analyzing spatial data. *Journal of Visual Languages & Computing*, v. 22, n. 3, p. 173–193.
- FEE (2011). Corede Serra.
http://www.fee.tche.br/sitefee/pt/content/resumo/coredes_detalhe.php?corede=Serra, [accessed on Jun 22].
- Ferreira, K. R., Casanova, M. A., Queiroz, G. R. De and Oliveira, O. F. De (2005). Arquiteturas e linguagens. *Bancos de Dados Geográficos*. Curitiba: MundoGEO. p. 33.
- Goovaerts, P. (2012). Geostatistical analysis of health data with different levels of spatial aggregation. *Spatial and Spatio-temporal Epidemiology*,
- Kimball, R. and Ross, M. (2002). *The data warehouse toolkit: guia completo para modelagem dimensional*. Rio de Janeiro: Campus. p. 494
- MacEachren, A. M., Jaiswal, A., Robinson, A. C., et al. (oct 2011). SensePlace2: GeoTwitter analytics support for situational awareness. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*. . IEEE.
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6102456>, [accessed on Apr 18].
- Rivest, S. and Marchand, P. (2001). Toward better support for spatial decision making: defining the characteristics of spatial on-line analytical processing (SOLAP). *GEOMATICA*, v. 55, n. 4, p. 539 to 555.
- Silva, R. and Santos, M. Y. (2011). Spatial Clustering to Uncluttering Map Visualization in SOLAP. In *ICCSA'11 Proceedings of the 2011 international conference on Computational science and its applications*.
- Tyner, J. A. (2010). *Principles of Map Design*. 1. ed. New York, New York, USA: The Guilford Press. p. 259
- Young, S. G. and Jensen, R. R. (may 2012). Statistical and visual analysis of human West Nile virus infection in the United States, 1999–2008. *Applied Geography*, v. 34, p. 425–431.