

Um Framework para Coleta e Filtragem de Dados Geográficos Fornecidos Voluntariamente

João Carlos Tavares da Silva¹, Clodoveu Augusto Davis Jr.²

¹Instituto de Informática – Pontifícia Universidade Católica de Minas Gerais

²Departamento de Ciência da Computação – Universidade Federal de Minas Gerais

jsilva27@gmail.com, clodoveu@dcc.ufmg.br

Resumo. *Diversas aplicações na Web permitem que usuários contribuam com dados textuais, fotos, opiniões e recomendações. O mesmo tem ocorrido em sites ligados a informação geográfica. A contribuição voluntária gera freqüentes dúvidas quanto à sua confiabilidade e qualidade. No entanto, é grande a importância potencial desse tipo de contribuição para a expansão e o melhoramento das fontes de dados geográficos atualmente disponíveis. Este trabalho apresenta uma iniciativa de criação de um framework para coleta de contribuições voluntárias sobre informação geográfica na Web, que será usado para pesquisa sobre aspectos como qualidade, confiabilidade, mecanismos de filtragem e validação, e integração a infra-estruturas de dados espaciais.*

1. Introdução

O compartilhamento de grandes volumes de informação espacial vem se tornando cada vez mais importante. Sistemas de Informação Geográficos (SIG) estão se tornando núcleos de ambientes computacionais que envolvem grandes quantidades de usuários distribuídos em rede [7]. Várias ferramentas passaram a permitir o acesso a grandes volumes de dados no formato de mapas e imagens de satélites, viabilizando o uso de informação geográfica em problemas do cotidiano. Essas iniciativas contribuíram também para desmistificar antigos conceitos a respeito da viabilidade do acesso a SIG através da Internet.

Uma consequência da grande oferta de dados espaciais básicos online foi o surgimento de aplicações que permitem que uma pessoa acrescente dados sobre o espaço em que vive [2]. Isto resulta, entre outras coisas, na possibilidade de localizar informações de interesse pessoal e anotar melhoramentos e correções que podem ser úteis em futuras revisões de mapas. A proliferação destes aplicativos, baseados no conceito da Web 2.0, abre as portas para a inovação. Milhões de usuários não são mais apenas vistos como consumidores, mas também são colaboradores e criadores de informação [4].

Um problema que surge a partir dessa perspectiva é filtrar as contribuições legítimas, separando-as de conteúdo falso ou mal-intencionado. Existem algumas iniciativas a esse respeito em aplicativos típicos da Web 2.0, mas ainda não há mecanismos consagrados voltados para informação geográfica. Em vista disso, este trabalho propõe o desenvolvimento de um *framework* para coleta de contribuições voluntárias a dados geográficos na Web. Como forma de avaliar a confiabilidade das contribuições, o aplicativo prevê a incorporação, para experimentação e testes, de diferentes mecanismos de filtragem do conteúdo,

voltados para estabelecer a confiabilidade dos dados e de seus fornecedores. O objetivo é determinar quais são os mecanismos de controle que se mostram mais eficazes para essa finalidade. Além disso, como se trata de um *framework*, o aplicativo desenvolvido permite que Web sites voltados para contribuição voluntária sejam criados e publicados por pessoas ou organizações interessadas.

Este artigo está organizado da seguinte forma. A seção 2 descreve iniciativas atuais em torno da contribuição de usuários para a formação de conteúdo na Web. A seção 3 apresenta o framework proposto. Na seção 4 são feitas considerações finais e apresentado o que se planeja para a condução das próximas etapas do trabalho.

2. Contribuições voluntárias para dados geográficos

Tem crescido rapidamente o interesse das pessoas por recursos e ferramentas que permitem que conteúdo produzido por indivíduos e grupos seja publicado na Internet, obtendo imediatamente alcance mundial. São inúmeras as iniciativas de criação de blogs, publicação de vídeos domésticos, registro de comentários e opiniões sobre produtos, serviços, livros e filmes, entre muitos outros. O crescimento explosivo de sites que permitem algum tipo de participação dinâmica dos usuários é indicação clara dessa tendência. Como exemplo, pode-se citar sites como YouTube, Orkut, Flickr, del.icio.us, Blogger e muitos outros.

Um exemplo fundamental das possibilidades da criação de conteúdo por parte de usuários é a Wikipedia. Os mecanismos de criação e manutenção da Wikipedia a colocam no centro de uma grande polêmica, envolvendo questões como qualidade, confiabilidade, completeza e atualidade dos artigos, em comparação com fontes tradicionais de informação. Mas é inegável que o modelo *wiki* de participação voluntária, irrestrita e, na maioria das vezes, anônima, pode levar a resultados que estão se tornando impraticáveis para organizações tradicionais. Basta tentar imaginar o custo e o tamanho do esforço necessário para criar algo como a Wikipedia sob a égide de uma empresa comercial, que teria que pagar pelas contribuições e cobrar pelo acesso ao resultado.

A analogia da Wikipedia com uma enciclopédia convencional, como a Encyclopaedia Britannica, é inevitável. Embora a Britannica tenha desenvolvido uma versão online, o conteúdo integral é reservado para assinantes; embora disponha de uma interface para que leitores cadastrados contribuam com sugestões, não é possível saber se a sugestão foi ou será aceita, e se alguma modificação ou atualização resultou da contribuição do leitor. Enquanto a Britannica é autoritativa, em muitos artigos da Wikipedia há um aviso informando o leitor de que há polêmica sobre o conteúdo apresentado. Mesmo assim, existe ainda muita discussão sobre a confiabilidade de partes do conteúdo da Wikipedia [15].

Os mecanismos que levam as pessoas a dispender tempo e esforço para produzir de graça algo que pode ou não ser usado por outros ainda não são claros [4, 9]. No entanto, a regra básica dos mecanismos *wiki* (e originária do desenvolvimento colaborativo de software livre) parece ser válida para uma ampla gama de situações: “se houver olhos suficientes, todos os *bugs* são visíveis” [14]. Ou seja, se um conjunto de informações disponível livremente na Web for de interesse para um grande número de pessoas, uma parcela razoável dessas pessoas tende a retribuir o serviço prestado por meio de uma predisposição a colaborar para que o serviço melhore. Tal fenômeno tem sido estudado na área de Comunicação,

sendo conhecido como “ação coletiva” [3, 8], e voltado para acervos públicos de informação, chamados de *information commons* [13].

Com relação à informação geográfica, um fenômeno semelhante vem ocorrendo recentemente. Tendo por base acervos de mapas e imagens georreferenciadas disponíveis gratuitamente pela Web (tais como Google Maps, Microsoft Windows Live Local, Yahoo Maps), vários *Web sites* têm sido criados contando com a perspectiva da colaboração do usuário na criação e manutenção de um acervo temático ou complementar de dados geográficos. Por meio desses mecanismos, um usuário qualquer pode tentar reconhecer aspectos da realidade que o cerca e criar uma anotação, comentário, dar um nome, ou conectá-lo a alguma outra fonte de informação complementar. Um exemplo é o site Wikimapia¹, em que usuários delimitam áreas de seu interesse com retângulos ou polígonos e associam a eles descrições e comentários. Muitas vezes, o local indicado recebe também um link para um artigo na Wikipedia, usando o mecanismo de geotags. Outro exemplo é o site OpenStreet-Map², em que usuários podem editar mapas de ruas a partir de imagens de satélite de alta resolução, dentre as quais as do Yahoo! Maps. Os mapas resultantes podem ser exportados e usados em outras ferramentas de geoprocessamento, estando livres de licenciamento. No Brasil, uma experiência semelhante é o projeto Tracksource³, iniciado por jipeiros interessados em trocar dados geográficos sobre estradas vicinais e trilhas, coletados por receptores GPS de navegação [11]. O projeto inclui muitos mapas urbanos roteáveis, contendo informação sobre regras de trânsito para planejamento de deslocamentos. Os mapas disponíveis no Tracksource podem ser usados em computadores pessoais ou carregados em receptores GPS para uso móvel em navegação. O Tracksource realiza hoje mais de 50.000 downloads de mapas por mês, totalizando cerca de 400 GB/mês. O site frequentemente ultrapassa 1.500.000 “hits” mensais, demonstrando o interesse que existe atualmente nesse tipo de informação – bem como as deficiências dos mapas disponíveis para receptores GPS.

Existem precedentes para a contribuição voluntária de dados, inclusive em projetos de pesquisa científica. É conhecida nos Estados Unidos uma iniciativa anual de observação e monitoramento de aves migratórias, em que a posição de bandos de aves em determinado instante é registrada e comunicada aos cientistas aficionados do *hobby* de observação de pássaros. Um exemplo conhecido de contribuição voluntária para mapeamento é o projeto *National Map Corps*, do *United States Geological Survey* (USGS), a agência de mapeamento nacional norte-americana [1, 5]. O USGS iniciou esse projeto na década de 1990, chegando a reunir mais de 3.000 voluntários em 2001. Recentemente, o projeto passou a receber apenas contribuições via Web⁴ e que fossem adequadamente acompanhadas de coordenadas obtidas por GPS, mas mesmo assim existe um *backlog* de contribuições que precisam ser tratadas. Na recepção de contribuições via Web, o projeto foi inspirado em uma iniciativa da NASA, denominada criativamente *Clickworkers*⁵.

¹ <http://www.wikimapia.com>

² <http://www.openstreetmap.org>

³ <http://www.tracksource.org.br>

⁴ <http://ims.er.usgs.gov/vfs/faces/index.jspx>

⁵ <http://clickworkers.arc.nasa.gov/top>

Em resumo, pode-se observar que existem muitas oportunidades para uma avaliação mais científica das iniciativas de participação voluntária na criação e manutenção de acervos de informação geográfica, visando prover melhores serviços no futuro.

3. Framework para coleta e filtragem de contribuições voluntárias

O propósito desta pesquisa é projetar e implementar um *framework* para coleta e dados geográficos fornecidos voluntariamente, através da Web. Por meio do *framework*, uma pessoa ou organização publica na Web um conjunto de dados geográficos, e abre a outros usuários a possibilidade de contribuir com novos dados, propostas de correção ou de atualização. Existe a opção de permitir apenas a carga de novos dados em feições previamente modeladas, ou permitir a criação de temas novos. A Figura 1 apresenta esquematicamente o framework proposto e desenvolvido até o momento.

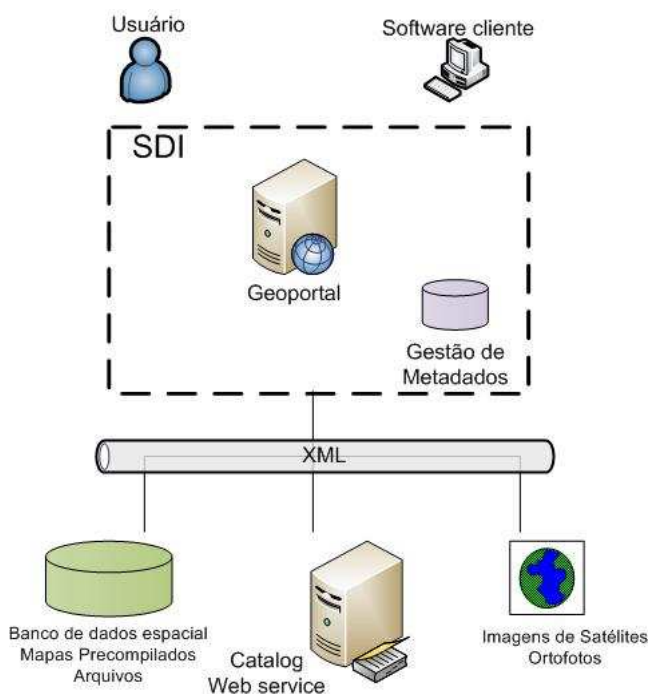


Figura 1 – Diagrama esquemático do framework desenvolvido

A solução, implementada sob um Geoportál, permite ao usuário a customização de sua interface, tornando-a aderente às particularidades e demandas de seu interesse, através de funções que possibilitam a criação de feições personalizadas, a carga de dados a partir de arquivos georreferenciados e a tematização dos dados (Figura 2). A criação de feições personalizadas torna a solução ajustável às necessidades do usuário, permitindo a inclusão de dados cujos atributos e propriedades são definidos pelos próprios usuários. É possível determinar o nome da feição, seu tipo de geometria, o número de atributos, o tipo e o valor padrão associado a cada atributo, as informações principais para exibição no mapa, a escala de visualização e a simbologia relacionada a cada faixa de escalas. Para cada feição personalizada, o sistema cria uma camada que permite a exibição dos dados correspondentes no mapa. Após a criação de uma feição, os objetos espaciais relacionados a ela podem ser adicionados manualmente ou por carga de dados a partir de um arquivo.

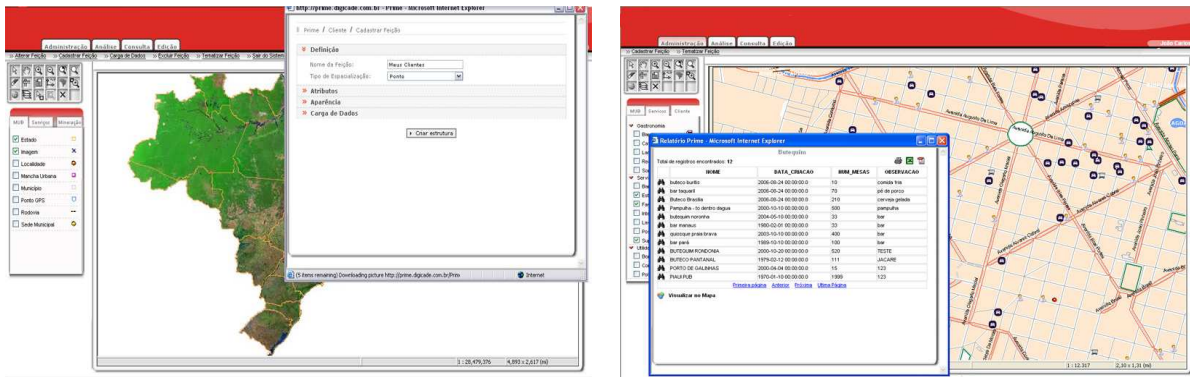


Figura 2 – Interfaces da solução implementada: telas para cadastro e inserção de dados

O *framework* utiliza um gerenciador de bancos de dados geográficos para armazenar o esquema criado pelo usuário para suas feições. No mesmo gerenciador são armazenados os dados de fundo (mapas básicos) e os dados fornecidos pelo usuário. A Figura 3 exibe um fluxograma de funcionamento da aplicação. Criada a estrutura, pode ser liberado o acesso para que outras pessoas possam contribuir, ou ainda continuar o cadastro de novas feições de interesse. O objetivo é permitir o acesso por meio de um visualizador, instalado como componente de um geoportal, e também publicar os dados fornecidos pelo usuário utilizando serviços Web padrão OGC. Com isso, os dados fornecidos pelos voluntários podem se tornar parte de uma infra-estrutura de dados espaciais (IDE) [6].

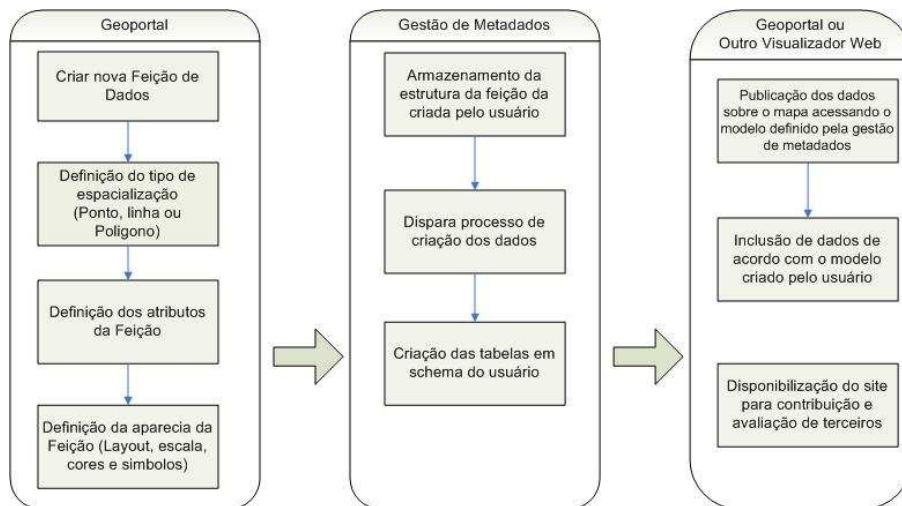


Figura 3 - Fluxograma de funcionamento da solução

4. Considerações finais e trabalhos futuros

Este projeto está em andamento. No estágio atual, usuários serão apresentados ao *framework*, e mecanismos de filtragem estão sendo implementados. A idéia é criar comunidades de usuários que contribuam com dados geográficos sobre um tema de interesse comum, em um ambiente controlado. Será também desenvolvida uma adaptação para que o usuário possa visualizar seus dados fora do ambiente do Geoportal, usando serviços Web OGC. O *framework* incluirá ainda um conjunto de recursos para que o criador de uma camada de dados ou feição possa solicitar, aos próprios voluntários, uma avaliação da qualidade dos

dados fornecidos [8, 10]. Diversos mecanismos como esses foram propostos e estão disponíveis para validação de contribuições voluntárias, geográficas ou não, em Web sites do tipo *wiki*. Esses mecanismos serão implementados e integrados ao *framework*. A idéia é permitir que o criador de um *site* ou feição possa especificar o tipo de mecanismo que considera mais interessante, levando em conta a natureza dos dados e da comunidade envolvida no processo. Esse recurso abre diversas possibilidades interessantes para pesquisa, na avaliação comparativa entre a qualidade percebida dos dados fornecidos e a qualidade efetiva, determinada por especialistas. Também permite comparar mecanismos de filtragem de contribuições voluntárias, de modo a apontar os mais eficientes, e pode levar à implementação de sistemas de reputação e recomendação para dados geográficos na Web [12].

5. Referências

1. Bearden, M.J. *The National Map Corps*. in *Workshop on Volunteered Geographic Information*. 2007. Santa Barbara, California, USA.
2. Bell, T., Cope, A.S., and Catt, D. *The third spatial revolution*. in *Workshop on Volunteered Geographic Information*. 2007. Santa Barbara, California, USA.
3. Bimber, B., Flanagan, A.J., and Stohl, C., *Reconceptualizing collective action in the contemporary media environment*. *Communication Theory*, 2005. **15**(4): p. 365-388.
4. Budhathoki, N.R. *Reconceptualization of user is essential to expand the voluntary creation and supply of spatial information*. in *Workshop on Volunteered Geographic Information*. 2007. Santa Barbara, California, USA.
5. Cowen, D.J. *Why not a geo-wiki corps?* in *Workshop on Volunteered Geographic Information*. 2007. Santa Barbara, California, USA.
6. Davis Jr, C.A. and Alves, L.L. *Local Spatial Data Infrastructures Based on a Service-Oriented Architecture*. in *8th Brazilian Symposium on GeoInformatics (GeoInfo 2005)*. 2005. Campos do Jordão (SP).
7. Davis Jr., C.A. and Fonseca, F.T., *Considerations from the Development of a Local Spatial Data Infrastructure*. *Information Technology for Development*, 2005. **12**(4): p. 273-290.
8. Frew, J. *Provenance and volunteered geographic information*. in *Workshop on Volunteered Geographic Information*. 2007. Santa Barbara, California, USA.
9. Goodchild, M.F., *Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0*. *International Journal of Spatial Data Infrastructures Research*, 2007. **2**: p. 24-32.
10. Gouveia, C. *How to make sense of citizen participation in environmental monitoring?* in *Workshop on Volunteered Geographic Information*. 2007. Santa Barbara, California, USA.
11. Lima, M.F.M., *O Projeto Tracksource*, in *Instituto de Educação Continuada*. 2008, Pontificia Universidade Católica de Minas Gerais.
12. Maué, P. *Reputation as a tool to ensure validity of VGI*. in *Workshop on Volunteered Geographic Information*. 2007. Santa Barbara, California, USA.
13. Onsrud, H., Câmara, G., Campbell, J., and Chakravarthy, N.S., *Public Commons of Geographic Data: research and development challenges*, in *GIScience 2004 (LNCS 3234)*, Egenhofer, M.J., Freksa, C., and Miller, H.J., Editors. 2004, Springer-Verlag: Berlin/Heidelberg. p. 223-238.
14. Raymond, E. *The Cathedral and the Bazaar*. 1999 [cited; Available from: Available at <http://www.tuxedo.org/~esr/writings/cathedral-bazaar/>].
15. Waters, N.L., *Why you can't cite Wikipedia in my class*. *Communications of the ACM*, 2007. **50**(9): p. 15-17.