

Um Modelo Espacial para Integração dos Dados da Saúde Implementado em Software Livre

Vania Bogorny¹, Felipe Nievinski², Nara Bigolin³

¹ Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS) - Campus do Vale, Av. Bento Gonçalves, 9500, Bloco 4, Bairro Agronomia, Porto Alegre, RS, CEP: 91.501-970, Brasil

² Instituto de Geociências - Universidade Federal do Rio Grande do Sul (UFRGS) - Campus do Vale, Av. Bento Gonçalves, 9500, Bloco 4, Bairro Agronomia, Porto Alegre, RS, CEP: 91.501-970, Brasil

³ Faculdade de Informática – (ULBRA) - Rua Miguel Tostes, 101 – Canoas – RS -Brasil
vania@inf.ufrgs.br, fnievinski@terra.com.br, nara@ulbra.tche.br

Resumo: *o software com o código aberto está sendo desenvolvido com agilidade e está ganhando mercado em várias áreas da informática, incluindo sistemas operacionais, bancos de dados convencionais e geográficos, linguagens de programação, sistemas de descoberta de conhecimento, entre outros. Este artigo apresenta um modelo de dados, orientado a objetos, que integra e referencia, geograficamente, as bases de dados heterogêneas da Secretaria Estadual de Saúde do Estado do Rio Grande do Sul (SES/RS), que foi implementado em software livre.*

Abstract: *open source software is being increasingly developed in a number of Information Technology areas, including operating systems, traditional database management systems, spatial DBMS, programming languages, knowledge discovery systems, among many others. This paper presents an object data model that integrates and spatially represents the heterogeneous databases related to public health of Rio Grande do Su, implemented in free software tools.*

1. Introdução

Os Sistemas de Informação Geográfica (SIG) são sistemas especiais que tratam dados convencionais/descritivos e a parte espacial dos dados geográficos, dados estes que têm uma localização com relação a superfície terrestre [Câmara 1996]. Os dados geográficos estão sendo cada vez mais utilizados por empresas privadas e órgãos governamentais para gerenciar questões ambientais, redes de telefonia e energia elétrica, áreas urbanas, epidemias, questões de saúde pública, entre outras. A Coordenadoria de Política de Atenção Integral a Saúde do Trabalhador da SES/RS optou por um SIG para fazer o mapeamento de intoxicações, internações e mortes decorrentes de acidentes e doenças com uma possível causa relacionada ao ambiente de trabalho. Além disso, o sistema permite, também, o mapeamento de acidentes e doenças não relacionadas ao local de trabalho, podendo assim, ser utilizado em diversas áreas de saúde pública.

O modelo de dados do SIG, denominado GEOSIST (Georreferenciamento do Sistema de Informações em Saúde do Trabalhador) integra as seguintes bases de dados: Sistema de Informações de Mortalidade (SIM), Sistema de Informações Hospitalares (SIH), Sistema de Informações de Notificação de Agravos Notificáveis (SINAN) e o Sistema de Informações em Saúde do Trabalhador (SIST). Através de sua interface WEB [Bogorny, Nievinski, Bigolin, Kalil e Corrêa 2002], o usuário pode fazer consultas e análises espaciais cruzando dados de agravo (acidentes e doenças) e dados de risco (possíveis causas que ocasionaram o acidente ou a doença). Os últimos foram coletados de órgãos externos à SES como o IBGE (Instituto Brasileiro de Geografia e Estatística), o Ministério do Trabalho, a Secretaria de Agricultura e Abastecimento (SAA) e o Programa Pró-Guaíba. A análise dos

No diagrama de classes da figura 1, estão representados os *agravos* mais relevantes para a Coordenadoria de Política de Atenção Integral à Saúde do Trabalhador, entre os quais estão as intoxicações causadas por agrotóxicos, as internações hospitalares, as mortes, os acidentes e os casos de violência. O estereótipo da classe expressa o sistema de informação que os originou os objetos.

Os dados de *risco* foram modelados com base em entrevistas realizadas com usuários de órgãos externos à SES/RS. Entre os principais elementos desse modelo estão os acidentes de cargas perigosas, a distribuição agrária, os pontos de coleta de água, as principais fontes poluidoras, os pontos de lançamento de líquido e a medição da poluição atmosférica. O modelo completo de riscos está descrito em [Bogorny e Nievinski 2002a].

A base de dados geográfica é formada por três mapas distintos, contudo, apresentados na mesma escala, permitindo a realização de camadas de sobreposição. Esses mapas são os Limites Políticos Municipais, que incluem os atuais 499 municípios gaúchos (fonte –SAA), o mapa Hidrográfico e o mapa Rodoviário estaduais. O mapa de Limites Políticos Municipais é a base para a espacialização e o mapeamento temático dos dados de atributos.

3. Dificuldades Encontradas na Realização do Projeto

O GEOSIST engloba dados de múltiplas fontes que, por conseqüência, são em grande parte heterogêneos. As diferentes fontes de dados utilizam codificações e padrões próprios, dificultando a sua integração. Em virtude do tempo previsto para o término do projeto, um modelo de dados foi criado para integrar as diferentes bases de dados, porém, apenas os dados de município e doenças foram integrados para fazer as consultas e análises espaciais. Os municípios foram codificados de acordo com os padrões estabelecidos pelo IBGE e as doenças de acordo com o Código Internacional de Doenças (CID) [Bogorny, Nievinski, Bigolin, Kalil e Corrêa. 2002].

Um fator importante que influencia diretamente nas consultas e análises espaciais é a quantidade de dados com ruído, ou seja, dados nulos e dados preenchidos inadequadamente. Outro aspecto relevante é o número de informações coletadas e tratadas de forma manual, ou seja, que não estão disponíveis em formato digital. Outros dados de fundamental importância para o sistema como, por exemplo, a quantidade de agrotóxico fabricado/importado e consumido nos diferentes municípios do estado, não são controlados pela SAA.

Além das dificuldades mencionadas, não existe, ainda, um consenso entre empresas e órgãos governamentais para padronizar, disponibilizar e compartilhar informações. Isso dificulta a total implementação de projetos como o GEOSIST. Em virtude desse tipo de problema, muitos dados, principalmente os de risco, foram mapeados na fase de concepção do sistema, criados no projeto lógico, porém não instanciados no projeto físico.

4. Benefícios/Facilidades propiciadas pelo uso do software livre

O GEOSIST foi totalmente desenvolvido em software livre. Sua arquitetura [Bogorny, Nievinski, Bigolin, Kalil e Corrêa 2002] é formada de uma interface WEB desenvolvida em PHP [PHP 2003], através da qual o usuário acessa a base de dados descritiva e espacial que é gerenciada pelo PostgreSQL [PostgreSQL 2003] e PostGIS [PostGIS 2003]. Os mapas são gerados pelo UMN-MapServer [MapServer 2003].

Os principais benefícios da utilização de software livre para este trabalho foram:

- *existência e disponibilidade dos recursos de software*: foram encontrados programas adequados para a implementação de cada etapa do projeto;
- *otimização de consultas*: o acesso ao código fonte permitiu adaptar os programas de acordo com as necessidades do projeto. Mais especificamente, a conexão do UMN-MapServer com o PostgreSQL foi adaptada para acelerar a execução de consultas complexas sobre os dados espaciais. O acesso ao

código fonte também permitiu a correção de dois *bugs* pequenos no *UMN-MapServer*, cujas correções foram incorporadas à distribuição oficial desse software;

- *documentação*: apesar de não haver uma empresa responsável pelo suporte, como existe para softwares proprietários, os principais programas utilizados disponibilizam uma ótima documentação e oferecem acervos como listas de discussão que facilitam a resolução de diversos problemas.

5. Conclusões e trabalhos futuros

Apesar das dificuldades encontradas na busca pela informação, no mapeamento e na padronização dos dados, o GEOSIST pode auxiliar o especialista em saúde pública a tomar medidas preventivas com relação às mortes, internações hospitalares e acidentes e doenças possivelmente relacionadas ao ambiente de trabalho.

O modelo de dados, apresentado neste artigo foi utilizado para fazer o mapeamento espacial dos agravos em saúde pública. Ele pode servir de base para integrar os diferentes sistemas de informação existentes na SES, desde que os dados como sexo, raça/cor, escolaridade, entre outros, sejam padronizados. O mapeamento desses dados está descrito em [Bogorny e Nievinski 2002a].

Com relação aos trabalhos futuros, inúmeras melhorias podem ser realizadas para que o GEOSIST se torne um SIG completo para o cruzamento espacial de agravos e riscos. Inicialmente é necessário padronizar dados como faixa etária, sexo, raça, escolaridade, ocupação, etc. Um outro aspecto de fundamental importância é a importação de todos os dados de risco, que foram modelados, porém, não foram disponibilizados pelos órgãos responsáveis.

Dados como uso de agrotóxicos, quantidade de água e percentual de esgoto tratado, classe social, entre outros, são informações que, de uma maneira geral, aumentariam a qualidade do trabalho.

Os dados do GEOSIST, principalmente os que se referem às mortes e internações estão sendo utilizadas como fontes de pesquisa na área de Inteligência Artificial, mais especificamente na Descoberta de Conhecimento em Bases de Dados (DCBD) [Lucas, Bigolin e Alvares 2002]. Nesta área, também estão sendo utilizadas ferramentas de software livre como, por exemplo, o Weka [Weka 2003]. Futuramente o GEOSIST deve ser utilizado como base para a Descoberta de Conhecimento em Bases de Dados Geográficos.

6. Referências Bibliográficas

- Bogorny, V. Nievinski, F.; Bigolin, N.; Kalil, F.; Corrêa, M.J.. (2002), Desenvolvimento de um SIG para Saúde Pública usando Software Livre. Workshop de Software Livre.
- Bogorny, V.; Nievinski, F. (2002), Projeto Conceitual, incluindo análise de requisitos e levantamento dos dados e dos softwares livres para o projeto de SIG. Relatório Técnico.
- Bogorny, V.; Nievinski, F. (2002a), Criação Física do Banco de Dados. Relatório Técnico.
- Booch, G.; Rumbaugh, J.; Jacobson, I. (2000), UML: user guide. Rio de Janeiro: Campus.
- Câmara, G.; Medeiros, C.B.; Casanova, M.A.; Hemerly, A.; Magalhães, G. (1996) “Anatomia de Sistemas de Informação Geográfica”, Instituto de Computação, UNICAMP – Campinas.
- Lucas, A. M.; Bigolin, N. M.; Alvares, L. O. Mineração em dados reais da saúde considerando aspectos temporais. Uruguiana-RS, p. 97-104, 2002.
- PHP (2003), <http://www.php.net/>.
- PostgreSQL (2003), <http://www.postgresql.org>.
- PostGIS (2003), <http://postgis.refractor.net>.
- UMN-MapServer (2003), <http://mapserver.gis.umn.edu>.
- Weka 3: Machine Learning Software in Java (2003), <http://www.cs.waikato.ac.nz/ml/weka>.