

MODELAGEM SEMÂNTICA: COMPREENDENDO AS DIFERENÇAS ENTRE SISTEMAS DE GEOPROCESSAMENTO

Uma das grandes barreiras à maior difusão da geo-informação reside na grande dificuldade de conversão de dados entre os diferentes sistemas de informação geográfica (SIG). Qualquer usuário de geoprocessamento que trabalha com projetos de porte médio já se defrontou com o desafio de traduzir e converter dados produzidos em outro sistema. Como enfrentar este problema?

Uma abordagem óbvia e direta para o problema de conversão consiste no uso de formatos intermediários, geralmente sob forma de arquivos textuais, que são importados e exportados pela maior parte dos sistemas. Nesta abordagem, no entanto, os arquivos do banco de dados original são exportados um a um, e ficamos sem responder uma uma pergunta crucial: como estavam organizados estes dados? Qual o relacionamento entre as entidades? Qual o conteúdo do banco de dados original?

Para realizar a conversão plena entre dados de SIGs distintos, é preciso entender a estrutura de cada banco de dados geo-referenciado. Isto não é tarefa fácil, em função da grande diversidade conceitual entre os sistemas de informação geográfica. Estas diferenças não existem porque os usuários e companhias assim o desejam, mas em função da tecnologia ter sido implementada antes de se ter estabelecido uma teoria sólida sobre a natureza da informação espacial. Em outras palavras: partindo de visões conceituais muito distintas, os desenvolvedores de sistemas produziram tecnologias com baixo grau de compatibilidade.

Neste contexto, é instrutivo traçar um paralelo com a área de bancos de dados relacionais. Em 1970, um pesquisador da IBM, E. F. Codd, lançou um trabalho de pesquisa em que propunha as bases do modelo relacional, que prima pela simplicidade, ao representar dados descritivos em tabelas (relações), organizadas em

linhas e colunas. Através do uso de uma poderosa e simples abstração (*a tabela*), este modelo garante a independência entre programas de aplicação e as estruturas de dados utilizadas no banco de dados. Em trabalhos subsequentes, a comunidade de banco de dados estabeleceu uma teoria sólida e rigorosa, que fundamenta o uso do modelo relacional. Deste modo, quando as diversas implementações de bancos de dados relacionais chegaram ao mercado, todas comungavam dos mesmos fundamentos; assim, a conversão de dados entre aplicativos como ORACLE, INFORMIX, ACCESS e SQL SERVER é problema relativamente simples de ser resolvido.

No caso de Geoprocessamento, somente em meados da década de 90 surgiram os primeiros trabalhos de fundamentação teórica destes sistemas, quando uma parcela significativa de usuários já investira na compra de software e na montagem de bases de dados geo-referenciadas. Este cenário perdura até hoje, e é responsável pelas grandes barreiras culturais (e até quase "religiosas") que separam os diferentes instituições. Quantos de nós já ouviram as frases: "Sou especialista no sistema X ?" "Temos uma equipe treinada no sistema Y, e vai demorar vários anos para retreiná-la no software Z ?" "Quantos cursos de "Iniciação às Geotecnologias" são na verdade apenas cursos de treinamento no sistema W ?

A superação deste cenário insular requer uma abordagem completamente diferente de todos os especialistas. Em particular, nossas universidades e grupos de pesquisa tem uma responsabilidade específica: ensinar os fundamentos teóricos gerais, e não apenas o uso de um único sistema.

A chave reside em duas palavras: *modelo semântico*, termo que denota o conjunto de conceitos próprio a cada sistema. Por exemplo, o MGE/INTERGRAPH emprega, entre outros, os conceitos de "categoria", "classe de feição", "feição" e "índice", enquanto o ARC/VIEW utiliza os conceitos de "visão", "tema", "conjunto de formas (*shapefile*)", "tabela", e o SPRING possui termos como "categoria", "geo-classe", "objeto", "numérico", "temático" e "cadastral". Porque estas diferenças ? A razão é que estes sistemas se baseiam em visões diferentes da geo-informação. Se não

compreenderm cada um destes conceitos, dificilmente teremos plena condição de estabelecer o melhor mecanismo possível de conversão de dados entre sistemas.

Em resumo, é fundamental que o profissional da geo-informação, para não ficar escravo de um único fornecedor, compreenda bem o modelo semântico de cada um dos principais SIGs do mercado brasileiro. Como contribuição neste necessário esforço de aprendizado, recentemente orientamos uma dissertação de Mestrado no INPE sobre "Modelos Semânticos de Sistemas de Informação Geográfica e Tradução para o Padrão OPEN GIS". Este trabalho analisa em detalhe o modelo semântico de três SIG disponíveis no Brasil (ARC/INFO, MGE e SPRING), e propõe uma estratégia para conversão de modelos entre estas tecnologias. Este trabalho está disponível para os interessados em <http://www.dpi.inpe.br/teses/thome>.

Gostaríamos de encorajar outros colegas e profissionais a realizar um grande esforço coletivo de compreensão dos modelos semânticos dos diferentes sistemas, condição necessária para superar a situação indesejável de dependência de fabricantes em que a maioria de nossas instituições ainda se encontra.

Gilberto Câmara é pesquisador da Divisão de
Processamento de Imagens do INPE, e foi um dos
coordenadores do desenvolvimento dos sistemas SGI e
SPRING.