

Dicionário de tipos de feições geoespaciais para o território brasileiro

Ivanildo Barbosa^{1,2}, Marco Antonio Casanova²

¹Seção de Ensino de Engenharia Cartográfica – Instituto Militar de Engenharia (IME)
Praça General Tibúrcio, 80 – CEP 22290-270 – Rio de Janeiro – Brasil

²Departamento de Informática – Pontifícia Universidade Católica do Rio de Janeiro
(PUC – Rio)

Rua Marquês de São Vicente, 225 - CEP 22451-900 – Rio de Janeiro – Brasil

{ibarbosa,casanova}@inf.puc-rio.br

Resumo. *Este trabalho apresenta uma análise das especificações vigentes no Brasil para a elaboração de produtos geoespaciais, visando à verificação da completude das mesmas no que diz respeito à diversidade de feições geoespaciais representáveis. A relação de termos de cada uma das especificações é comparada com o dicionário de feições geoespaciais da Alexandria Digital Library (ADL), verificando combinação exata entre termos, sinônimos, ou omissões em qualquer das partes. Foram detectadas diferenças relevantes nas abordagens para classificação das feições, bem como a omissão de termos classificando elementos que não ocorrem em território brasileiro.*

Abstract. *This paper presents an analysis of the current Brazilian specifications for creating geospatial products, aiming at checking their completeness concerning the diversity of representable geospatial features. The list of terms of each specification is compared with the Alexandria Digital Library (ADL) Feature Type Thesaurus, checking for the exact match for terms, for synonyms and for missing terms. The analysis detected important differences on the approaches for feature classification, as well the omission of certain terms in the specifications, mostly corresponding to types of geographic features not occurring in Brazil.*

1. Introdução

Existe uma crescente demanda por dados espaciais e uma consequente demanda pela sua disponibilização. Um dos objetivos da Infraestrutura Nacional de Dados Espaciais (INDE) é o compartilhamento dos dados geoespaciais de origem federal, estadual, distrital e municipal [Brasil 2008].

Os dados geoespaciais abrangem bases em mídia analógica, como as cartas topográficas impressas, ou dados digitais em formato matricial ou vetorial. No caso das cartas do mapeamento sistemático, produzidas pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e pela Diretoria de Serviço Geográfico (DSG) do Exército Brasileiro, as feições são classificadas em categorias de informação conforme especificado no manual T34-700 [DSG 1998], destinado a “*normatizar a representação dos acidentes*

naturais e artificiais em cartas topográficas e similares nas escalas de 1:25.000, 1:50.000, 1:100.000 e 1:250.000". A atribuição de simbologia às feições depende da pertinência dessas feições a uma das categorias de informação pré-definidas, respeitadas as condições de representatividade nas escalas mencionadas, muito embora nem todos os dados geoespaciais de interesse geral não estejam restritos a essas escalas, nem ao conjunto de dados representáveis em mapas topográficos.

Com o processo de informatização da cartografia, a DSG e o IBGE realizaram esforços para especificar a estruturação dos dados espaciais vetoriais, surgindo as versões da Mapoteca Topográfica Digital [IBGE 1996] e da Tabela da Base Cartográfica Digital [DSG 1997]. Contudo, esses padrões eram incompatíveis, demandando mecanismos de conversão entre arquivos produzidos por aqueles órgãos.

Atualmente, está em vigor a segunda versão das Especificações Técnicas para Estruturação de Dados Geoespaciais Digitais Vetoriais (ET-EDGV) [CONCAR 2007]. Nelas estão definidos os esquemas aplicáveis às diferentes categorias de informação disponíveis em formato vetorial. Em cada categoria são descritas as classes e respectivos atributos, seus tipos e domínios. A recuperação das informações contidas nos produtos gerados com base nessas especificações está condicionada à perfeita associação do significado das categorias às representações das feições, permitindo a usuários de diversas especialidades conhecer de maneira inequívoca o conteúdo disponibilizado.

Os termos empregados na caracterização das feições geoespaciais representáveis conforme as especificações apresentadas formam um vocabulário adotado pelas instituições citadas em Brasil [2008], que pode ser usado como base para a construção de um dicionário (*thesaurus*) – *vocabulário de uma linguagem de indexação controlada, estruturada de modo que os relacionamentos entre conceitos são explicitados* [ISO, 1986 *apud* Breitman *et al*, 2007]. Uma vez delimitado o domínio de atuação, são estabelecidos os critérios de categorização dos termos que, neste trabalho, serão baseados naqueles adotados em cada especificação, aplicando alterações pontuais visando à busca por diferenças de conteúdo entre o ADL FTT e o dicionário de feições geoespaciais baseado nas especificações brasileiras.

Este trabalho tem por objetivo avaliar a abrangência do vocabulário criado a partir das especificações nacionais para produção de dados geoespaciais (verificando ausência ou excesso de termos) comparando-o com o dicionário de tipos de feições (*Feature Type Thesaurus – FTT*) da Biblioteca Digital Alexandria (*Alexandria Digital Library – ADL*), desenvolvida pela Universidade da Califórnia. Essa comparação será realizada pelo alinhamento sintático e semântico dos vocabulários, respeitando as relações entre os termos (sinônimos, generalização e especialização, *etc*) e a manutenção do significado durante a tradução dos termos.

A correlação entre termos em diferentes idiomas facilitará a associação das feições a ontologias adequadas, agregando semântica aos dados, assim como fornecerá subsídios para o desenvolvimento de mediadores e mecanismos de busca na web para descoberta de produtos e serviços.

2. Análise dos termos relativos ao domínio

Cada uma das especificações nacionais apresenta um critério primário de classificação das feições em categorias, de acordo com as suas particularidades. Neste tópico, serão comparadas as definições dadas para as categorias especificadas no manual T34-700, na MTD e na ET-EDGV. A seguir, encontram-se listadas as categorias empregadas em cada especificação.

- a) *T34-700* (9 categorias): Hidrografia, Vegetação, Limites, Pontos de Referência, Localidades, Sistema de Transporte, Altimetria, Edificações e Infraestrutura;
- b) *MTD* (8 categorias): Hidrografia, Vegetação, Limite, Ponto de Referência, Localidade, Sistema Viário, Hipsografia, Obra e Edificação;
- c) *ET-EDGV* (13 categorias): Hidrografia, Vegetação, Limites, Pontos de Referência, Localidade, Sistema de Transportes, Relevo, Energia e Comunicações, Abastecimento de Água e Saneamento Básico, Educação e Cultura, Estrutura Econômica, Administração Pública, Saúde e Serviço Social.

Uma análise imediata permite correlacionar, pela semelhança sintática, nomes de categorias comuns às 3 especificações: *Hidrografia*, *Localidade(s)*, *Limite(s)*, *Ponto(s) de Referência* e *Vegetação*. O termo *Sistema de Transportes* aparece explícito em duas especificações, e na terceira ocorre apenas o termo *Sistema* acompanhado do qualificador *Viário*. Contudo, a correlação entre *Sistema Viário* e *Sistema de Transportes* precisa ser avaliada com base nas descrições, da mesma forma que os demais itens sem semelhança sintática. O termo *Edificação* aparece no plural ou associado ao termo *Obra*, o que sugere a correlação ente duas categorias de especificações diferentes.

Analisando outras descrições, é possível estabelecer alguns vínculos entre alguns dos termos (grifando palavras relevantes em comum):

- *Altimetria*: elementos **hipsográficos**, que representam o **relevo** da superfície terrestre, relativamente ao *datum* vertical de referência. Este relevo é representado por meio de curvas de nível e pontos de altitude;
- *Hipsografia*: não foi encontrada descrição dessa categoria, mas uma menção ao uso de cores hipsométricas como forma de representação do relevo; e
- *Relevo*: Categoria que representa a forma da superfície da Terra e do fundo das águas tratando, também, os materiais expostos, com exceção da cobertura vegetal.

A correlação do termo *Edificação* leva a associar as categorias *Edificações* (TBCD) e *Obra e Edificação* (MTD). Dentre as categorias ainda não analisadas nas ET-EDGV, esse termo aparece na descrição de *Educação e Cultura*, *Estrutura Econômica*, *Administração Pública* e *Saúde e Serviço Social*. Pode-se interpretar que houve especialização da classe *Edificações* (TBCD), condicionando a pertinência da feição à categoria à sua finalidade.

Prosseguindo com a análise das categorias especificadas no T34-700, observa-se que o termo *Infraestrutura* ocorre parcialmente (na forma de estrutura) nas categorias *Energia e Comunicações* e *Abastecimento de Água e Saneamento Básico*, ambas das ET-EDGV. A correlação semântica acarretada pelo termo *Estrutura* ainda é

fraca, uma vez que o mesmo aparece na descrição de *Sistema de Transportes* (ET-EDGV). Contudo a maior afinidade da categoria com o termo *Transporte* e a descrição de *Infraestrutura* (T34-700) empregando um conjunto de feições relacionadas a atividades de infraestrutura torna a primeira associação mais apropriada. Algumas camadas de informação contidas na MTD correlacionam elementos referentes a *Energia e Comunicações e Abastecimento de Água e Saneamento Básico* como *Obra e Edificação*, pela própria definição de *Infraestrutura* (T34-700) como “edificações de”.

Outros termos são introduzidos na criação de classes especializadas ou como elementos de listas controladas, domínio de um atributo de uma classe, devido às suas semelhanças segundo o critério de classificação adotado.

3. Hierarquização de termos

Para verificar a abrangência da classificação empregada nas especificações nacionais para elaboração de documentos cartográficos, foi criada uma estrutura de dicionário conforme o padrão internacional contendo os termos referentes a feições e classes geoespaciais. Com isso, é possível identificar o contexto geoespacial de cada termo.

Foram desconsiderados na composição do vocabulário (consequentemente, do dicionário) termos referentes a classes não instanciáveis das ET-EDGV e outras que, reunidas, compõem uma feição, não representando, isoladamente, elementos geoespaciais. Neste caso, foram considerados os termos relativos às feições integrais (exemplo, considerar *rio* e descartar *trecho de massa d’água*).

Outro fator a ser considerado nas especificações é que o nome das classes são escritos visando à implementação (não havendo espaços, acentuação ou caracteres especiais, ou contendo abreviaturas). Para fins de composição do dicionário, foram selecionados os termos mais próximos já existentes nas outras especificações ou aqueles empregados na descrição da classe.

Com base nas definições contidas em ISO (1986) *apud* Breitman *et al* (2007), as feições geoespaciais podem ser classificadas como *Top Terms* (TT), *Broader Terms* (BT) e *Narrowed Terms* (NT). Essa classificação se baseia nos relacionamentos entre as feições representadas pelos termos, retratando principalmente *pertinência* e *similaridade*.

Estabelecer critérios para descrever o relacionamento entre termos é uma tarefa delicada, uma vez que estão sendo analisadas feições reais no terreno. Porém, alguns fatos podem auxiliar a escolha de termos relacionados (como os relacionamentos existentes entre as classes, no caso específico de termos das ET-EDGV).

A escolha de *top terms* exige um grau de abstração que deve levar em consideração os objetivos propostos na elaboração das especificações. Com o objetivo de facilitar a comparação entre os dicionários, os termos contidos nas especificações nacionais será associado à estrutura do FTT da ADL. Desta forma, a primeira categorização separa os elementos geoespaciais em Elementos Naturais e Elementos Artificiais (*man-made*). Seguindo esse critério, elementos de **Relevo**, **Hidrografia** e **Vegetação** integrariam o primeiro grupo, enquanto o segundo agrega as categorias *Sistema de Transporte*, *Edificações*, *Infraestrutura*, *Limites* e *Pontos de Referência*, que são **Elementos Construídos**. Nesta mesma categoria se encontram os elementos de

Relevo, Hidrografia (*Estruturas Hidrográficas*) e Vegetação produzidos por ação humana. *Localidades*, que são definidas e agrupadas segundo critérios político-administrativos, e outros tipos de particionamento do espaço com finalidades administrativas (setores censitários, por exemplo) são abrangidas pelo termo *Áreas Administrativas*, existente na ADL.

Todo TT já é um BT de um conjunto de termos e não possui um BT para si, por estarem no topo da classificação das feições. Dentro de cada TT, pode-se observar uma primeira divisão, com base nas categorias originais. Isso significa que *Sistema de Transporte, Limites, Edificações, Infraestrutura, Pontos de Referência, Localidades e Estruturas Hidrográficas* formam o primeiro conjunto de NT.

Dentro de cada uma das categorias das ET-EDGV (e em alguns casos no T34-700), podem ocorrer especializações que indicam correlação entre os termos envolvidos (outras classes). Os termos especializados tornam-se NT dos termos generalizados, que passam a ser os BT daqueles. Por outro lado, há classes que abrangem diversos tipos de feições sem que haja especialização (diferentes instâncias de uma classe).

Foram considerados sinônimos os termos *Hipsografia* (com *Altimetria*), *Sistema de Transportes* (com *Sistema Viário*) e aqueles qualificados como *temporário* (com o correspondente *intermitente*). Em geral, os termos possuem *descrição excludente*, mesmo em casos que a diferença entre os termos seja apenas na extensão. Nesses casos, o termo pode ser considerado quase sinônimo (UF).

4. Conclusões

Reunidos todos os termos relacionados a feições geoespaciais elencados nas especificações analisadas, recorreu-se à tradução dos mesmos de modo que esse conjunto pudesse ser comparado aos vocábulos e expressões constantes no ADL FTT. Foram empregados os recursos de tradução do *Google*, do Glossário de Gestão Costeira Integrada e, no caso de ambiguidade ou dúvida sobre a corretude da tradução, o termo foi consultado no glossário *WordNet*. A tabela 1 apresenta algumas das comparações realizadas.

Tabela 1 – comparação de termos descritos na ET-EDGV e no ADL FTT

| Termo EDGV | Tradução | Termo ADL FTT | Significado combina? |
|------------|------------------|--|----------------------|
| rio | <i>river</i> | <i>rivers</i> | Sim |
| cachoeira | <i>waterfall</i> | <i>Waterfalls (Used for: cascades, cataracts, falls)</i> | Em parte |
| fenda | <i>gap</i> | <i>gap</i> | Não |

Foi observado que a diferença no critério de categorização dos termos dificultou o trabalho, principalmente separando elementos naturais de elementos artificiais (construídos, antrópicos ou *man-made*). Portanto, a classificação em TT, BT e NT pode ser feita de maneira mais eficiente, sob o ponto de vista funcional das especificações

analisadas, adotando o critério de categorização das ET-EDGV por ser mais uniforme e objetivo.

Por se tratar de especificações para representação de feições presentes no Brasil, não são encontrados elementos relacionados aos ambientes polares (tundras, fiordes, geleiras, etc.) de regiões específicas do globo (*savanas, desertos, vulcões, áreas de atividade tectônica*, etc). Também foi observado que alguns termos usados no Brasil, quando traduzidos, tornam-se um único termo, assim como foi detectada a ausência de termos como *córrego, cânion e arquipélago*. Em contrapartida, algumas categorias apresentam termos de aplicação muito específica, algumas delas não presentes no FTT da ADL, encontradas apenas em glossários de domínios específicos. Isso reflete uma disparidade na abordagem das categorias, com grande profundidade em alguns casos e superficialmente em outros.

Como sugestões de trabalhos futuros, pode-se destacar a busca por ontologias de domínio que possam abranger as feições relacionadas nas especificações com suas propriedades e predicados textuais e espaciais, assim como a implementação de um dicionário de termos geoespaciais retratando seus relacionamentos e associando-os a URI (*Uniform Resource Identifier*), cuja codificação também é objeto de estudo.

Referências

- BRASIL, Decreto n. 6.666, de 27 de novembro de 2008. “Institui, no âmbito do Poder Executivo federal, a Infra-Estrutura Nacional de Dados Espaciais - INDE, e dá outras providências”.
- BREITMAN, Karin K., CASANOVA, Marco A., Truskowski, Walter, “Semantic Web: Concepts, Technologies and Applications”, London: Springer-Verlag London Limited. 2007. 327 p.
- COMISSÃO NACIONAL DE CARTOGRAFIA, “Especificações Técnicas para Estruturação de Dados Geoespaciais Digitais Vetoriais”. http://www.concar.ibge.gov.br/arquivo/94@ESPECIFICACOES%20TECNICAS%20PARA%20ESTRUTURACAO%20DE%20DADOS%20GEOESPACIAIS%20VETORIAIS_V2.0_10_10_2007.pdf. Brasília, 2007.
- DIRETORIA DE SERVIÇO GEOGRÁFICO, “Manual Técnico T34-700: Convenções Cartográficas”, 2ª Ed., Brasília: 1998.
- DIRETORIA DE SERVIÇO GEOGRÁFICO, “Tabela da Base Cartográfica Digital”, Brasília: 1997.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, “Mapoteca Topográfica Digital: Documento Geral – Versão 3.0.” IBGE, Diretoria de Geociências, Departamento de Cartografia. Rio de Janeiro: 1996.
- University of California – Santa Barbara, “Alexandria Digital Library Feature Type Thesaurus”, <http://www.alexandria.ucsb.edu/~lhill/FeatureTypes/ver070302/index.htm>. Abril.
- Revista de Gestão Costeira Integrada, “Glossário das Zonas Costeiras”, <http://www.aprh.pt/rgci/glossario/index.html>. Maio 2010.