

Mapeando o Perfil de Metadados Geoespaciais Brasileiro para Ontologias de Alto Nível

Helena Serrão Piccinini ^{1,2}

Marco Antonio Casanova ²

Karin Breitman ²

Rubens Nascimento Melo ²



Sumário

- Motivação
- Objetivo
- Alinhamento de Esquemas de BD
- Estratégia Utilizada
- Metadados Geoespaciais
- Perfil de Metadados Geoespaciais Brasileiro (MGB)
- Padrões de Metadados e Ontologias Utilizados
- Exemplos de Mapeamentos
- Dificuldades
- Conclusões
- Referências

Motivação

- Programas de *Governo Eletrônico* – promover integração de dados e serviços públicos
- INDE-Brasil (*Infra-estrutura Nacional de Dados Geoespaciais Brasileira* - decreto lei 6.666 de novembro de 2008)
- Definição de um padrão de metadados geoespaciais brasileiro → **Perfil MGB** (*Perfil de Metadados Geoespaciais Brasileiro* - baseado no padrão ISO 19115:2003)



Motivação

- Padrões de Metadados

- Funcionam como descrições complementares durante o processo de aquisição do dado
- Não são interpretados por máquina

- Web Services

- WSDL (*Web Services Description Language*) – são limitadas aos aspectos sintáticos

Motivação

Problemas de integração e localização de informações dos serviços públicos disponíveis na Web.

=

Problemas de integração e localização de informações que Web Semântica se propõe a resolver.

Problemas de integração de dados na Web Semântica.

=

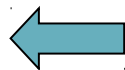
Problemas tratados pela área de Bancos de Dados.

Objetivos

- Esclarecer o significado atribuído a cada conceito do Perfil MGB
- Minimizar problemas de interoperabilidade semântica entre aplicações que usam o perfil e outras aplicações

Conceitos

- **Alinhamento de Esquemas de Banco de Dados**



- Metadados Geoespaciais
- Ontologias

Alinhamento de Esquemas de BD

- **Abordagem a priori**

- padrão para guiar o projeto do esquema integrador
- se não existir → proposta para um esquema global
- pode contornar o problema de heterogeneidade semântica entre esquemas de bancos de dados

- **Abordagem a posteriori**

- parte de esquemas já definidos e tenta alinhá-los

Alinhamento de Esquemas de BD

- **Abordagem sintática**

- usa indicações sintáticas: ex.: tipos de dados dos atributos e similaridades sintáticas entre os nomes dos elementos dos esquemas
- depende de uma suposição de que a proximidade sintática → em similaridade semântica
- pode levar a alinhamentos incorretos

- **Abordagem semântica**

- usa indicações semânticas para gerar hipótese sobre o alinhamento dos esquemas
- tenta detectar como os objetos do mundo real são representados em diferentes fontes de dados

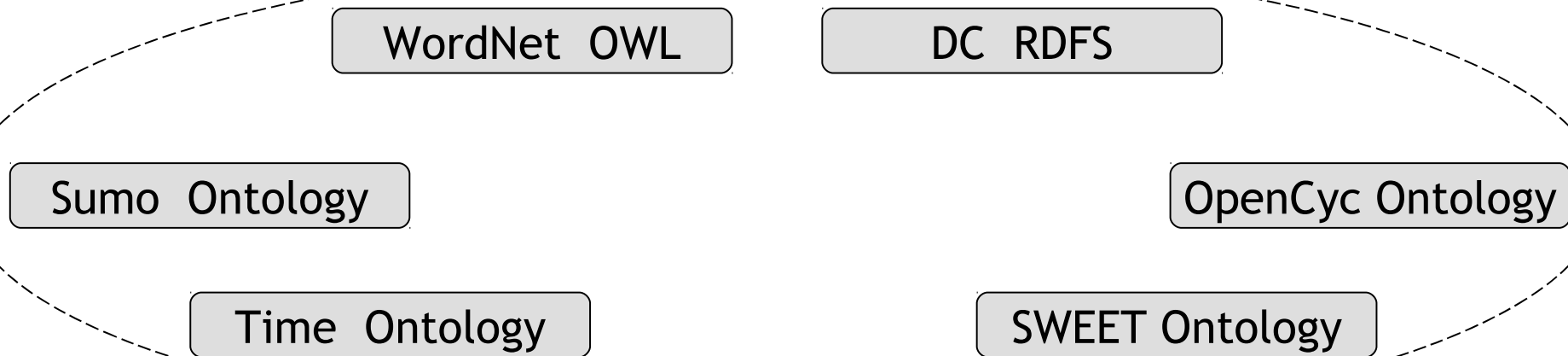
Estratégia Utilizada

Abordagem de Integração Semântica a Priori

- selecionar fragmentos de ontologias conhecidas que possuam conceitos pertinentes ao domínio Geográfico
- alinhar os conceitos destes fragmentos distintos, unificando-os
- publicar os conceitos unificados como uma ontologia
 - Cada termo do perfil → conceito de uma ontologia → identificação única e estável na Web
- Conjunto dos fragmentos das ontologias → Integrador entre a ontologia criada e outras ontologias

Estratégia Utilizada

Fragmentos das Ontologias



Modelo Integrador

Perfil MGB Mapeado

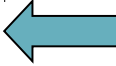
Geo Ontologia IBGE

Geo Ontologia INPE

Geo Ontologia IBAMA ...

...

Conceitos

- Alinhamento de Esquemas de Banco de Dados
- **Metadados Geoespaciais** 
- Ontologias

Metadados Geoespaciais

- **Metadados Geoespaciais** - descrevem os dados geoespaciais através de informações estruturadas
- **Dados Geoespaciais** - bases cartográficas em diversas escalas, dados geodésicos, atlas, mapas temáticos relativos à geografia, meio ambiente, etc

Metadados Geoespaciais

- **Estruturais** - definem a estrutura empregada na aquisição e armazenamento dos dados geoespaciais
 - Ex: Especificações Técnicas para a Estruturação de Dados Geoespaciais Digitais Vetoriais Brasileiros - Diretoria de Serviços Geográficos / Exército + CONCAR
- **Referência** - descrevem os dados geoespaciais através das informações de identificação, restrições legais, qualidade do dado, práticas de manutenção, extensão geográfica, representação espacial, sistema de referência, de tipo de conteúdo e distribuição
 - Ex: Perfil de Metadados Geoespaciais Brasileiro – ISO 19115 - CONCAR

Perfil de Metadados Geoespaciais Brasileiro

- É um fragmento da norma ISO 19115
 - ISO (*International Organization for Standardization*)
 - maior editora e promotora de padrões internacionais no mundo
 - ISO 19115
 - padrão para o armazenamento e distribuição de metadados geográficos
 - fornece informações sobre identificação, qualidade, referência espacial, distribuição de dados geográficos digitais, ...
 - padrão descritivo, não é interpretado por máquina
- Inserido no contexto da criação de INDE-Brasil
 - Cooperativa de dados Geoespaciais como a INSPIRE (*Infrastructure for Spatial Information in the European Community*) e NSDI (*National Spatial Data Infrastructure*)

Perfil de Metadados Geoespaciais Brasileiro

CONCAR (*Comissão Nacional de Cartografia*)

Comitê de Estruturação de Metadados Geoespaciais

- **Ministério do Planejamento, Orçamento e Gestão** – Instituto Brasileiro de Geografia e Estatística – **IBGE**
- **Exército** - Diretoria de Serviços Geográficos – **DSG**
- **Marinha** - Diretoria de Hidrografia Náutica – **DHN**
- **Aeronáutica** - Instituto de Cartografia Aeronáutica – **ICA**
- **Ministério do Meio Ambiente** - Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis - **IBAMA**
- **Ministério das Minas e Energia** - Serviço Geológico Brasileiro – **CPRM**
- **Ministério da Ciência e Tecnologia** - Instituto Nacional de Pesquisas Espaciais - **INPE**

Perfil de Metadados Geoespaciais Brasileiro

Seções:

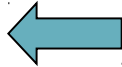
- Identificação
- Restrição
- Qualidade
- Manutenção
- Representação Espacial
- Sistema de Referência
- Conteúdo
- Distribuição
- Metametadados

Perfil de Metadados Geoespaciais Brasileiro

Sumário do Perfil MGB:

- | | |
|--|---|
| 1. Título (obrigatório) | 2. Data (obrigatório) |
| 3. Responsável (obrigatório) | 4. Extensão Geográfica (condicional) |
| 5. Idioma (obrigatório) | 6. Código de Caracteres do Conjunto de Dados Geoespaciais (condicional) |
| 7. Categoria Temática (obrigatório) | 8. Resolução Espacial (opcional) |
| 9. Resumo (obrigatório) | 10. Formato de Distribuição (obrigatório) |
| 11. Extensão Temporal e Altimétrica (opcional) | 12. Tipo de Representação Espacial (opcional) |
| 13. Sistema de Referência (obrigatório) | 14. Linhagem (opcional) |
| 15. Acesso Online (opcional) | 16. Identificador Metadados (opcional) |
| 17. Nome Padrão de Metadados (opcional) | 18. Versão da Norma de Metadados (opcional) |
| 19. Idioma dos Metadados (condicional) | 20. Código de Caracteres dos Metadados (condicional) |
| 21. Contato para Metadados (obrigatório) | 22. Data dos Metadados (obrigatório) |
| 23. Status (obrigatório) | |

Conceitos

- Alinhamento de Esquemas de Banco de Dados
- Metadados Geoespaciais
- **Ontologias** 

Padrões de Metadados e Ontologias Utilizados

Critérios para seleção das ontologias e padrões:

- conceitos relativos ao domínio Geoespacial
- não proprietários
- ferramentas de manipulação para facilitar seu manuseio
- estáveis
- desenvolvidos ou recomendados por entidades confiáveis (W3C)
- URI PURLs (*Persistent Uniform Resource Locators*)

Padrões de Metadados Utilizados

Padrão DC (*Dublin Core Metadata Element Set*)

- É uma norma ISO 15836-2003
- Possui quinze propriedades para descrição de recursos Web
- Genérico → bastante utilizado para descrever uma ampla gama de recursos da Web
- Disponível em RDF e RDFS
- Extensão do DC de janeiro de 2008 - disponibiliza um número maior de propriedades

Ontologias de Alto Nível

- São grandes modelos de representação do conhecimento humano e servem como referência para as ontologias de domínio
- Possuem hierarquias de classes e propriedades muito bem definidas

Ontologias de Alto Nível Utilizadas

- **SUMO** (*Suggested Upper Merged Ontology*) – desenvolvida pelo grupo de trabalho Suggested Upper Ontology da IEEE. Consolida conceitos com outras ontologias de domínio público
- **WordNet** – grande banco de dados léxico da língua inglesa
- **OpenCyc** – versão gratuita da ontologia Cyc, desenvolvida através do projeto Cyc (enCYClopedia) que é uma tentativa de criação da maior base de conhecimento do mundo

Ontologias de domínio Utilizadas

- **SWEET (*Semantic Web for Earth and Environmental Terminology*)** - (NASA) framework semântico criado para auxiliar a localização e uso de informações relacionadas às Ciências da Terra presentes em páginas da Web. Ontologias de domínio: Earth Realm, Physical Phenomena, Space, Physical Property e Biosphere
- **Time Ontology** - ontologia de conceitos temporais, usada para descrever o conteúdo temporal de páginas da Web e as propriedades temporais de Web Services

Exemplo de Mapeamento

Perfil MGB	ISO 19115
Latitude Limítrofe Sul - coordenada meridional do limite da extensão do conjunto de dados geoespaciais, expressa em latitude, utilizando graus decimais (e valores positivos a norte)	southBoundLatitude - southern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north)

Exemplo de Mapeamento

SUMO

Latitude - the angular distance between an imaginary line around a heavenly body parallel to its equator and the equator itself.

Bound - the line or plane indicating the limit or extent of something.

South - a location in the southern part of a country, region, or city.

<http://sigma.ontologyportal.org:4010/sigma/WordNet.jsp?synset=108595531>

<http://sigma.ontologyportal.org:4010/sigma/WordNet.jsp?synset=108512259>

<http://sigma.ontologyportal.org:4010/sigma/WordNet.jsp?synset=108561583>

OPENCYC

latitudeMin - means that the [SpatialThing_Localized](#) OBJ has its most southern part(s) geographically located at LAT, where LAT denote a latitude line [[LatitudeLine](#)] in the "geographical universe".

[LatitudeLine](#) - The collection of lines around the earth which are parallel to the [EarthsEquator](#).

[EarthsEquator](#) - The equator of [PlanetEarth](#). [EarthsEquator](#) is the [LatitudeLine](#) which has exactly 90 degrees separation from the North and South Poles. [LatitudeLines](#) are conventionally measured by their separation from the Earth's equator in degrees, minutes, and seconds. See also [equatorOfSpheroid](#).

<http://sw.opencyc.org/concept/Mx4rvZq7yJwpEbGdrcN5Y29ycA>

SWEET

```
<owl:Class rdf:about="#Colatitude">
<rdfs:subClassOf rdf:resource="#HorizontalCoordinate"/>
<rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#dir;#hasDirection"/>
    <owl:allValuesFrom rdf:resource="#dir;#South"/>
  </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
```

<http://sweet.jpl.nasa.gov/2.0/spaceCoordinates.owl#Colatitude>

Exemplo de Mapeamento

MGB	DC	WordNet	SUMO	OpenCyc	SWEET
Title	Title (específico)	Title (geral)	Title (geral)	Title document (geral)	—
tin 004	—	triangulation + irregular + network (gerais)	—	—	—
Grid	—	Grid (geral)	Grid (geral)	—	Grid (específico)
Vertical Datum	—	Datum (geral)	Datum (geral)	VerticalDatum (específico)	Datum (geral)
Reference System	—	—	—	—	Spatial Reference System (específico)
Projection System	—	—	—	surface projection (específico)	—
Ellipsoid	—	Ellipsoid (geral)	Ellipsoid (geral)	Ellipsoid (específico)	—
Spatial resolution	—	Scale (específico)	Scale (específico)	Scale (específico)	Spatial resolution (específico)
Lineage	Provenance (específico)	—	—	—	—
South Bound Latitude	—	South + Bound + Latitude (gerais)	South + Bound + Latitude (gerais)	latitudeMin (específico)	Colatitude (específico)

Dificuldades

- ontologias de alto nível possuem conceitos muito gerais
- ontologias de domínio podem possuir poucos comentários
 - examinar uma linguagem escrita para computadores
 - navegação em sua estrutura pode ser confusa ou demorada
- pode não existir uma correspondência direta entre conceitos - combinação de vários conceitos para uma melhor aproximação com o conceito que está sendo descrito na ontologia de domínio
- mapeamentos podem requerer várias ontologias de alto nível e de domínio
- mapeamentos podem requerer a participação de especialistas
- decisão de quais ontologias serão utilizadas não é trivial

Conclusões

- Apesar das dificuldades este trabalho é executado apenas uma vez
- A aplicação da estratégia apresentada atende os objetivos
 - clarifica o significado dos conceitos do Perfil MGB
 - minimiza problemas de interoperabilidade entre aplicações que usam o perfil e outras aplicações
- Se o alinhamento a priori for executado criteriosamente, o problema de heterogeneidade semântica entre ontologias pode ser contornado

Referências

- Bednar P, Furdik K., Kleimann M., Klischewski R., Skokan M., Ukena S.. Semantic Integration of e-Government Services in Schleswig-Holstein - Proceedings of the 7th international conference on Electronic Government -Turin, Italy – 2008.
- Berners-Lee, Tim; Hendler, James; Lassila, Ora. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American. 2001.
- Berners-Lee, Tim; Weitzner, D.; Hall, W.; Shadbolt, N; Hendler, J.; - Web Science: An Interdisciplinary Approach to Understanding the Web – communications of the ACM – julho de 2008; vol. 51; no.7.
- Breitman, K., Casanova, M., and Truszkowski, W. (2007) *Semantic web: concepts, technologies, and applications*. Springer, London.
- CONCAR (Comissão Nacional de Cartografia) <http://www.concar.ibge.gov.br/> - acessado set/2009.
- Casanova, M. A.; Breitman, K. K.; Marins, A. (2007) – Database Conceptual Schema Matching – IEE Computer, volume 40, págs. 102-104 – outubro de 2007.
- Dublin Core Metadata Element Set <http://dublincore.org/documents/dces/> acessado set/2009.
- FGDC - <http://www.fgdc.gov/> acessado em set/2009.
- Gazola, A.; Brauner, D. e Casanova, M. A. - Gazetteer.br: Um Serviço de gazetteer via Web – PUC-Rio-2007.
- Governo Eletrônico Brasileiro - egov <http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padres-de-interoperabilidade> - acessado em set/2009.
- IFLA (2006) Digital Libraries: Metadata Resources. Available at: <http://www.ifla.org/II/metadata.htm>.
- ISO 19115:2003 – Geographic Information – Metadata.
- Marins, A.L.Almeida – Modelos Conceituais para Proveniência – dissertação de Mestrado – PUC-Rio – 2008
- Ontology Browser - <http://code.google.com/p/ontology-browser/> acessado em set/2009.
- OpenCyc - <http://www.opencyc.org> - acessado em setembro de 2009.
- Perfil de Metadados Geoespaciais do Brasil (Perfil MGB) versão preliminar - agosto de 2009 - <http://www.concar.ibge.gov.br/plenaria14/PerfilMGB-Versao1.pdf>
- PURLs (Persistent Uniform Resource Locators) <http://purl.org/docs/index.html> - acessado em set/2009.
- SIGMA <http://sigma.ontologyportal.org:4010/sigma/Browse.jsp?kb=SUMO> - acessado em set/2009.
- SUMO - <http://www.ontologyportal.org/> - acessado em set/2009.
- SWEET - <http://sweet.jpl.nasa.gov/ontology/> - acessado em set/2009.
- SWOOGLE Semantic Web Search - <http://swoogle.umbc.edu/> acessado em set/2009.
- Time Ontology in OWL - <http://www.w3.org/TR/owl-time/> acessado em set/2009.
- WordNet - <http://wordnetweb.princeton.edu/perl/webwn> - acessado em set/2009.
- W3C - <http://www.w3.org/> acessado em set/2009
- United Nations e-Government Survey (2008) - From e-Government to Connected Governance. Disponível <http://unpan1.un.org/intradoc/groups/public/documents/UN/UNPAN028607.pdf> acessado em jun/2009.

Obrigada

Helena S. Piccinini ^{1/2}
hpiccinini@inf.puc-rio.br

Marco A. Casanova ²
casanova@inf.puc-rio.br

Karin K. Breitman ²
karin@inf.puc-rio.br

Rubens N. Melo ²
rubens@inf.puc-rio.br

(2)

