

Uma Abordagem de Gerenciamento Semântico de Experimentos Meteorológicos em Pluviometria

Thiago M. da S. Barbosa¹, Sérgio Manuel Serra da Cruz^{1,2}

¹ Departamento de Matemática– Universidade Federal Rural do Rio de Janeiro (UFRRJ)

² Programa de Pós Graduação em Modelagem Matemática e Computacional (UFRRJ)

thiago@ufrrj.br, serra@ufrrj.br

Abstract. *This paper presents a semantic approach to aid meteorologists to organize their data. The approach relies on the cycle of knowledge management and reasoning supported by foundational ontologies. We present a tool that aid meteorologists to develop semantic search to assess the quality of the rainfall data used in their experiments.*

Resumo. *Este trabalho apresenta uma abordagem de gerenciamento semântico de dados meteorológicos que se apóia no ciclo de gestão do conhecimento e em ontologias de fundamentação. Apresentamos uma ferramenta que auxilie os meteorologistas a desenvolverem buscas semânticas para avaliar a qualidade dos dados de precipitação.*

1. Introdução

Meteorologia é um empreendimento global que consome muitos recursos computacionais e humanos para estudar o tempo e o clima. Os experimentos em meteorologia fazem uso intensivo de estatística, modelos matemáticos, simulações computacionais e são divididos em três etapas subseqüentes e distintas. Na etapa de *pré-processamento*, foco deste trabalho, os dados são transformados, graças a um conjunto de técnicas específicas, de brutos e não tratados em dados curados e consistentes. Os dados meteorológicos não são indexados, podendo ser armazenados em unidades de medida distintas e com descontinuidade cronológica. Outra dificuldade diz respeito à forma de se preencher as possíveis falhas nas séries de dados. Entenda-se por falha um dado errôneo ou ausente.

Este trabalho tem como objetivo apresentar uma abordagem de gerenciamento semântico de dados meteorológicos que se apóia no ciclo de gestão do conhecimento e nas ontologias bem fundamentadas [Guizzardi *et al.* 2005; Cruz 2011] para o desenvolvimento de uma ferramenta que permita aos meteorologistas desenvolverem buscas semânticas para avaliar a qualidade dos dados de precipitação utilizados em seus experimentos científicos.

2. Materiais e Métodos

2.1. Dados Meteorológicos de Precipitação Pluvial e seus Pré-Processadores

Os dados utilizados nesta pesquisa são públicos e fazem parte de longas séries meteorológicas (maiores ou iguais a 20 anos e com início a partir de 1960) coletados em 77 estações meteorológicas geograficamente dispersas em todo o estado do Rio de Janeiro (mantidas pela Agência Nacional de Águas - ANA). Esta região dispõe de totais de precipitação dentre os

maiores da região Sudeste do Brasil e está sujeita a ocorrência de eventos extremos de chuva.

As séries foram extraídas dos sistemas FAO e HidroWeb (2005). Os dados são muito heterogêneos, apresentam limitações estruturais e inconsistências semânticas, por exemplo, o HidroWeb possui 11 definições distintas para o termo "bacia", porém apenas o termo "bacia hidrográfica" é fiel a sua definição no domínio da meteorologia [Mangan et al 2001]. Outras inconsistências semânticas são verificadas nos termos "estado", "municípios", "entidades", entre outros. Os pré-processadores utilizados neste trabalho foram anteriormente desenvolvidos por nosso grupo de pesquisa. Maiores detalhes em [Filho et al 2013].

2.2. Ontologias de Fundamentação e de Proveniência

Este trabalho utilizou a metamodelagem de ontologias bem fundamentadas como referencial teórico, adotou-se a *Unified Foundational Ontology* (UFO) [Guarino 2008, Guizzardi 2005]. Este trabalho também utilizou a ontologia *Open Provenance Ontology* (OvO) como suporte teórico para os conceitos relacionados com a proveniência dos experimentos científicos. OvO é uma ontologia de domínio proposta e desenvolvida por [Cruz, 2011 e Cruz et al. 2012], ela tem como pilares a UFO e a especificação OPM (*Open Provenance Model*) [Moreau et al. 2013].

3. Resultados

Meteoro é uma ontologia de aplicação, bem fundamentada, que trata dos conceitos ligados a transformação de dados brutos de precipitação em dados curados e metadados de proveniência para o domínio da pluviometria. *Meteoro* estende a ontologia OvO para o domínio específico da precipitação, além disso reutiliza os seus conceitos ligados a proveniência de experimentos científicos em larga escala.

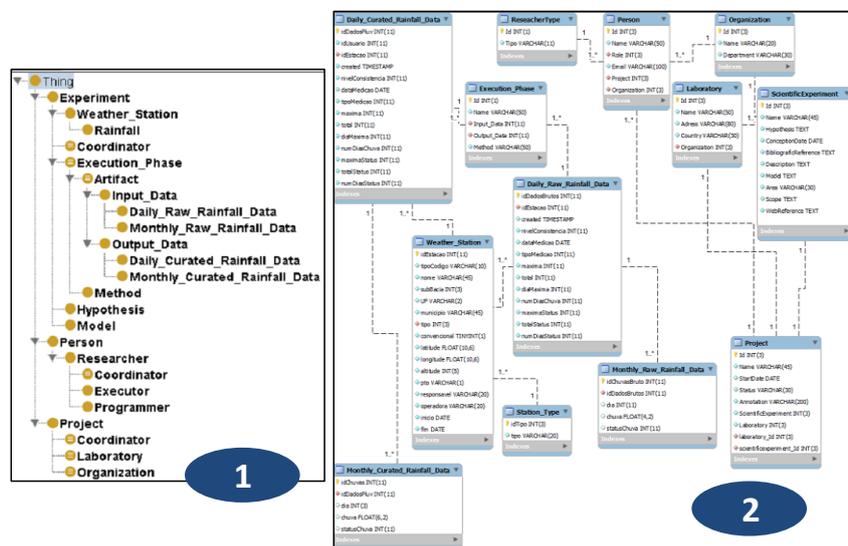


Figura 1 e 2 - (1) Fragmento da ontologia *Meteoro* e (2) Modelo de Dados (adaptado de [Filho et al, 2013]).

3.1. Ontologia *Meteoro* em OWL

Para que a ontologia *Meteoro* seja processada ela deve ser transposta para uma linguagem que apóie inferências. Neste caso se codificou em OWL-DL através do editor Protégé. Este processo é complexo, adotaram-se as regras de mapeamentos de classes UFO-A para OWL DL anteriormente definidas por Cruz (2011).

A ontologia *Meteoro* (modelada em OntoUML) e as Figuras 1 e 2 estão correlacionadas. Por exemplo, o conceito *Weather Station* identificado como um sortal do tipo *<Kind>* representa unicamente as instâncias das estações meteorológicas da ANA; seus subtipos *Rainfall* são *<Subkind>* que herdam seus princípios de identidades. Na Figura 2 verificam-se as classes em OWL. Essas classes e suas propriedades são mapeadas para a tabela “*Weather_Station*” no modelo de dados (Figura 3). Ou seja, é possível identificar unicamente uma estação do tipo pluviométrica dentre todas as 77 estações avaliadas no experimento. Raciocínio análogo se aplica aos demais conceitos correlacionados com aos dados (brutos e curados), agentes, processos e proveniência.

3.2. Mapeamentos entre os Conceitos da Ontologia e o Banco de Dados

Para que a buscas semânticas e a navegação sobre dados e descritores de proveniência sejam processados pelo computador faz-se necessário um segundo nível de mapeamentos. É necessário associar os conceitos (classes e propriedades codificadas em OWL DL) com as tabelas e atributos do modelo de dados (Figura 2). Este segundo nível de mapeamento é possível graças à linguagem D2R própria do servidor D2RQ, ele permite o acesso ao conjunto de dados como se fossem triplas RDF.

3.3. A Ferramenta *WebOntology*

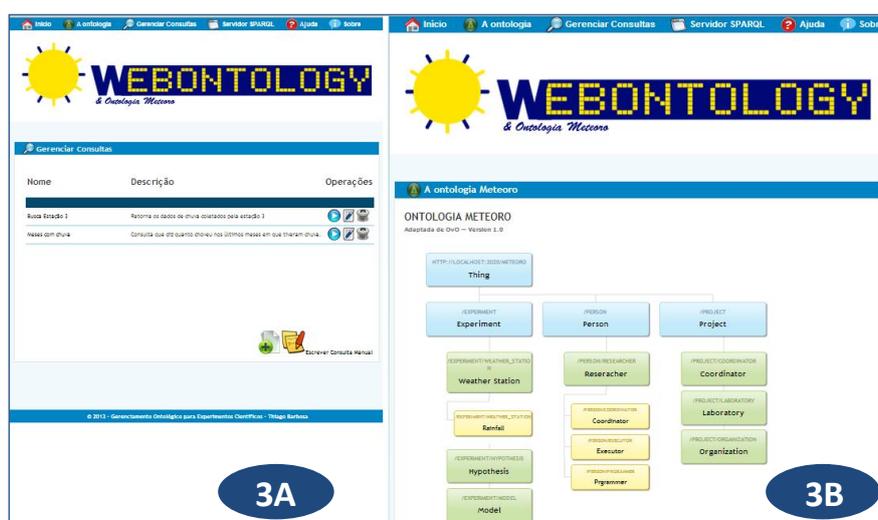


Figura 3A e 3B – Gerenciamento de consultas e navegação através dos conceitos.

Atualmente existem interfaces e motores de execução que permitem o processamento de consultas SPARQL através da Web. No entanto, seu uso ainda não é trivial para os meteorologistas. Como prova de conceito foi desenvolvido um protótipo da *WebOntology* em linguagem PHP e MySQL. A ferramenta Web utiliza o servidor D2RQ, capaz de carregar ontologias em OWL e os seus mapeamentos D2R para processar as buscas semânticas sobre o banco de dados relacional. Das funcionalidades codificadas, mencionamos as duas mais relevantes para o momento (Figuras 3A e 3B). (i) **Gerenciar Consultas**: Visando a reduzir o retrabalho, esta funcionalidade (Figura 3A) permite ao pesquisador gerenciar graficamente suas buscas semânticas (codificar, cadastrar, executar, excluir ou alterar consultas SPARQL) mais utilizadas. (ii) **Visualizar Ontologia**: Outra funcionalidade é a possibilidade de carregar, navegar e visualizar os conceitos da ontologia *Meteoro* em forma de um diagrama hierárquico

sem a necessidade do uso de editores de ontologias (Figura 3B). Além de obter informações sobre a ontologia e seus conceitos, com uma baixa curva de aprendizado.

4. Trabalhos Relacionados

Atualmente existem trabalhos que exploram a gestão do conhecimento e ontologias em diversas áreas. Dentre os trabalhos que consideram a importância dos descritores de proveniência no que diz respeito à qualidade do dado científico, destacam-se o Cuebee [Mendes 2007], uma aplicação flexível e extensível que fornece uma interface amigável para orientar os usuários durante o processo de formulação de consultas complexas. Na área da computação quântica, Medeiros (2010) consorcia uma ontologia de aplicação e um *framework* nas atividades de gerenciamento de dados relacionados aos testes experimentais de circuitos quânticos e criptografia. Percebe-se que esses trabalhos não atuam ao longo de todas as quatro etapas do ciclo da gestão do conhecimento nem incorporam metamodelagem baseadas em ontologias de fundamentação ou não tiveram continuidade em seu desenvolvimento.

5. Conclusão e Trabalhos Futuros

Este trabalho propôs uma abordagem de gerenciamento semântico de dados meteorológicos que se apóia no ciclo de gestão do conhecimento e em ontologias. Apresentou uma primeira versão de uma ontologia de aplicação e um protótipo de uma ferramenta semântica que facilita o desenvolvimento de buscas semânticas baseadas em consultas RDF-SPARQL.

6. Referências

- Cruz, S. M. S. (2011) "Uma Estratégia De Apoio À Gerência De Dados De Proveniência Em Experimentos Científicos". Tese de Doutorado. PESC/COPPE-UFRJ.
- Cruz, S. M. S, Campos, M. L. M. Mattoso, M. (2012) "A Foundational Ontology to Support Scientific Experiments". ceur-ws.org/Vol-728/paper6.pdf.
- Filho, G. R. L, Precinoto, R. S, Correia, T. P, Santos, E. O, Lyra, G. B, Cruz, S. M. S. (2013). "Assimilação, Controle de Qualidade e Análise de Dados de Meteorológicos Apoiados por Proveniência", VII e-science Workshop, XXXIII CSBC.
- Guarino, N. (1998) "Formal Ontology and Information Systems", In: Proceedings of the 1st International Conference Formal Ontology in Information Systems, pp. 3-15.
- Guizzardi, G. (2005) "Ontological Foundations for Structural Conceptual Models", PhD Thesis, University of Twente, Netherlands.
- HidroWeb (2005). "Sistemas de Informação Hidrológicas". <http://hidroweb.ana.gov.br/>.
- Mangan et al (2001) "Modelos de Domínio e Ontologias: uma comparação através de um estudo de caso prático em hidrologia". Int. Symp. on Knowledge Management/Document Management, pp.149-172.
- Medeiros, L. de F. (2010). "Framework para Engenharia e Processamento de Ontologias Utilizando Computação Quântica". Tese de Doutorado. UFSC, Florianópolis.
- Mendes, P. N, Sheth A. P, "Complex Queries for Hypothesis Validation on the Web", Knoesis Center, Computer Science & Engineering, Wright State University, <http://knoesis.wright.edu/students/pablo/download/BiomedWeb2007-CuebeePoster-Print.pdf>.
- Moreau, L. et al (2013)., Open Provenance Model (OPM) OWL Specification. <http://openprovenance.org/model/opmo>