

Ontologia de Aplicação para o Lago Batata

Adriano Neves de Souza, Adriana Pereira de Medeiros

Instituto de Ciência e Tecnologia – Universidade Federal Fluminense – Rio das Ostras
Rio de Janeiro – RJ – Brazil

adriano_souza@id.uff.br, adrianamedeiros@puro.uff.br

Abstract. *This paper presents an application ontology that describes the data of the research done by researchers from UFRJ - Macaé in Lake Batata, an aquatic ecosystem that suffered environmental impacts. The ontology is the base for the creation of a data repository in RDF, which will enable the development of Web applications to support research about this lake.*

Resumo. *Este trabalho apresenta uma ontologia de aplicação que descreve os dados das pesquisas feitas pelos limnólogos da UFRJ-Macaé no Lago Batata, um ecossistema aquático que sofreu impactos ambientais. A ontologia proposta é base para a criação de um repositório de dados em RDF que permitirá o desenvolvimento de aplicações Web para apoiar essas pesquisas.*

1. Introdução

A complexidade ecológica dos ecossistemas aquáticos, causada pelo grande volume de dados de coleta, cria dificuldades na busca pela compreensão desses ecossistemas. Essa compreensão gera conhecimentos científicos que propiciam alternativas de recuperação ou mitigação dos impactos externos causados [Bozelli et al. 2000]. Alguns esforços têm sido feitos para alavancar o conhecimento sobre ecologia pela organização de dados de pesquisas. Um exemplo é o projeto PELD [Esteves et al. 2004], uma iniciativa governamental para incentivar soluções através do levantamento e organização de dados de pesquisas em ecossistemas. Pesquisadores limnólogos da Universidade Federal do Rio de Janeiro (UFRJ), em Macaé-RJ, têm tido dificuldades com a falta de formalização e estruturação dos dados de pesquisas que vêm sendo feitas há décadas no Lago Batata [Bozelli et al. 2000]. Este lago é um ecossistema aquático amazônico, localizado na região de Porto Trombetas em Oriximiná/PA, que sofreu impactos ambientais por depósitos de rejeitos de produção de bauxita, minério utilizado na produção de alumínio.

Muitos ramos da biologia possuem seus domínios descritos por ontologias, e o uso delas em biodiversidades tem sido apontado como solução para a obtenção de conhecimento científico [Campos et al. 2007]. Ontologias permitem modelar parte de uma realidade, suas entidades, relações e restrições, visando definir um entendimento comum sobre um domínio [Daltio 2007]. As pesquisas feitas por [Moura et al. 2012], [Campos et al 2007] e [Daltio e Medeiros 2007] buscam soluções para integração de dados e ferramentas de Web Semântica, mas tais soluções especificam ontologias que não descrevem termos propostos neste trabalho.

Este trabalho apresenta uma ontologia de aplicação para o Lago Batata, definida a partir dos dados de análise e coleta levantados pelos limnólogos da UFRJ-Macaé, a

fim de auxiliá-los em suas pesquisas. A descrição semântica desses dados possibilitará uma análise mais rica e detalhada do lago através de inferências no modelo de dados estruturado obtido. Além disso, apoiará a análise da recuperação do lago e a busca por soluções que mitiguem os impactos ambientais. A seção 2 apresenta a ontologia de aplicação proposta. A seção 3 apresenta exemplos de instâncias e consultas viabilizadas pela ontologia. A seção 4 apresenta algumas conclusões e trabalhos em andamento.

2. Ontologia de Aplicação do Lago Batata

A ontologia proposta foi criada usando a metodologia *Ontology Development 101* [Noy e McGuinness 2001] e a ferramenta Protégé 4.3 (<http://protege.stanford.edu/>). Ela foi especificada com a linguagem OWL (*Ontology Web Language*), especificamente OWL DL 2, com 35 classes e 222 axiomas. O domínio foi definido como Lago Batata e o escopo foi estabelecido pelas seguintes questões de competência, validadas pelos pesquisadores: i) Qual é o período de coleta com maior concentração de clorofila em um dado ano? ii) Qual pulso de inundação apresentou o maior valor de turbidez em um determinado ano? iii) Qual pulso de inundação apresentou maior percentual de matéria orgânica no sedimento em um dado ano? iv) Qual é o pulso de inundação de um determinado período? v) Quais coletas foram feitas em áreas impactadas em um dado período? Pesquisas foram feitas em alguns repositórios de ontologias: <http://www.daml.org/>, http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library, http://datahub.io/pt_BR/dataset/schemapedia e <http://swoogle.umbc.edu/>. Nas ontologias *hydroBodyOfWater* (<http://sweet.jpl.nasa.gov/2.0/hydroBodyOfWater.owl>) e *Geography* (<http://www.daml.org/ontologies/412>) foram encontrados alguns termos genéricos relacionados à ontologia proposta, com caráter descritivo. No entanto, não foram encontradas ontologias para o domínio pesquisado neste trabalho. A Figura 1 apresenta a visualização em grafo das principais classes da ontologia (vértices) e das relações entre as classes (arestas), descritas por propriedades de objetos em OWL.

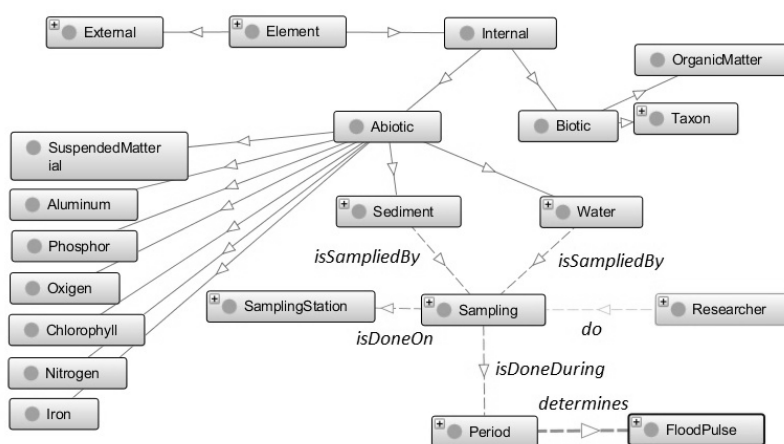


Figura 1- Classes e Propriedades da Ontologia (parcial)

A classe *Sampling* descreve a coleta feita pelo pesquisador (*Researcher*) nas estações de coleta, representadas pela classe *SamplingStation*. A classe *SamplingStation* possui duas propriedades de dados, *coordinates* do tipo *literal* e *impacted* do tipo *boolean*, que especificam respectivamente a localização geográfica da estação de coleta e se é uma área impactada ou não. A propriedade de objeto *isDoneOn* determina a relação entre *Sampling* e *SamplingStation*. A classe *FloodPulse* especifica os pulsos de

inundação do lago, que são as etapas do processo de enchimento e esvaziamento do lago, e não possui propriedade de dados, pois a identificação das instâncias é feita pela própria URI (*Enchente, AguasAltas, Vazante, AguasBaixas*). A classe *Period*, através da propriedade de dados *Date*, especifica a data (mês e ano) em que a coleta é feita e está relacionada à classe *FloodPulse* pela propriedade de objeto *determines* (*Period->determines->FloodPulse*). Essa propriedade descreve a relação entre os meses do ano e as fases de inundação do lago, que podem sofrer mudanças ao longo dos anos por não existir um padrão que estabeleça que um mês terá um pulso de inundação específico. A propriedade *isDoneDuring* (*Sampling->isDoneDuring->Period*) descreve que uma coleta é feita durante um determinado período, sendo que um mesmo período pode ter várias coletas, de acordo com a quantidade de estações de coleta de onde são retiradas as amostras. As classes *Sediment* e *Water* representam todos os dados coletados de sedimento e água no lago e estão relacionados à coleta (*Sampling*) através da propriedade de objeto *isSampledBy*. Todos os dados de coleta relacionados à água são descritos pelas propriedades de dados das classes *Water, SuspendedMaterial, Aluminum, Chorophyll, Iron, Nitrogen, Oxigen e Phosphor*. Por limitações de espaço, essas propriedades não serão descritas.

Toda propriedade possui uma direção, de *Domain* para *Range*, e pode ter restrições. O Quadro 1 mostra a propriedade de objeto *isDoneOn* entre as classes *Sampling* e *SamplingStation* em OWL, para a qual é definida a restrição *FunctionalProperty*. Nessa restrição, uma coleta *x* pode ser feita em somente uma estação de coleta *y*. Usando a tripla *Sampling->isDoneOn->SamplingStation* é possível buscar informações de coletas agrupadas por estações de coleta. Essas restrições são garantidas pelos motores de inferência (*reasoners*), que garantem que as instâncias de dados obedecerão às regras da ontologia [Hitzler et al. 2009].

```
<owl:ObjectProperty rdf:about="isDoneOn">
<rdf:type rdf:resource="FunctionalProperty"/>
<rdfs:domain rdf:resource="Sampling"/>
<rdfs:range rdf:resource="SamplingStation"/>
</owl:ObjectProperty>
```

Quadro 1 - Exemplo de propriedade funcional

```
<owl:ObjectProperty rdf:about="isDeterminedBy">
<owl:inverseOf rdf:resource="determines"/>
</owl:ObjectProperty>
```

Quadro 2 - Exemplo de Propriedade Inversa

O Quadro 2 mostra a propriedade de objeto *isDeterminedBy* com a restrição *inverseOf* que a define como inversa da relação *determines*. Assim, a relação *Period->determines->FloodPulse* também pode ser descrita como *FloodPulse->isDeterminedBy->Period*. Restrições como essas possibilitam detalhamento semântico do modelo de dados e, com o apoio dos *reasoners*, consultas podem obter resultados mais apurados, como mostra o exemplo da próxima seção.

3. Exemplo de Consulta utilizando a Ontologia

Consultas em dados RDF são realizadas com a Linguagem SPARQL. Nela, as condições são descritas com o padrão sujeito->predicado->objeto. Por limitação de espaço, apresentamos apenas um exemplo de consulta a ser executada por uma aplicação web na análise dos dados de pesquisa do Lago Batata. O Quadro 4 mostra a consulta para a questão “Qual é o pulso de inundação de um determinado período?”. *SELECT* especifica os dados que serão retornados no resultado. A cláusula *WHERE*

especifica as condições da consulta, definidas com base nas propriedades da ontologia. Abaixo da consulta é apresentado o resultado, mostrando que o pulso de inundação do período consultado foi águas baixas, de acordo com os dados da instância da classe *Period* ilustrada no Quadro 3. Note que, embora os dados estejam descritos com a relação *isDeterminedBy* no repositório ao invés de *determines*, o resultado da consulta é o mesmo, dado que as duas propriedades de objeto foram definidas como inversas.

```
<owl:NamedIndividual about="Period5c01092593a8901f">
<type resource="Period"/>
<localhost:date datatype="string">12/2001</date>
<localhost:isDeterminedBy resource="AguasBaixas"/>
</owl:NamedIndividual>
```

Quadro 3 - Exemplo de Instância

```
select distinct ?date ?floodpulse
WHERE {
  ?period rdfs:type lago:Period.
  ?period lago:determines ?floodpulse.
  ?period lago:date ?date.
  FILTER (?date = "12/2001")
}
```

date	floodpulse
12/2001	AguasBaixas

Quadro 4 - Exemplo de Consulta

4. Conclusões e Trabalhos Futuros

Neste trabalho propomos uma ontologia de aplicação para descrever semanticamente os dados das pesquisas feitas pelos pesquisadores limnólogos da UFRJ-Macaé no lago Batata. Sua principal contribuição é a criação de um repositório de dados de pesquisa em RDF, que será utilizado pela aplicação web que está sendo desenvolvida para apoiar a análise desses dados. O objetivo é a produção de conhecimento científico a partir das análises feitas por consultas semânticas. Após a conclusão da aplicação, será realizado um experimento junto aos pesquisadores envolvidos para validar a aplicação.

5. Referências

BOZELLI, REINALDO L.; ESTEVES, FRANCISCO A.; ROLAND, F. Lago Batata: Impacto e Recuperação de um Ecossistema Amazônico. UFRJ/SBL- RJ, 2000.

CAMPOS, J. L.; NETTO, J. F. D. M.; CASTRO, A. N. DE; ALBUQUERQUE, A. C. F. Ontologias para Interoperabilidade de Modelos e Sistemas de Informação de Biodiversidade, 2007.

DALTIO, J. Aondê: Um Serviço Web de Ontologias para Interoperabilidade em Sistemas de Biodiversidade, 2007.

ESTEVES, F. A. ; SCARANO, F. R. ; ROCHA, C. F. D. Pesquisa de Longa Duração na Restinga de Jurubatiba: Ecologia, História Natural e Conservação. 1. ed. Rio de Janeiro: RiMA Editora, 2004. v. 1. 376p.

HITZLER, P. et al. "OWL 2 Web Ontology Language Primer", 2009, <http://www.w3.org/TR/2009/REC-owl2-primer-20091027>.

MOURA, A.; PORTO, F.; POLTOSI, M. Integrating Ecological Data Using Linked Data Principles. ONTOBRAS-MOST 2012: 156-167.

NOY, N.; MCGUINNESS, D. Ontology development 101: A guide to creating your first ontology. Development, v. 32, p. 1–25, 2001. Disponível em: <http://protege.stanford.edu/publications/ontology_development/ontology101.pdf>.