

Implementação de um geocatálogo para auxílio na localização e recuperação de dados geográficos abertos



Drausio Gomes dos Santos

é servidor do Tribunal de Contas da União, mestre em Computação Aplicada pela Universidade de Brasília, especialista em Gestão da Tecnologia da Informação pelo Cin/UFPE, tecnólogo em Processamento de Dados pela FATEC/SP e bacharel em Geografia pela USP.



Alexandre Zaghetto

é mestre e doutor em Engenharia Elétrica pela Universidade de Brasília (UnB), possui graduação em Engenharia Elétrica, com ênfase em Eletrônica, pela Universidade Federal do Rio de Janeiro, além de professor adjunto no Dept. de Ciência da Computação da UnB e líder do Grupo de Pesquisa em Sistemas Biométricos (BiTGroup).

RESUMO

Este artigo trata da construção de um geocatálogo que utiliza a semântica e ontologias para auxílio no processo de descoberta de recursos geográficos em serviços de uma infraestrutura aberta de dados espaciais. Sua arquitetura incorpora um banco de dados não relacional orientado a grafos e é passível de implementação em ambiente de nuvem. Para sua concepção foi utilizada a metodologia de desenvolvimento de sistemas SERVUS, especializada em arquitetura orientada a serviços geoespaciais.

Palavras-chave: Engenharia de Software. Geocatálogo. SERVUS. IDE. Semântica. Dados abertos.

1. INTRODUÇÃO

Tem sido um desafio para os órgãos da administração pública obter uma visão espacial do resultado das ações governamentais de forma simplificada e atualizada, uma vez que os dados são descentralizados, produzidos por fontes variadas e em momentos diferentes. O governo tem tomado algumas medidas na direção da padronização, otimização de recursos e integração de dados, sendo uma delas a criação da Infraestrutura Nacional de Dados Espaciais do Brasil (INDE-BR).



A INDE-BR consiste em uma iniciativa de criação de uma arquitetura para disseminação e regramento do uso de dados geográficos no âmbito da administração pública. Cabe ressaltar um dos objetivos definidos na sua concepção, o de

“evitar a duplicidade de ações e o desperdício de recursos na obtenção de dados aeroespaciais pelos órgãos da administração pública, por meio da divulgação dos metadados relativos a esses dados disponíveis nas entidades e nos órgãos públicos das esferas federal, estadual, distrital e municipal” (BRASIL, 2008).

Camboim (2013) ressalta a iniciativa governamental de criação da Infraestrutura Nacional de Dados Abertos (INDA) e do Portal Brasileiro de Dados Abertos como uma estratégia de adoção de Dados Interligados Abertos (*Linked Open Data*). Esse autor propõe uma arquitetura capaz de tornar os dados da INDE-BR disponíveis na INDA de forma integrada. Essa proposta de arquitetura destaca-se pelo uso de camadas semânticas e ontologias, fazendo uma junção de técnicas da web semântica com dados geoespaciais, ingressando no campo da geosseântica.

As arquiteturas das IDE atuais são, via de regra, orientadas a serviço e adotam os padrões tecnológicos estabelecidos pela *Open Geospatial*

Consortium (OGC). Alguns desses padrões definem o funcionamento dos geosserviços, que permitem a obtenção de mapas cartográficos, mapas temáticos, dados geográficos e metadados.

Um dos componentes essenciais numa arquitetura de IDE é o catálogo de serviços. A metodologia SERVUS de desenvolvimento de sistemas (USLÄNDER, 2010), feita especialmente para tratar do uso e composição de serviços geoespaciais, propõe a utilização de um geocatálogo semântico para auxílio na localização de recursos geográficos em rede.

Para estudar esse aspecto, este documento foi dividido da seguinte forma: na Seção 2 são apresentados alguns trabalhos correlatos; na Seção 3 é apresentada a fundamentação teórica, tratando dos temas IDE e metodologia SERVUS; na Seção 4 é apresentado o modelo proposto, detalhes do geocatálogo e um exemplo de utilização com a metodologia SERVUS e, por fim, na Seção 5 temos as conclusões.

2. TRABALHOS CORRELATOS

Andrade e Baptista (2011) trilharam um caminho que incluía busca semântica, listagem de grande quantidade de informações oriundas de fontes diversas, estabelecimento de métricas e geração de um ranking de recursos disponibilizados numa

infraestrutura de dados espaciais. Os autores partiram do pressuposto de que o uso de ontologias seria o caminho para tornar mais precisas as buscas e facilitar o processo de automatização. Como contribuições, tivemos o estabelecimento de uma métrica combinada e a construção de uma rede semântica.

Daltio e Carvalho (2012) propuseram a construção de um *framework* para a recuperação semântica de dados espaciais. O *framework* era baseado em um processo de anotação semântica dos recursos geográficos e na utilização de um serviço de gerenciamento de ontologias. Cabe ressaltar neste trabalho a importância dada ao processo de seleção das ontologias.

Gimenez, Tanaka e Baião (2013) apresentam uma proposta de integração semântica para a INDE-BR usando geo-ontologias de domínio incorporadas em camadas semânticas atuando sobre as IDEs.

3. FUNDAMENTAÇÃO TEÓRICA

3.1 INFRAESTRUTURA DE DADOS ESPACIAIS (IDE)

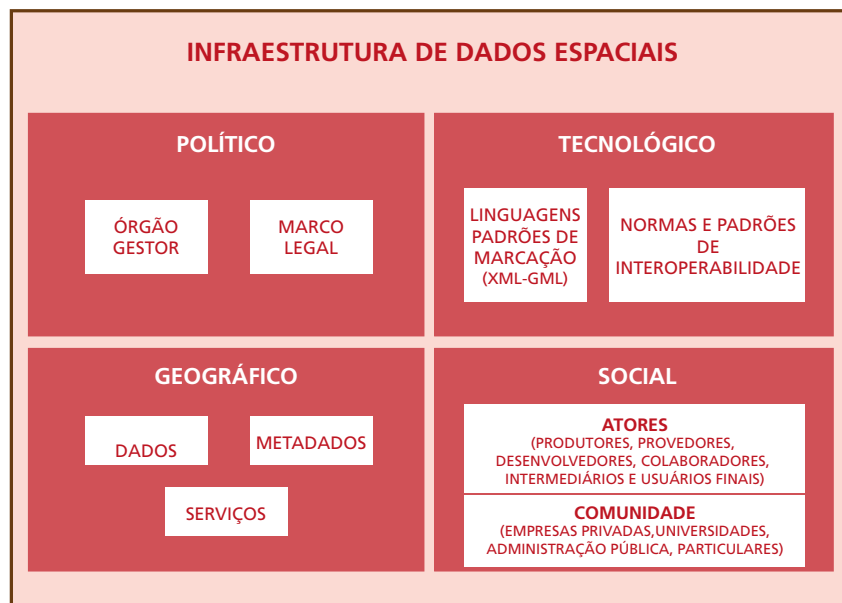
Bernabé-Poveda e López Vázquez (2012, p. 57-59) conceituam Infraestrutura de Dados Espaciais (IDE) como uma infraestrutura necessária para o acesso, compartilhamento, troca, combinação e

análise dos dados geográficos, de uma forma padrão e interoperável. Também consideram a necessidade de que esses dados estejam disponíveis em rede, por meio de um conjunto de sistemas que utilizem protocolos e interfaces padrões, propiciando a criação de aplicações que possam ser vistas pelo usuário como um único sistema. IDE pode ser classificada como uma estrutura de componentes tecnológicos, geográficos, sociais e políticos conforme ilustrado na Figura 1.

O componente tecnológico é representado por uma arquitetura baseada em padrões de interoperabilidade e capaz de compartilhar dados e informações geográficas. As linguagens de marcação XML e GML têm um papel fundamental nesse componente. O componente social é representado por um conjunto de atores, dentre eles, produtores de dados, provedores de serviços, usuários, desenvolvedores de software e responsáveis por padrões e normas, além de uma grande comunidade formada por empresas privadas, governo, universidades e a sociedade em geral. O componente geográfico é representado pelos dados, seus metadados e os geosserviços. O componente político é representado pelas pessoas e pelo órgão responsável por estabelecer o marco regulatório e as regras de funcionamento.

Figura 1:

Componentes de uma IDE



OGC *Web Services Common Standard* é um padrão tecnológico que propõe uma interface comum para um conjunto de geosserviços: *Web Map Service* (WMS), *Web Feature Service* (WFS) e *Web Coverage Service* (WCS). Nessa proposta busca-se padronizar as formas de acesso, operações, parâmetros obrigatórios e opcionais, além de estruturas de dados, a fim de diminuir os esforços de interoperabilidade. Uma das operações definidas é a *GetCapabilities* que recupera os metadados referentes às capacidades dos geosserviços (GREENWOOD; WHITE-SIDE, 2010).

WMS consiste num serviço que produz mapas dinâmicos, em que as informações geográficas são organizadas em camadas (*layers*). O serviço WFS trabalha com entidades geográficas com dados discretos ou vetoriais em formato GML, representando atributos e geometrias. WCS é o serviço que dá suporte a recuperação de dados espaciais como coberturas.

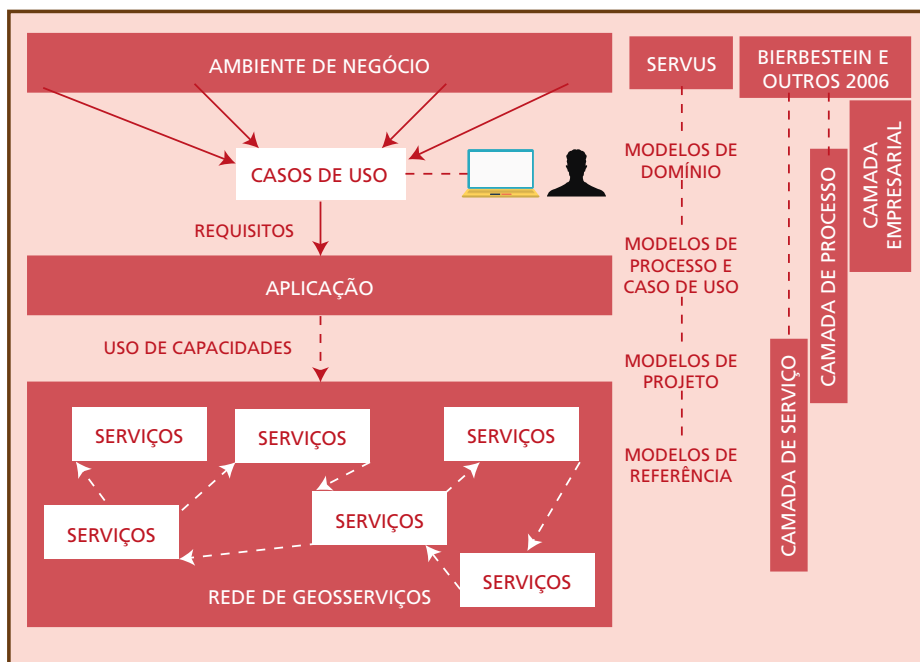
Um catálogo de serviço permite a localização de dados ou serviços geográficos através de uma gama de operações, dentre elas a operação *GetRecords* que recupera um conjunto de registros de metadados (OGC, 2016).

3.2 METODOLOGIA DE DESENVOLVIMENTO SERVUS

A sigla SERVUS é oriunda de dois termos da engenharia de software: “serviço” (*SERVice*) e “caso de uso” (*USE case*). O problema central que a metodologia se propõe a resolver é –considerando-se um conjunto de casos de uso que atendam aos requisitos de negócio e demandem capacidade de recursos em uma rede de serviços geoespaciais – como descobrir e associar recursos requisitados a recursos ofertados, compondo uma aplicação que supra os requisitos de qualidade, funcionais e informacionais. Dois outros elementos compõem esse cenário: descrições semânticas em uma rede de recursos semânticos e condições inerentes (*side conditions*) ao desenvolvimento de sistemas ambientais (processos de descoberta de recursos, uso de padrões OGC, combinações de recursos) (USLÄNDER, 2010).

Segundo Usländer (2010, p. 90), “SERVUS compreende o projeto de sistemas de informação ambientais como uma atividade iterativa de descoberta e associação: capacidades disponíveis são descobertas e associadas a requisitos do usuário formulados como casos de uso”.

Figura 2:
Hierarquia de modelos SERVUS



Fonte: Adaptado de Usländer (2010, p. 92)

A metodologia é composta por uma linguagem de modelagem, um processo de desenvolvimento e uma arquitetura de referência. Durante o processo de desenvolvimento são criados artefatos tais como: um modelo de domínio, um modelo de processos/casos de uso e um modelo de projeto. O modelo de projeto é o principal deles e expressa os requisitos e as capacidades sob a forma de recursos. Os modelos gerados são baseados nas camadas de abstração propostas por Bieberstein (2006). A Figura 2 exibe o mapeamento entre requisitos e capacidades, os modelos produzidos e a relação com as camadas de abstração.

O modelo de domínio da metodologia SERVUS representa o domínio temático do problema a ser resolvido. Ele formalmente define a parte do mundo que compreende o universo de discurso entre o usuário e o projetista do sistema, isto é, compreende o conhecimento compartilhado acerca do domínio da aplicação. Tipicamente esses conhecimentos compartilhados são representados pela especificação de uma ontologia (USLÄNDER, 2010).

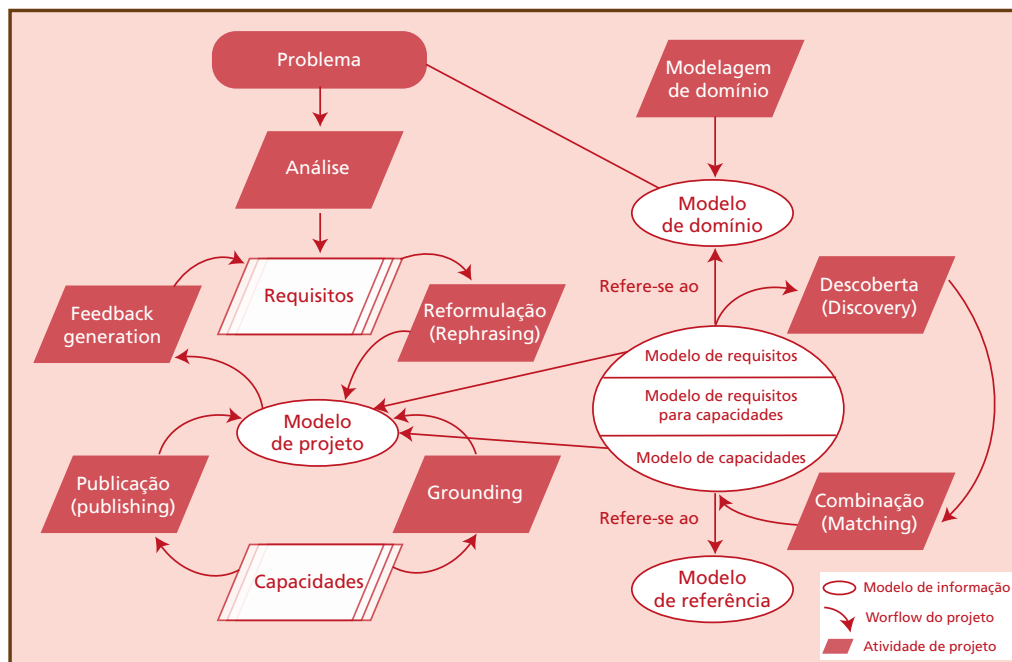
O modelo de referência é composto por um *framework* arquitetural, responsável pelas orientações e regras de como especificar o sistema, e por um modelo conceitual. O modelo conceitual de SERVUS é um metamodelo segundo os preceitos de *Model Driven Architecture* (MDA) e *Meta-Object*

Facility (MOF). As principais metaclasses são: feição (*feature*), *interface*, serviço e recurso. O modelo conceitual é composto por três subconjuntos de metaclasses, o primeiro está associado à visão de serviço, o segundo à visão de informação e o terceiro à visão de recurso. O modelo de projeto pode ser visto como uma composição de três modelos: o de requisitos (REQ's), o de capacidades (CAP's) e o de mapeamento entre os dois primeiros (REQ2CAP). A Figura 3 apresenta um quadro geral dos modelos e atividades da etapa de projeto.

Segue a seguir uma descrição das principais atividades do processo de desenvolvimento.

- a. **Publishing:** atividade, manual ou automatizada, de busca de capacidades na rede de serviços geoespaciais e de incorporação ao modelo de capacidades. As capacidades são traduzidas para recursos ofertados, ou seja, disponíveis na rede de geosserviços;
- b. **Rephrasing:** atividade responsável por traduzir os requisitos, de forma a extrair um conjunto de recursos necessários para realização dos casos de uso, os recursos podem ser associados aos conceitos de ontologia e em seguida carregados na rede de recursos semânticos, compondo o modelo de requisitos;

Figura 3: Modelos e atividades de projeto



Fonte: Adaptado de Usländer (2010, p. 99)

- c. **Discovery:** atividade responsável por selecionar, no rol de capacidades levantadas na fase publishing, um conjunto que atenda aos recursos solicitados levantados na fase de rephrasing, para cada recurso requisitado pode existir um conjunto de recursos ofertados que atendam às necessidades;
- d. **Matching:** atividade responsável por associar requisitos a capacidades, ou seja, recursos requisitados a recursos ofertados, avalia dentre as capacidades candidatas, elencadas na etapa de discovery, as mais adequadas aos requisitos. O resultado final dessa etapa é o modelo de mapeamento requisitos/capacidades;
- e. **Grounding:** atividade que disponibiliza uma nova capacidade na rede de serviços geoespaciais.

SERVUS propõe a criação de um catálogo semântico que compõe uma arquitetura de implementação e um ambiente de projeto. O papel do catálogo semântico é o de se comunicar com os geosserviços da rede e processar as consultas por recursos. Funciona como uma extensão semântica que possibilita consultas baseadas em ontologias e a avaliação da proximidade semântica dos resultados obtidos. No

que se refere ao suporte à metodologia, o catálogo semântico responde pelas atividades de *harvesting*, que consiste na coleta de metadados dos serviços, e de *publishing*, publicação dos recursos disponíveis para uma rede de recursos semânticos.

6. MODELO PROPOSTO

O trabalho foi desenvolvido conforme as seguintes etapas: revisão de literatura, com a realização de estudos sobre IDEs, metodologia de desenvolvimento SOA, metodologia SERVUS, banco de dados orientado a grafos, geocatálogo e busca semântica; levantamento de requisitos baseado nos geocatálogos estudados, principalmente no proposto pela metodologia SERVUS, nas necessidades dos *stakeholders* da área de meio ambiente e na infraestrutura tecnológica atual do órgão; definição da arquitetura de alto nível, considerando-se, dentre outros aspectos, fontes de dados, arquiteturas de IDEs e geosserviços da INDE-BR; desenvolvimento do protótipo, incluindo a definição das linguagens de desenvolvimentos, dos componentes, das bibliotecas de software e do ambiente de desenvolvimento; implantação com definição do ambiente e instalação do protótipo; elaboração de um cenário de utilização do geocatálogo típico de auditoria em meio ambiente, validação do protótipo e, por fim, avaliação do resultados (SANTOS, 2016).

Figura 4:
Arquitetura de implementação

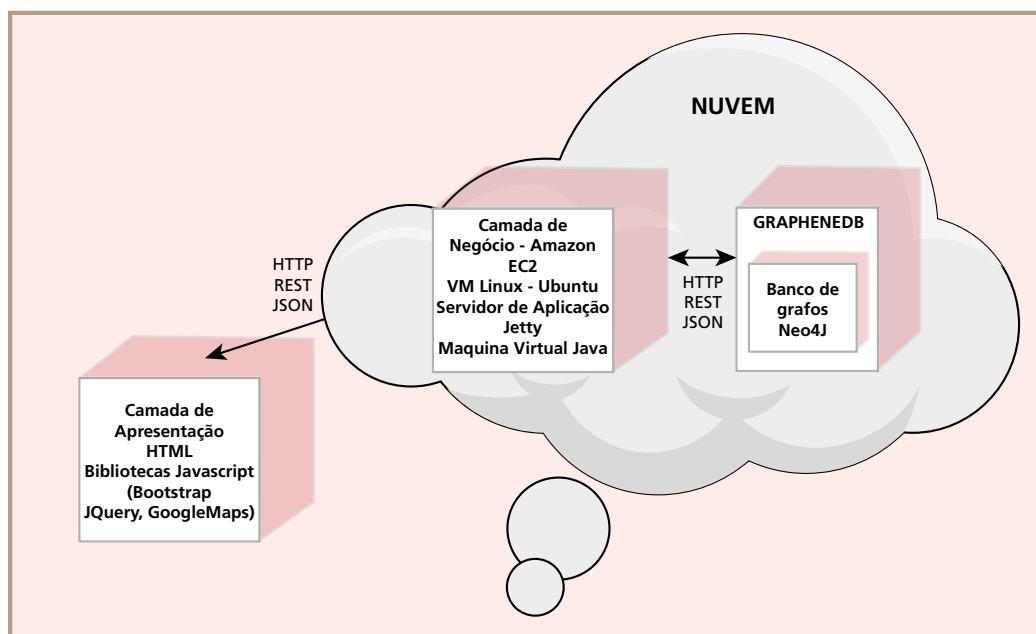
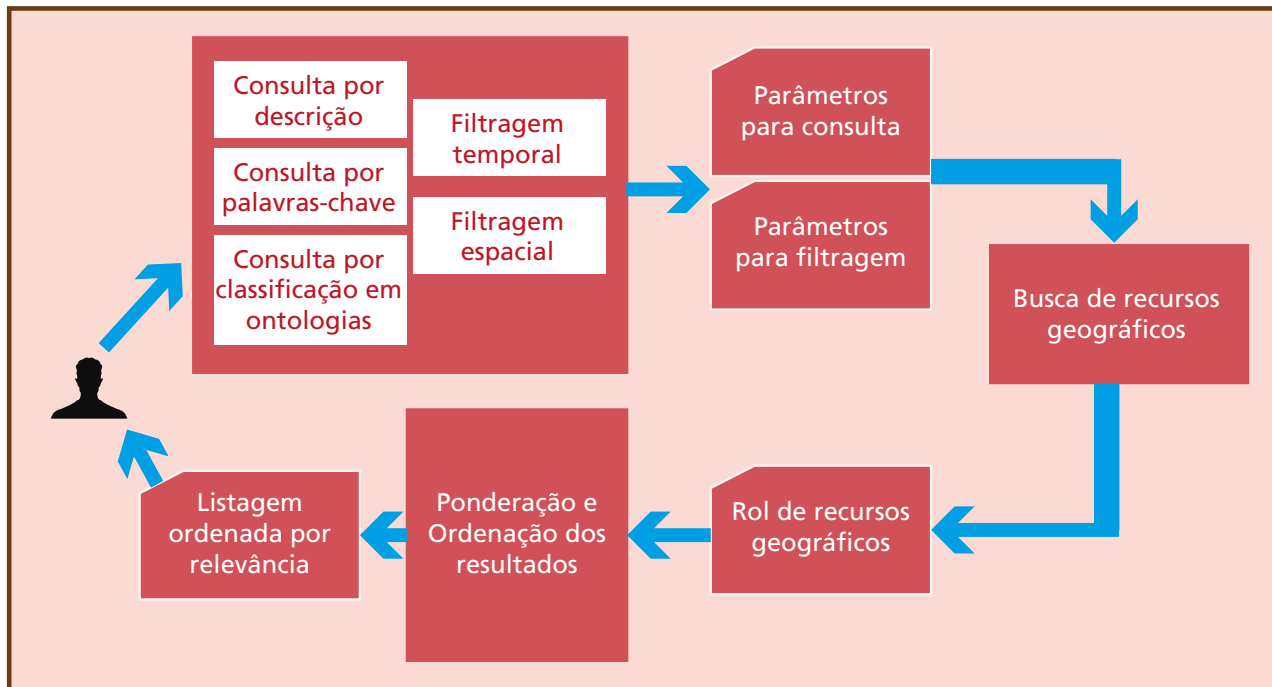


Figura 5:

Modelo de busca do catálogo



A arquitetura escolhida para implementação do protótipo do geocatálogo foi a de uma aplicação web de três camadas. A camada de apresentação foi desenvolvida em HTML5 utilizando bibliotecas *JavaScript: Bootstrap 3.3.5, jQuery 2.1.0, Google Maps 3.0* e o componente *bootstrap-slider.js*. A camada de negócio foi implementada em Java versão 1.7 e a aplicação empacotada no formato de *web archive – WAR*. As requisições da camada de apresentação são feitas pelo protocolo *http* e acessam um servidor padrão REST (*Representa-*

tional State Transfer) que retorna documentos no formato JSON (*JavaScript Object Notation*). A camada de negócio foi hospedada num serviço de nuvem EC2 da *Amazon*. A camada de persistência foi implementada através de um banco de dados NoSQL de grafos, o *Neo4J 2.3.1*. A hospedagem da camada de persistência foi realizada na página web *GrapheneDB*, que implementa os bancos *Neo4J* num ambiente de nuvem. A comunicação entre camada de negócio e camada de persistência também ocorre por meio de interfaces REST e no

Tabela 1:

Cálculo do índice de relevância

Termos de pesquisa	Campo com termo localizado	Peso	Peso Máximo
Descrição: "vegetação" - (peso 8)	Nome - (peso 8)	64	216
Ontologia VCE (termo raiz): "vegetação" - (peso 6)	Descrição - (peso 6)	36	162
Palavra-chave: "agrícola" - (peso 5)	Resumo - (peso 5)	25	135
	Peso total	125	513
	Índice de relevância calculado (125/513)	0,24	

formato JSON. A Figura 4 representa a arquitetura implementada.

As consultas podem ser realizadas por descrição (texto livre), palavras-chave e por termos das ontologias. Para cada um dos itens anteriores as frases ou termos são divididos em *tokens* que serão os termos utilizados na busca. Podem ser realizadas filtragens por coordenadas geográficas (filtro espacial) e por ano (temporal). Para cada par ordenado, composto pelo termo e o campo pesquisado, é atribuído um peso que será a base para o procedi-

mento de geração do ranking. A Figura 5 apresenta o modelo de busca do geocatálogo.

O resultado efetivo de uma busca depende da qualidade dos metadados e do valor de negócio atribuído a cada campo. Para lidar com esse aspecto foi elaborado um esquema de ponderação dos campos de pesquisa e dos campos de metadados e o estabelecimento de um índice de relevância que serve de base para o processo de geração do ranking de resultados. A Tabela 1 apresenta um exemplo de cálculo do índice de relevância.

Figura 6:
Tela de busca do geocatálogo

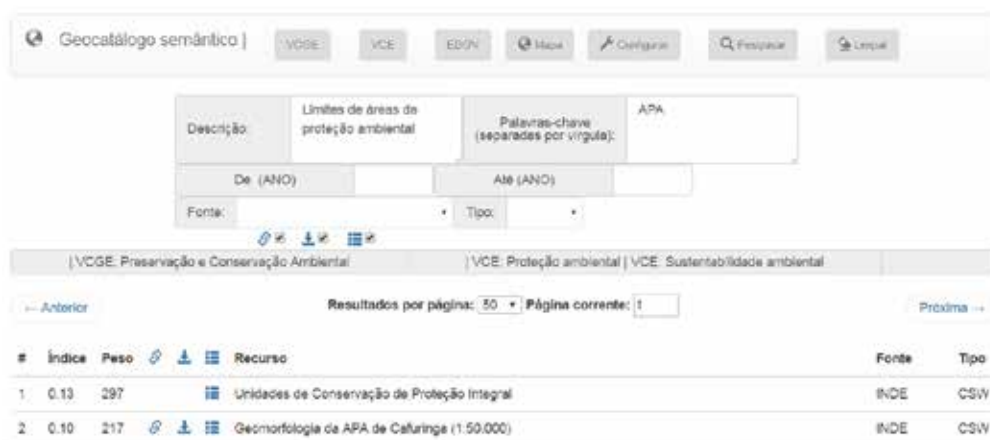
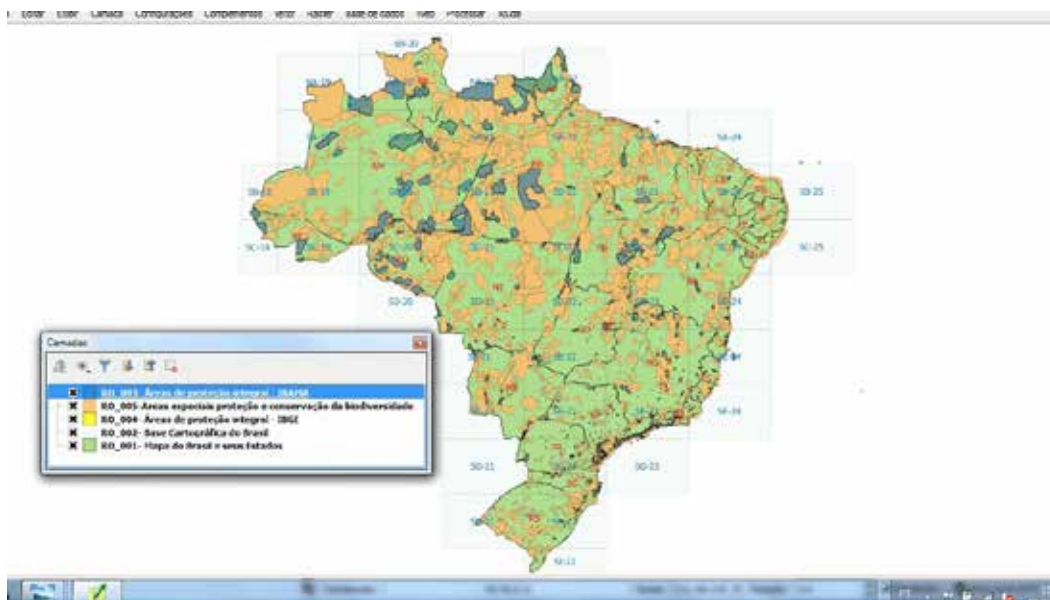


Figura 7:
Áreas de proteção ambiental importadas para ferramenta QGIS



Os registros dos recursos são recuperados mediante requisições REST para o banco de dados remoto, compostas pelos comandos na linguagem *Cypher*. Os principais tipos de interação são: consulta a ontologias, criação de relacionamentos entre os recursos e geração do ranking dos recursos geográficos. A Figura 6 exibe o ranking na forma de uma listagem com os links para visualização do recurso, download e visualização dos metadados.

Para validação do uso do geocatálogo foram utilizadas as etapas da metodologia SERVUS, utilizamos como base um problema relacionado a uma auditoria na área ambiental, que assim apresentamos: como obter os dados necessários, a nível brasileiro, para o levantamento da localização de áreas de proteção ambiental, utilizando como base os geosserviços disponíveis na rede de infraestrutura de dados espaciais nacional, de instituições e órgãos públicos?

O resultado final foi a localização e visualização dessas áreas. A Figura 7 apresenta as áreas de proteção ambiental localizadas pelo geocatálogo, cujos dados foram baixados de serviços WFS e importados para a ferramenta Quantum GIS.

7. CONCLUSÃO

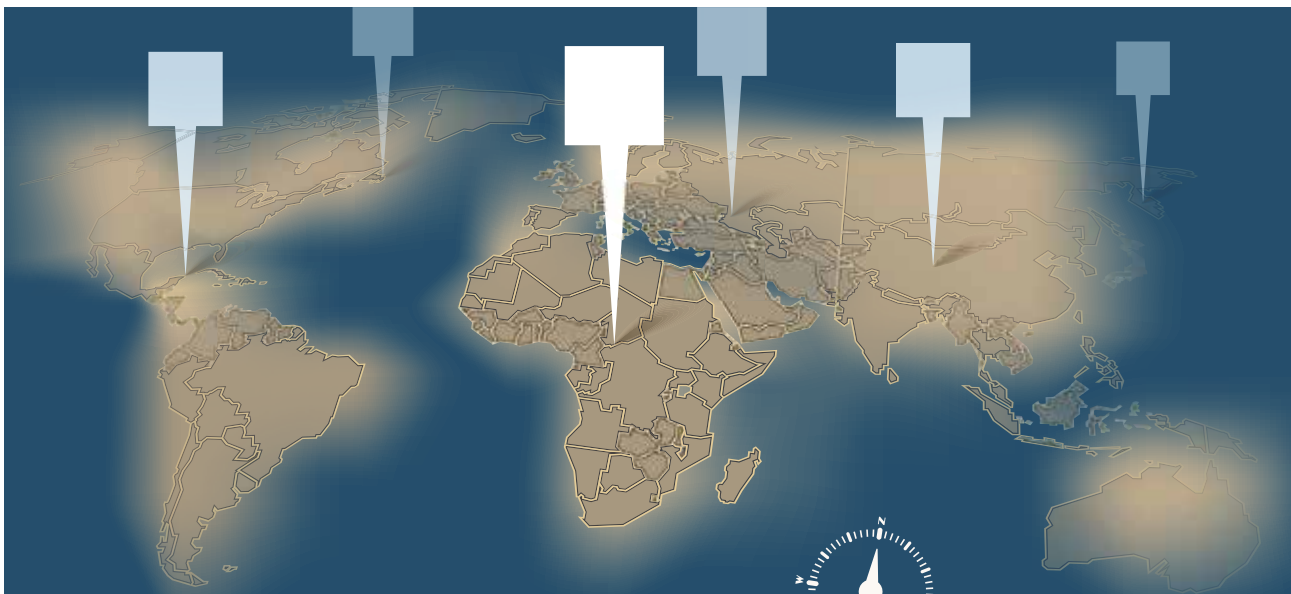
Este artigo abordou aspectos da implementação de um geocatálogo que possibilita a utilização de ontologias como parâmetros de pesquisa. As consultas são realizadas nos metadados de geosserviços de infraestruturas de dados espaciais. Foram aborda-

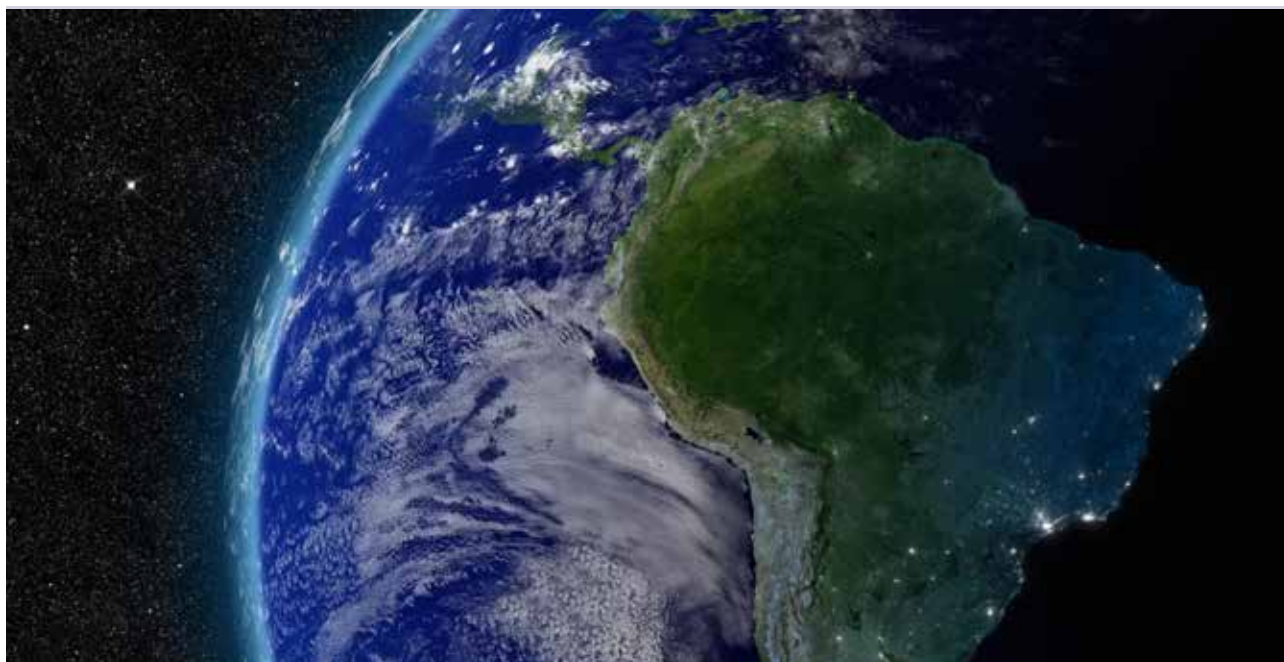
dos aspectos da implantação tais como: arquitetura e modelo de dados. No que se refere à arquitetura, o geocatálogo foi implementado como uma aplicação web de três camadas, passível de implantação em ambiente de nuvem. O modelo de dados foi implementado em banco de dados orientado a grafos.

O geocatálogo se mostrou viável como ferramenta de busca de recursos geográficos e de apoio à metodologia SERVUS, possibilitando a atuação nas etapas de *publishing* e *matching*. Podem ser consideradas como contribuições a utilização da linguagem *Cypher* no manuseio de rede de recursos, ontologias e metamodelo, de forma integrada, em banco de dados não relacionais orientado a grafos. O estabelecimento de mecanismos de parametrização das consultas, mediante ponderação de parâmetros e campos de metadados e o estabelecimento de um índice de relevância, são contribuições para o campo da busca em bases de metadados. Como trabalhos futuros, podemos considerar a automatização do processo *harvesting* e a integração com uma infraestrutura de dados espaciais corporativa, de forma que o geocatálogo atue conjuntamente com os geoportais.

REFERÊNCIAS

ANDRADE, F. G.; BAPTISTA, C. S. Using semantic similarity to improve information discovery in spatial data infrastructures. *Journal of Information and Data Management*, v. 2, n. 2, p. 181, 2011.





BERNABÉ-POVEDA, M. Á.; LÓPEZ-VÁZQUEZ, C. M. (Eds.). Fundamentos de las infraestructuras de datos espaciales (IDE). Madrid: Universidad Politécnica de Madrid, 2012. 593p.

BIEBERSTEIN, N. Service-oriented architecture compass: business value, planning, and enterprise roadmap. Indiana: FT Press, 2006.

BRASIL. Decreto n. 6.666, de 27 de novembro de 2008. Institui, no âmbito do Poder Executivo Federal, a Infraestrutura Nacional de Dados Espaciais – INDE, e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm>. Acesso em: 22 out. 2016.

CAMBOIM, S. F. Arquitetura para integração de dados interligados abertos à INDE-BR. 2013. 140f. Tese (Doutorado em Ciências Geodésicas) – Universidade Federal do Paraná, Curitiba, 2013.

DALTIO, J.; CARVALHO, C. A. Um framework para recuperação semântica de dados espaciais. In: BRAZILIAN SYMPOSIUM ON GEOINFORMATICS, 13, 2012, Campos do Jordão. Proceedings... São José dos Campos: MCTI/INPE, 2012. p. 60-65.

GIMENEZ, P. J. A.; TANAKA, A. K; BAIÃO, F. Uma proposta de integração semântica para a Infraestrutura Nacional de Dados Espaciais usando geo-ontologias. In: WORKSHOP DE COMPUTAÇÃO APLICADA AO GOVERNO ELETRÔNICO (WCGE), 2013, João Pessoa. Anais... João Pessoa: UFPB, 2013. v. 5. p. 25-32.

GREENWOOD, J.; WHITESIDE, A. OGC Web Services Common Standard. 2010. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.248.6351&rep=rep1&type=pdf>>. Acesso em: 22 out. 2016.

OGC. Catalogue Services 3.0 Specification: HTTP Protocol Binding. 2010. Disponível em: <<http://docs.opengeospatial.org/is/12-176r7/12-176r7.html>>. Acesso em: 22 out. 2016.

SANTOS, D. Implementação de um geocatálogo utilizando banco de dados orientado a grafos para apoio à metodologia SERVUS. 2016. 82f. Dissertação (Mestrado em Computação Aplicada) – Universidade de Brasília, Brasília, 2016. Disponível em: <http://alexandre.zaghetto.com/wp-content/uploads/2016/07/Disserta%C3%A7%C3%A3o_Vers%C3%A3o_FinalParaE-di%C3%A7%C3%A3o.pdf>. Acesso em: 22 out. 2016.

USLÄNDER, T. Service-oriented design of environmental information systems. Vol. 5. Karlsruhe: KIT Scientific, 2010.