

Tratamento de Inferência em Banco de Dados Ecológicos

Maira Poltosi¹, Daniele Palazzi¹, Ana Maria de C. Moura¹,
Fabio Porto¹, Frederic Valentin¹, Jean Louis Valentin²

¹Extreme Data Lab (DEXL Lab)
Laboratório Nacional de Computação Científica (LNCC)
Petrópolis – RJ – Brasil

²Departamento de Biologia Marinha – Instituto de Biologia
Universidade Federal do Rio de Janeiro (UFRJ) – Rio de Janeiro, RJ – Brasil
{maira, dpalazzi, anamoura, fporto, valentin}@lncc.br,
jlv@biologia.ufrj.br

Abstract. *This paper presents the development of a database to support ecological data management and treatment. This database, developed in the context of the PELD project, is provided with inference capability, achieved from the relational model mapping to RDF. Taxonomical, spatial and trophic relations are explored by means of nested sets and explicit transitive-closure, which enable retrieving interesting information about existing marine species of Guanabara bay, illustrated by SPARQL queries.*

Resumo. *Este artigo apresenta o desenvolvimento de um banco de dados para auxiliar no gerenciamento e tratamento de dados ecológicos. Este banco, desenvolvido no contexto do projeto PELD, é dotado da capacidade de inferência, fruto de um mapeamento do modelo relacional para RDF. Relações taxonômicas, espaciais e de cadeia trófica são exploradas por meio de modelos de conjuntos aninhados e de fecho-transitivo explícito, que permitem a descoberta de informações interessantes sobre espécies marinhas da baía da Guanabara, ilustradas por consultas SPARQL.*

1. Introdução

A forte preocupação com a preservação do meio-ambiente no planeta e a necessidade de organizar e catalogar as reservas naturais existentes, tanto minerais como biológicas, a exemplo fauna, flora e recursos hídricos, tem dado origem a diversos sistemas para o repertoriamento e controle destas informações.

As descobertas observadas e relatadas em eventos científicos (ISEST¹, ECOTHEE², etc.) e publicações (Nature, Biotropica, Ecological Informatics, etc.) no domínio da Ecologia, juntamente com os avanços tecnológicos na área da Ciência da Computação, vêm contribuir significativamente para o desenvolvimento de sistemas de software adequados para lidar com dados de natureza ecológica e espaço-temporais, abrangendo os mais diversos aspectos como: modelagem, metadados, ontologias, estruturas multi-dimensionais, armazenamento, acesso, distribuição, mineração, visualização, interfaces gráficas, interoperabilidade, consultas, etc.. Segundo [Manzi 2009], os sistemas de informação para dados ecológicos ainda se apresentam em estágios iniciais de desenvolvimento e enfrentam alguns desafios técnicos, como o de criar

1 2011 International Symposium on Environmental Science and Technology: <http://www.isest11.com/>

2 2nd International Conference on ECOLOGICAL THEOLOGY AND ENVIRONMENTAL ETHICS: <http://www.conferencealerts.com/seeconf.mv?q=ca1i0xh6>

estruturas de dados verdadeiramente genéricas, podendo acomodar dados de qualquer tipo de levantamento.

No Brasil, encontra-se em desenvolvimento o projeto PELD³ (Pesquisas Ecológicas de Longa Duração) que tem entre as principais missões o avanço no conhecimento ecológico de forma a fornecer dados para embasar decisões relacionadas ao gerenciamento de recursos naturais além de disponibilizar estas informações aos diversos setores da sociedade. O PELD atualmente possui 29 sítios de coletas distribuídos em diversos biomas do Brasil, a fim de consolidar o conhecimento existente sobre a composição e funcionamento destes ecossistemas. Os dados provenientes das pesquisas do sítio PELD Amazônia estão sendo migrados para o padrão de metadados⁴ EML, adotado pelas redes de pesquisa mundiais LTER e ILTER, prevendo a futura integração de dados entre os sítios PELD.

Entre estes, está em desenvolvimento o PELD Guanabara⁵, que visa melhorar o entendimento da estrutura e do funcionamento do ecossistema da Baía de Guanabara e de suas respostas aos impactos antrópicos e climáticos. Além disso, esse projeto permitirá avaliar a adequação das medidas que estão sendo tomadas para a despoluição da Baía de Guanabara, no que diz respeito à ecologia marinha. De forma a viabilizar um estudo contínuo de longo prazo, e de prover meios computacionais que permitam realizar um prognóstico da tendência evolutiva desse ecossistema, está em desenvolvimento o projeto de um banco de dados para a organização, gerenciamento e recuperação de informações de dados ecológicos da Baía de Guanabara.

A grande contribuição desse trabalho em relação a outros sistemas de gerenciamento de dados ecológicos existentes (Ecoflora, [Manzi 2009], [Cavalcanti 2005]) é dotá-lo da capacidade de raciocínio. Por meio de uma representação ontológica do modelo, novos relacionamentos e instâncias podem ser inferidos a partir de propriedades transitivas e de hierarquias sobre os conceitos do modelo, oferecendo aos pesquisadores da área a descoberta de dados interessantes, tais como, os predadores, de diferentes níveis, de uma determinada espécie.

O restante desse trabalho está organizado da seguinte forma: a seção 2 apresenta um panorama geral sobre os trabalhos relacionados na área. A descrição do cenário de aplicação está na seção 3. A seção 4 formaliza e descreve o processo de mapeamento do modelo relacional para a representação RDF. Na seção 5 encontram-se exemplos de consultas à base realizadas em SPARQL. Finalmente, a seção 6 finaliza o artigo, com sugestões para trabalhos futuros.

2. Trabalhos relacionados

Vários trabalhos têm sido propostos com a finalidade de gerenciar e compartilhar dados ecológicos. Em especial, no contexto do projeto PELD [Manzi 2009], é proposta a criação de um banco de dados (BD) para gerenciamento de dados biológicos e ambientais georeferenciados, de modo a prover a recuperação simples e avançada de dados, incluindo a análise espacial dos mesmos.

Ecoflora⁶ é um BD ecológico sobre 2200 espécies de plantas das ilhas britânicas, construído a partir de inúmeras fontes e classificadas segundo uma taxonomia envolvendo 130 características ecológicas e morfológicas contendo informações sobre família, gênero, nome específico, autor, sinônimos, etc. Os dados incluem ainda informações sobre espécies distribuídas no Reino Unido e demais países da Europa, além de informações sobre insetos e doenças provenientes de fungos, os quais são obtidos da literatura.

Em [Cavalcanti 2005] é descrito um sistema de BD relacional para o gerenciamento e análise dos dados coletados da Reserva de Desenvolvimento Sustentável do Lago Tupé. O

3 <http://ppbio.inpa.gov.br/Port/projetosassociados/peld/>

4 <http://knb.ecoinformatics.org/community.jsp>

5 <http://www.lncc.br/peldguanabara/index.php>

6 <http://www.ecoflora.co.uk/>

objetivo do banco é a sua integração com um sistema de informação distribuído sobre biodiversidade, que possa fornecer apoio às atividades de pesquisa e extensão em andamento na área. Dados específicos sobre inventários biológicos são coletados dos sistemas ECOLOG [Cavalcanti 1993], que contém dados sobre espécies, indivíduos, descritores morfológicos e variáveis ambientais; e ALINE [Cavalcanti 1998], que reúne informações sobre flora e a fauna da área de estudo, incluindo dados sobre taxonomia e nomenclatura, usos econômicos, nomes vulgares, ecologia, morfologia, distribuição geográfica e bibliografia relevante para cada espécie. Além disto, conta com o sistema DELTA [Cavalcanti 1996] que permite codificar e processar a informação taxonômica sobre os dados coletados, que posteriormente podem ser analisados com o auxílio de pacotes estatísticos.

Em [Campos et al 2009] é proposta uma modelagem para dados de zoneamento ecológico econômico do estado de Minas Gerais, com o objetivo posterior de criar um BD. Nesse banco serão armazenadas características ecológicas envolvendo temas mais genéricos tais como zonas, riscos ambientais, prioridade de conservação e prioridade de recuperação, todos geo-referenciados, mas que são a seguir expandidos em temas mais específicos, envolvendo Vulnerabilidade Natural, que é composta pelos subtemas Biótico e Abiótico.

Dentre os sistemas analisados, observou-se que a grande preocupação dos mesmos refere-se ao armazenamento e recuperação dos dados ecológicos. Interfaces flexíveis e amigáveis permitem consultar informações, cujos tópicos são indexados por categorias com base em taxonomias. No entanto, nenhum deles explora a recuperação de informações implícitas que podem ser obtidas através de um tratamento semântico e de inferência sobre os dados ali representados. O aspecto de inferência é sem dúvida uma questão interessante na análise das informações ecológicas do ponto de vista dos pesquisadores. Quando essas informações são descritas e representadas por uma linguagem ontológica, o mecanismo de raciocínio ali embutido pode trazer contribuições relevantes para área, como mostram alguns trabalhos que lidam com sistemas de anotação de documentos em bio-informática [Fontes 2011], área geo-espacial [Macário e Medeiros 2009], [Macário et al 2009] e de meio ambiente [Crivellini et al 2009].

3. Cenário da Aplicação PELD Guanabara

O PELD Guanabara visa obter dados bióticos e abióticos a partir de amostras extraídas da água da baía e de comunidades planctônicas e bentônicas, assim como de recursos pesqueiros. Este estudo utiliza um modelo parcial considerando apenas os aspectos relativos ao acompanhamento de biomassa de plancton.

De uma amostra de plancton são armazenadas informações temporais (data e hora), espaciais (latitude, longitude e profundidade), método de coleta da amostra e de conservação utilizados, assim como as condições atmosféricas e marítimas no momento da coleta. Para cada análise realizada, são registradas a data, amostra e o método aplicado.

As medições de biomassa dos organismos encontrados nas amostras podem ser realizadas em nível de espécie ou em nível mais alto da taxonomia. A taxonomia classifica os organismos em uma hierarquia, cujo primeiro nível é Reino, que é dividido em Filos e sucessivamente em classes, ordens, famílias, gêneros e espécies. Cada nível destes tem, por sua vez, suas respectivas subdivisões. Qualquer nível desta classificação é chamada taxon e tem como plural taxa. Há diferenças nos níveis preenchidos de um e outro organismo. Alguns organismos têm sido reclassificados, e neste caso, ambas classificações são mantidas, sendo estabelecida uma relação de sinônimos entre elas.

Do ponto de vista de análise, o BD ecológico apresenta algumas dimensões importantes. A informação de região geográfica permite selecionar e classificar eventos segundo sua localização. Por outro lado, as relações tróficas são fundamentais para o estudo do ecossistema. Finalmente a taxonomia permite uma análise hierarquizada das espécies. O ambiente de

inferência apresentado na seção a seguir explora de forma integrada estratégia de banco de dados e técnicas da Web Semântica possibilitando estas análises.

4. Mapeamento do modelo relacional para representação RDF

A disponibilização em um formato homogêneo dos dados produzidos pelos projetos PELD é um fator fundamental na formação de uma visão integrada dos ecossistemas brasileiros.

Neste contexto, vêm se utilizando cada vez mais do RDF⁷ (Resource Description Framework) como meio de interoperar, inferir e publicar dados, além de possibilitar que estes possam posteriormente se conectar a outros recursos de domínios similares. Permite ainda a exploração de dados e associações utilizando-se da linguagem de consulta SPARQL⁸. Optou-se pela representação do modelo através de assertivas RDF para incorporar expressividade semântica na camada de dados.

A utilização de RDF pode trazer, todavia, uma preocupação com o desempenho no tratamento de grandes volumes de dados, como é o caso do banco de dados ecológico. Nestas situações, uma estratégia comumente adotada é armazenar os dados em bancos de dados relacionais.

Deste modo, o banco de dados do PELD Guanabara está sendo armazenado no sistema PostgreSQL v. 8.4.5. e é mapeado para um grafo virtual RDF com auxílio da ferramenta D2RQ⁹, uma plataforma desenvolvida em Java que permite executar consultas SPARQL e acessar a informação do BD como *LinkedData*¹⁰ sobre a Web. Esta plataforma é constituída pelos seguintes componentes: D2RQ Mapping Language, linguagem declarativa que descreve o mapeamento; D2R Server, servidor HTTP que fornece uma visão Linked Data e permite consultas SPARQL; e D2RQ Engine, um plug-in para uso de raciocinadores Jena¹¹ e Sesame¹². A Figura 1 apresenta a representação em RDF do BD PELD Guanabara.

No processo de mapeamento é criado um arquivo padrão que analisa o esquema do BD. Todas as tabelas do modelo relacional são então transformadas em classes no grafo RDF virtual e todas as colunas em propriedades. Este arquivo pode ainda ser personalizado utilizando vocabulários conhecidos, como por exemplo, rdf, rdfs, foaf, dc, owl entre outros. Uma vez gerado o arquivo de mapeamento, é possível navegar sobre o grafo RDF, através do servidor D2R Server.

4.1. Formalização da arquitetura

Define-se um BD relacional como $BD = R \cup C$, onde $R = \{R_1, R_2, \dots, R_n\}$ é um conjunto de relações, de esquema $E = \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_n\}$, e $C = \{C_1, C_2, \dots, C_m\}$ é um conjunto de restrições do tipo chave primária e chave estrangeira.

Um domínio de aplicação pode ser modelado em RDF segundo um conjunto de assertivas do tipo $\langle s, p, o \rangle$, onde um predicado p descreve um recurso s , segundo um objeto o . Definimos $s \subseteq URI$, $p \subseteq URI$ e $o \subseteq URI \cup L$, onde URI é um conjunto de referências à *Unified Resource Identifiers*¹¹, identificando recursos do domínio, e L é um conjunto de literais. O conjunto A de assertivas RDF é chamado de grafo RDF, onde $A = (N, E)$, e $N \subseteq S \cup O$ e $E \subseteq P$, onde S é o conjunto de recursos, O é o conjunto de objetos e P é o conjunto de predicados. Recursos definidos em S ou em O podem ser agrupados em tipos através do predicado $rdf:type$, que associa recursos a Classes do domínio. Adicionalmente, predicados em P são instâncias do tipo $rdf:Property$.

7 www.w3.org/RDF

8 www.w3.org/TR/rdf-sparql-query/

9 <http://www4.wiwiw.fu-berlin.de/bizer/d2rq/spec/>

10 <http://linkeddata.org/>

11 <http://jena.sourceforge.net/>

12 <http://www.openrdf.org/>

4.3. Modelo de conjuntos aninhados

No modelo de conjuntos aninhados, considera-se o conjunto Is de instâncias de uma relação R , e uma função booleana $\acute{e}_pai(R, R)$, tal que $\forall i_i e i_j \in Is, \acute{e}_pai(i_i, i_j)$ se e somente se, i_i tem uma chave estrangeira para i_j [Navathe 2005].

A implementação do modelo de conjuntos aninhados é realizada construindo-se uma árvore que obedeça a relação $\acute{e}_pai(R,R)$ para todas as instâncias de R . Cada nó nesta árvore representa uma instância dessa relação. Processa-se a árvore em pré-ordem. A cada nó associam-se dois valores, esquerda e direita. O valor esquerda corresponde à posição em que o elemento foi lido. Adicionalmente, uma vez que todos os filhos de um nó tenham sido visitados, associa-se o valor *direita*, correspondendo à posição seguinte a de seu descendente mais à direita. A navegação continua em um irmão, a partir do valor à direita recém associado.

Como resultado do processamento acima descrito, adiciona-se ao esquema da relação R os atributos esquerda e direita. Assim, dada duas instâncias $i_i e i_j \in R, \acute{e}_pai(i_i, i_j)$ se e somente se $i_i.esquerda > i_j.esquerda$ e $i_i.direita > i_j.direita$, e não existe $i_k \in R, i_k \neq i_j$, tal que $i_i.esquerda > i_k.esquerda$ e $i_i.direita > i_k.direita$.

As seguintes relações foram implementadas utilizando o modelo de conjuntos aninhados para representação do fecho transitivo:

```
Região (idRegiao, ..., Esquerda, Direita)
Taxon (idTaxon, ..., Esquerda, Direita)
```

4.4. Modelo de fecho transitivo explícito

O relacionamento transitivo expresso pela associação \acute{e}_presa produz uma ordenação em grafo, impedindo sua implementação através do modelo de conjuntos aninhados. Nesta implementação decidiu-se explicitar todo o fecho transitivo. Define-se uma relação $R'(R,R)$, tal que, $\forall i_i, i_j \in R$, se $\acute{e}_presa(i_i, i_j)$ então existe $i_k \in R'$ e $i_k < i_i, i_j >$.

A relação Cadeia_trofica foi implementadas utilizando o modelo de fecho transitivo explícito:

```
Cadeia_trofica (idCadeia_trofica, idTaxon_presa, idTaxon_predador)
```

5. Consultas em SPARQL

A partir da modelagem apresentada anteriormente, algumas consultas SPARQL foram implementadas, com seus respectivos resultados. Elas visam a análise dos taxa sobre a região em que foram coletados, relacionando-os à cadeia trófica e aos sinônimos entre eles.

5.1. Consultas com inferência sobre BD PELD Guanabara

As consultas a seguir foram implementadas utilizando SPARQL, cujo código é apresentado na Figura 2. Os resultados correspondentes são listados na Figura 3.

- (i) Quais são os animais que correspondem aos predadores do peixe “Manjuba” na cadeia trófica?
- (ii) Obter as presas do taxon “Bagre”, incluindo os taxa sinônimos.
- (iii) Quais os taxa subníveis de um dado taxon? Recuperar a sub-árvore a partir do taxon informado, indicando o nível da hierarquia taxonômica.
- (iv) A ocorrência de um acidente ecológico em uma dada região pode causar a morte de plancton que são componentes dos primeiros níveis da cadeia trófica. É possível identificar quais os predadores das espécies de plancton encontrados nesta região para dimensionar o impacto deste evento.

<pre>(i) SELECT DISTINCT ?nome_pres ?nome_predador ?grau WHERE { ?s <URI/taxon_dstaxon_comum> "Manjuba"; <URI/taxon_dstaxon_comum> ?nome_pres. ?w <URI/cad_trof_Taxon_idTaxon_predador> ?pred. ?pred <URI/taxon_dstaxon_comum> ?nome_predador. ?resource <URI/cad_trof_inGrau_relacao> ?grau ; <URI/cad_trof_Taxon_idTaxon_presa> ?e. FILTER (?w = ?resource && ?s = ?e) } ORDER BY ?grau ?nome_pred</pre>	<pre>(ii) SELECT DISTINCT ?nome_predador ?nome_presa ?sinonimo WHERE { {?s <URI/taxon_dstaxon_comum> "Bagre". ?w <URI/cad_trof_Taxon_idTaxon_predador> ?s. ?s <URI/taxon_dstaxon_comum> ?nome_predador. ?w <URI/cad_trof_Taxon_idTaxon_presa> ?presa. ?presa <URI/taxon_dstaxon_cientifica> ?nome_presa. }UNION{ ?s <URI/taxon_dstaxon_comum> "Bagre". ?w <URI/cad_trof_Taxon_idTaxon_predador> ?s. ?s <URI/taxon_dstaxon_comum> ?nome_predador. ?w <URI/cad_trof_Taxon_idTaxon_presa> ?presa. ?presa <URI/taxon_dstaxon_cientifica> ?nome_presa. ?id_sinonimo <URI/taxon_taxon_idtaxon_sinonimo> ?presa. ?id_sinonimo <URI/taxon_dstaxon_cientifica>?sinonimo.}}</pre>
<pre>(iii) SELECT ?nome_taxom_nivel ?nivel WHERE { ?s <URI/taxon_dstaxon_cientifica> "Calanoida". ?s <URI/taxon_inleft> ?left. ?s <URI/taxon_inright> ?right. ?taxon <URI/taxon_inleft> ?left_t. ?taxon <URI/taxon_inright> ?right_t. ?taxon <URI/taxon_taxon_nivel_idtaxon_nivel> ?taxon_nivel. ?taxon <URI/taxon_dstaxon_cientifica> ?nome_taxom_nivel. ?taxon_nivel <URI/taxon_nivel_intaxon_nivel_ordem> ?ordem. ?taxon_nivel <URI/taxon_nivel_dstaxon_nivel> ?nivel FILTER (?left_t > ?left && ?right_t < ?right)} ORDER BY ?ordem</pre>	<pre>(iv) SELECT DISTINCT ?id_regiao ?id_ponto_coleta ?name_taxon ?nome_predador WHERE { ?id_regiao <URI/regiao_dsregiao> "da barra a ponte". ?id_ponto_coleta <URI/ponto_coleta_regiao_idregiao> ?regiao. ?id_amostra <URI/amostra_ponto_coleta_idponto_coleta> ?id_ponto_coleta. ?id_analise <URI/analise_amostra_idamostra> ?id_amostra. ?id_analise_bio <URI/analise_biomassa_analise_idanalise> ?id_analise; <URI/analise_biomassa_taxon_idtaxon> ?id_taxon. ?id_taxon <URI/taxon_dstaxon_cientifica> ?name_taxon; <URI/taxon_dstaxon_cientifica> ?nome_presa. ?w <URI/cad_trof_Taxon_idTaxon_predador> ?predador. ?predador <URI/taxon_dstaxon_cientifica> ?nome_predador. ?w <URI/cad_trof_inGrau_relacao> ?grau; <URI/cad_trof_Taxon_idTaxon_presa> ?e. FILTER (?id_regiao = ?regiao && ?id_taxon = ?e)} ORDER BY ?nome_taxon</pre>

Figura 2. Consultas SPARQL sobre BD Peld Guanabara

6. Conclusão

O projeto PELD Guanabara visa oferecer melhor entendimento da estrutura e do funcionamento do ecossistema da Baía de Guanabara e de suas respostas aos impactos antrópicos e climáticos. Como parte desta iniciativa, está sendo desenvolvido um sistema computacional para armazenar os dados ecológicos da Baía de Guanabara e permitir sua análise integrando informações geográficas, taxonômicas e sobre a cadeia alimentar das espécies deste ecossistema. O sistema utiliza um banco de dados relacional para armazenamento dos dados e uma visão RDF que homogeniza a publicação de hierarquias, conceitos e relacionamentos do domínio em uma estrutura do tipo grafo. De modo a oferecer a capacidade de inferência sobre relacionamentos transitivos, técnicas de conjuntos aninhados e de fecho-transitivo explícito estendem o esquema relacional adotado e são exploradas em consultas SPARQL.

Finalmente, como trabalhos futuros, pretende-se avaliar a adoção do padrão EML como

terminologia para a integração das bases PELD, bem como integrar a visão RDF ao contexto de dados associados conhecidos como *LinkedData*.

SPARQL results: (i)			SPARQL results: (ii)		
nome_pres	nome_predador	grau	nome_predador	nome presa	sinonimo
"Manjuba"	"Bagre"	1	"Bagre"	"Oithona hebes"	-
"Manjuba"	"Corvina"	1	"Bagre"	"Anchoa Marinii"	-
"Manjuba"	"Toninha"	1	"Bagre"	"Acartia tonsa"	-
			"Bagre"	"Brevoortia pectinata"	-
			"Bagre"	"Temora stylifera"	-
			"Bagre"	"Oithona plumifera"	-
			"Bagre"	"Temora turbinata"	-
			"Bagre"	"Acartia tonsa"	"Acartia (Acanthacartia) tonsa"

SPARQL results: (iii)		SPARQL results: (iv)			
nome_taxom_nivel	nivel	ld_regiao	ld_ponto_coleta	name_taxon	nome_predador
"Temoridae"	"Familia"	db:regiao/1511	db:ponto_coleta/1	"Acartia tonsa"	"Anchoa Marinii"
"Acartiidae"	"Familia"	db:regiao/1511	db:ponto_coleta/1	"Acartia tonsa"	"Brevoortia pectinata"
"Acartia"	"Genero"	db:regiao/1511	db:ponto_coleta/1	"Acartia tonsa"	"Micropogonias furnieri"
"Temora"	"Genero"	db:regiao/1511	db:ponto_coleta/1	"Acartia tonsa"	"Netuma barba"
"Acartia (Acanthacartia)"	"Subgenero"	db:regiao/1511	db:ponto_coleta/1	"Acartia tonsa"	"Pontoporia blainvillei"
"Acartia tonsa"	"Especie"	db:regiao/1511	db:ponto_coleta/1	"Oithona hebes"	"Anchoa Marinii"

Figura 3. Resultados das consultas

Agradecimentos

Este trabalho é parcialmente financiado pelo CNPq através de bolsas PCI (Proc. 382.489/09-8), produtividade de pesquisa (Proc. 309502/2009-8), PCI (Proc. 383344/2010-7) e DTI (Proc. 381663/2010-8).

Referências

- Campos, S.R., Martinhago A.Z., Massahud R.T., França A.M., Prieto L. E., Mendes J.D.C. (2009) Modelagem do Banco de Dados do Zoneamento Ecológico Econômico de Minas Gerais usando UML-GeoFrame. Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 25-30 abril, INPE, p. 4943-4949.
- Cavalcanti, M. J. (2005) Bancos de Dados sobre Biodiversidade na Amazônia: a Experiência do Projeto Biotupé. Biotupé: Meio Físico, Diversidade Biológica e Sociocultural do Baixo Rio Negro, Amazônia Central. Santos-silva, Aprile, Scudeller, Editora INPA, Manaus.
- Cavalcanti, M.J. (1993) ECOLOG: sistema gerenciador de bancos de dados para levantamentos ecológicos de campo. I Simpósio de Aplicações da Informática em Biologia, Campinas p. 44-45.
- Cavalcanti, M.J. (1996) Designing a graphical user interface for DELTA: some considerations. DATA Newsletter, 12: 2-4.
- Cavalcanti, M.J. (1998) A Windows-based data capture program for ATF. Alice Software News , 6: 13.
- Celko, J. SQL for Smarties, Advanced SQL Programming, Morgan Kaufmann, San Francisco, USA, 3 ed., 2005.
- Crivelenti, R.C., Coelho R.M., Adami S.F., Oliveira, S.R.M. (2009) Mineração de dados para inferência de relações solo-paisagem em mapeamentos digitais de solo. Pesq. agropec. bras., Brasília, v.44, n.12, p.1707-1715.
- Fontes, C.A. (2011) Explorando Inferência em um Sistema de Anotação Semântica, Dissertação de mestrado, Instituto Militar de Engenharia.
- Macário, C.G.N., Medeiros C.B. (2009) A framework for semantic annotation of geospatial data for agriculture. International Journal of Metadata, Semantics and Ontologies v.4 n.1/2, p.118-132.
- Macário, C.G.N., Sousa, S.R., Medeiros C.B. (2009) Annotating geospatial data based on its semantics. 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.
- Manzi, A. 2009. Gerenciamento de dados de Projetos Ecológicos de Longa Duração, Projeto de pesquisa, Edital MCT/CNPq N° 59/2009 - Propostas para apoio a projetos no âmbito do PELD.
- Navathe, S., Elmasri, R. (2005) Sistemas de Banco de Dados, Pearson Addison Wesley, São Paulo, 4 ed., 2005.