

Construção de Bases de Dados de Pesquisa em Ambientes de Desenvolvimento de Software Orientado a Domínio

Ana Mirtes M. Fouro¹

fouro@cpatu.embrapa.br

Káthia Marçal de Oliveira²

kathia@ucb.br

Ana Regina C. da Rocha³

darocha@cos.ufjf.br

Resumo. A construção de bases de dados de pesquisa é uma atividade em constante desenvolvimento em instituições com atividades em ciência e tecnologia. Acreditando que a utilização do conhecimento do domínio, previamente modelado, pode ser bastante útil na execução do projeto conceitual, propomos uma abordagem para uso de ontologias de domínio, no apoio à modelagem conceitual de bases de dados de pesquisa em ambientes de desenvolvimento de *software* orientados a domínio. Como produto desta abordagem, mostraremos a construção de Insecta, um ambiente de desenvolvimento de software orientado ao domínio de Entomologia e a utilização do conhecimento modelado na construção de uma base de dados de pesquisa - Dendrobase.

Abstract. The development of research databases is an activity in constant execution in institutions with focus on science and technology. Believing that the uses of domain knowledge, previously described and organized, can be useful in the execution of databases conceptual project, we consider an approach that uses domain ontologies to support research databases conceptual modeling in domain-oriented software development environment. To show this approach in real situation, we will present the Insecta definition process, an domain-oriented software development environment for Entomology and will present yet the knowledge defined uses in the development of a specific research database - Dendrobase.

1 INTRODUÇÃO

1 Empresa Brasileira de Pesquisa Agropecuária, Setor de Informação, Caixa Posta 48 – CEP 66095-100, Belém, PA-Brasil

2 Universidade Católica de Brasília, SGAN 916, módulo B, Asa Norte - CEP 70790-160 Brasília, DF-Brasil

3 Universidade Federal do Rio de Janeiro, Departamento, Caixa Posta 68511 – CEP 21945-970, Rio de Janeiro-RJ-Brasil

Bases de dados de pesquisa são construídas com o propósito específico de armazenar os dados coletados de situações reais para apoiar a realização de pesquisas científicas, estudos estatísticos, apoio a modelagens e simulações, dentre outros. Sua construção é uma atividade em constante desenvolvimento em instituições que desenvolvem projetos de pesquisa. Acreditamos que o conhecimento do domínio, previamente modelado, pode auxiliar no processo de construção deste tipo de base de dados, facilitando o entendimento dos requisitos a serem modelados e melhorando a comunicação e a produtividade da equipe.

Dentre as diversas atividades não triviais com as quais os desenvolvedores de *software* têm que lidar durante todo o processo de desenvolvimento, uma das mais críticas, depois da identificação correta dos requisitos (OLIVEIRA, 1999), é a modelagem conceitual. Mesmo com toda a evolução da tecnologia de banco de dados, a modelagem conceitual continua sendo a atividade mais crucial do processo de projeto de bases de dados, por tratar-se de uma atividade que pouco pode ser apoiada por ferramentas CASE (*Computer Aided Software Engineering*), uma vez que o projetista tem total responsabilidade pelo processo de entendimento e transformação dos requisitos em modelos conceituais (BATINI, 1992).

O grupo de engenharia de software da COPPE/UFRJ tem trabalhado na construção de ambientes de desenvolvimento de *software* orientados ao domínio (ADSOD) (OLIVEIRA, 1999b). Um ADSOD é um ambiente de desenvolvimento de software (ADS), que utiliza o conhecimento do domínio, modelado através de ontologias de domínio (OLIVEIRA, 1999c) para auxiliar, de várias formas, mas principalmente no entendimento do domínio, os desenvolvedores de *software* nas várias fases do processo de desenvolvimento.

Neste trabalho, buscamos mostrar como apoiar a construção de bases de dados de pesquisa em ambientes de desenvolvimento de software orientados a domínio. Serão apresentadas: a abordagem utilizada para a construção de *Insecta*, um ambiente de desenvolvimento de software orientado ao domínio de Entomologia¹; e a utilização do conhecimento modelado no apoio à construção de uma base de dados de sistemas genéticos para espécies arbóreas tropicais.

Além desta seção, este artigo está dividido em mais cinco seções. Na seção 2 é apresentada a relação entre Ontologias e ADSOD. A seção 3

mostra o processo de construção de bases de dados utilizando ontologias. Na seção 4 apresentamos Insecta, um ADSOD para o domínio de entomologia. Na seção 5 é mostrada a utilização do conhecimento do domínio de entomologia na construção de uma base de dados de sistemas genéticos para espécies arbóreas tropicais - dendrobase. Finalizando com a seção 6, apresentamos nossas conclusões e trabalhos futuros.

2 ONTOLOGIAS E ADSOD

Ontologia é um termo originário da filosofia e para o qual existem diferentes interpretações. A definição adotada no presente trabalho é a mesma adotada por Oliveira, (1999), então, ontologia é uma teoria sobre um domínio que especifica um vocabulário de entidades, classes, propriedades, predicados e funções; e um conjunto de relações que necessariamente interliga esse vocabulário, considerando ainda, que essa ontologia é uma descrição parcial projetada para ser compartilhada dentro de uma comunidade que concorda com a sua definição, e para o fim específico de desenvolvimento de software em um dado domínio.

Ontologias de domínio (GUARINO, 1998) são construídas para expressar conceitos de áreas de conhecimento particulares. Uma vez que o uso de ontologias neste trabalho tem por objetivo apoiar a modelagem de bases de dados de pesquisa em domínio específico, estamos diretamente interessados nesta categoria de ontologias, visto que, estas permitem modelar o conhecimento específico de uma área de conhecimento, podendo ainda serem utilizadas em ambientes de desenvolvimento de *software* orientados ao domínio (ADSOD).

Por sua vez, ambientes de desenvolvimento de software objetivam apoiar, de forma automatizada, o desenvolvimento de produtos de *software*, de forma integrada e controlada (OLIVEIRA, 1999). Ambientes de desenvolvimento de *software* orientados a domínio (ADSOD), como definidos por Oliveira, 2000, têm como característica principal, a introdução e uso do conhecimento do domínio, modelado através de ontologias de domínio, no ADS, provendo apoio aos desenvolvedores de *software*, em domínios específicos, através das várias fases do processo de desenvolvimento (OLIVEIRA, 2000; SILVA, 2000).

Para a construção de um ADSOD é necessário que os engenheiros do ambiente² definam o processo de software que será utilizado e o modelo de

conhecimento do domínio (teoria do domínio) que é utilizado para assistir o desenvolvedor ao longo do processo de desenvolvimento de software (OLIVEIRA, 2000).

3 PROCESSO DE CONSTRUÇÃO DE BASES DE DADOS DE PESQUISA USANDO ONTOLOGIAS

O processo de construção de bases de dados de pesquisa estabelece um conjunto de atividades definidas com base na estrutura comum estabelecida pela norma NBR ISO/IEC 12207 (1998) e no projeto de banco de dados utilizando a abordagem entidade-relacionamento (BATINI, 1992; CHEN, 1976).

Seu ciclo de vida é constituído de cinco fases sequenciais. Para cada fase, além da descrição de seu objetivo, são descritas todas as atividades que as compõem. Para cada atividade é feita a sua descrição, sendo identificadas as sub-atividades que devem ser executadas para que a atividade seja realizada, os produtos que devem ser gerados e as responsabilidades de execução, participação e aprovação dos produtos gerados (FOURO, 2002). A FIG. 1 mostra a estrutura geral do processo, em que também é destacado o procedimento para modelagem conceitual com base em ontologias de domínio, que mostram como ocorre a utilização do conhecimento do domínio numa atividade específica do processo definido.

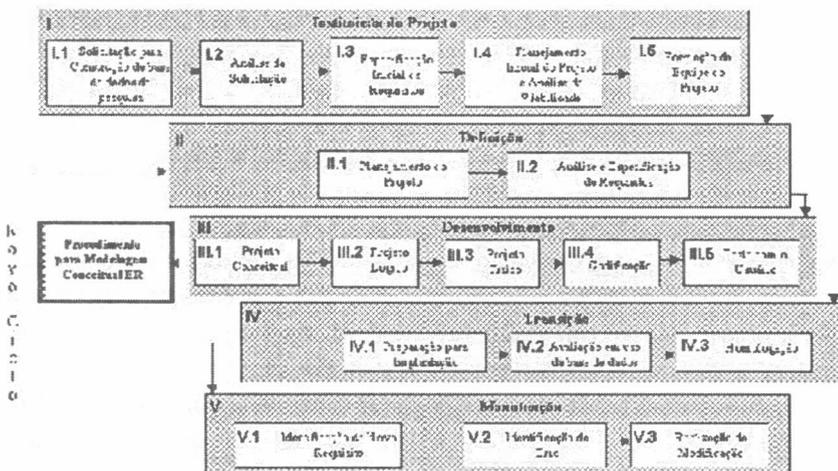


FIGURA 1 – Processo para Construção de Bases de Dados de Pesquisa

Na organização do procedimento para modelagem conceitual, foi considerada a relação de equivalência entre os construtores básicos de ontologias (tipos de conceitos, propriedades de conceitos e as relações possíveis entre os conceitos) e os construtores da modelagem conceitual entidade-relacionamento (entidades, atributos e relacionamentos), que foi estabelecida por Fouro (2002), através de um mapeamento, com base nos trabalhos de Wand (1996); Parsons (1997); Wand (1999); Kolp (2000).

O procedimento é executado com base no conteúdo de uma ontologia de domínio, definida com a execução da atividade *II.3 – Definição da teoria do domínio*, do processo definido e que está disponível no ADSOD construído.

4 INSECTA: UM ADSOD PARA O DOMÍNIO DE ENTOMOLOGIA

Para a construção do ADSOD foi definido o processo de software que será utilizado, apresentado na seção anterior (FIG. 1). Também foi construída a teoria do domínio para entomologia, cujo objetivo é modelar o conhecimento básico de entomologia, necessário ao desenvolvimento das atividades do entomólogo, com vistas a apoiar a construção de bases de dados de pesquisa em ADSOD para este domínio. Na especificação inicial de requisitos da teoria do domínio para entomologia, foram identificadas quatro questões de competência gerais a que a teoria do domínio deve responder, bem como as subteorias identificadas e que respondem às questões identificadas. Para cada subteoria foi construída uma árvore de conceitos, um conjunto de axiomas e uma sub-base de conhecimento (FOURO, 2002).

Um dos componentes da teoria do domínio de entomologia é a subteoria ecologia de insetos que expressa, através de seus conceitos, o objeto de estudo dos ecologistas, que examinam a dinâmica da *população*, a interação entre insetos (como a competição e alimentação), o *meio ambiente* (que inclui a disponibilidade de nutrientes) e o fluxo de energia através do *ecossistema*, e os padrões de distribuição e sucessão.

Após o estabelecimento dos aspectos necessários, temos disponível o ADSOD do domínio especificado. Neste trabalho definimos *Insecta*, conforme apresentado na FIG. 2, em que podemos observar, as fases que foram definidas para o processo e algumas das atividades que com-

õem o processo do ADSOD *Insecta*. Em destaque temos a sub-atividade investigação do domínio, podendo ser observado ainda, do lado direito, as ferramentas que apóiam a execução da sub-atividade selecionada.

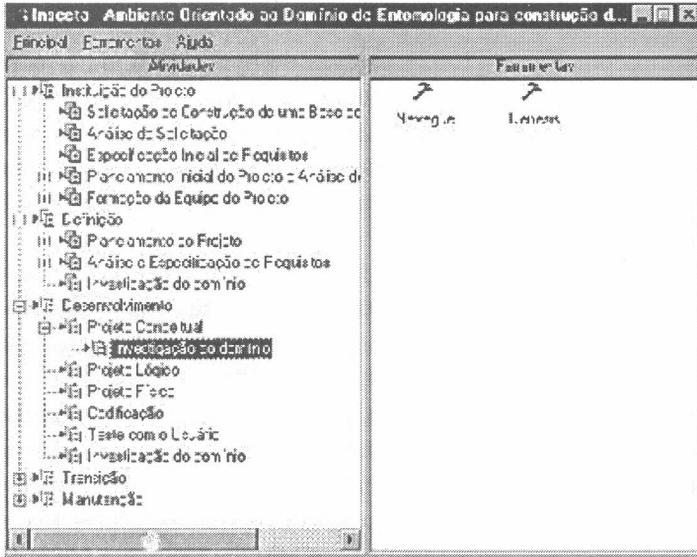


FIGURA 2 – ADSOD *Insecta*

Uma vez construído o ambiente de desenvolvimento, o mesmo poderá ser utilizado no apoio ao desenvolvimento de projetos no domínio considerado. No caso deste trabalho, o ADSOD *Insecta* poderá ser utilizado para a construção de bases de dados de pesquisa para o domínio de entomologia.

O uso do conhecimento do domínio ocorre através da execução da sub-atividade investigação do domínio, cujo objetivo é realizar a pesquisa, permitindo o uso do conhecimento do domínio necessário para a realização da atividade correspondente. No ADSOD construído sob a infraestrutura da estação TABA (ROCHA, 1990), essa investigação é apoiada pelo assistente de aprendizado do domínio NAVEGUE (SILVA, 2000).

Com esta ferramenta, o desenvolvedor tem acesso a informações sobre: a teoria do domínio e sua respectiva descrição; as subteorias que compõem a teoria do domínio e suas respectivas descrições; cada um dos conceitos que compõem as subteorias e sua respectiva descrição, sinônimos, termo

equivalente no idioma inglês, relações com outros conceitos, referências bibliográficas, projetos em que foi utilizado e exemplos; cada atributo dos conceitos, sua respectiva descrição, unidade e valores. A FIG. 3 mostra a utilização da ferramenta no ADSOD *Insecta*, em que podem ser observadas as subteorias que compõem a teoria do domínio para entomologia.

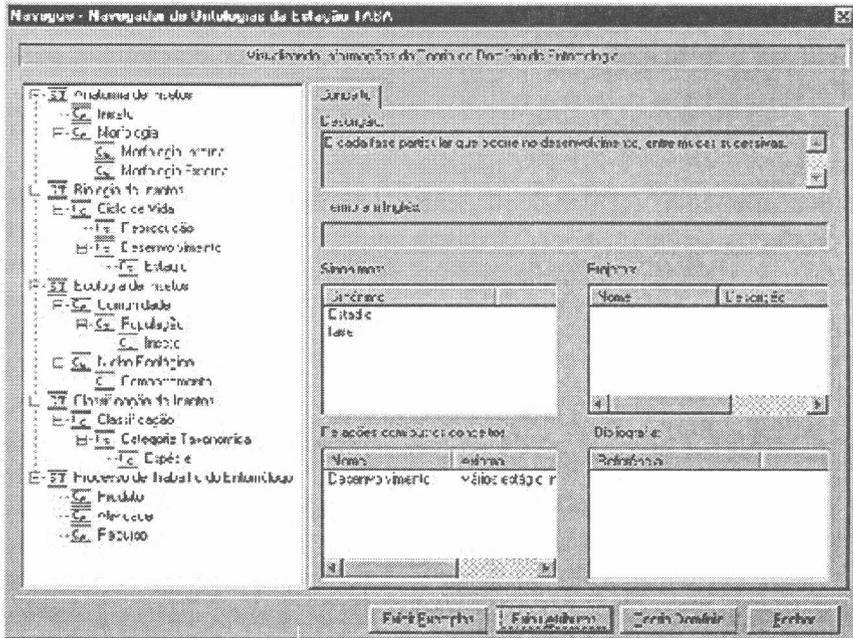


FIGURA 3 – Investigação do domínio no ADSOD *Insecta* utilizando a ferramenta Navegue

No contexto da Embrapa Amazônia Oriental³, para que uma base de dados possa ser construída é necessário que esta atividade tenha sido prevista no contexto de um projeto de pesquisa. Como esta abordagem foi definida fora do contexto de um projeto de pesquisa no âmbito da Embrapa, ainda não foi possível a sua aplicação numa situação real voltada exclusivamente para o domínio de entomologia. Mas, o conhecimento modelado através da teoria do domínio construída já foi utilizado no contexto de um projeto de pesquisa em andamento, conforme será apresentado na seção a seguir.

5 UTILIZAÇÃO DO CONHECIMENTO DO DOMÍNIO DE ENTOMOLOGIA NO APOIO À CONSTRUÇÃO DE UMA BASE DE DADOS DE SISTEMAS GENÉTICOS PARA ESPÉCIES ARBÓREAS TROPICAIS - DENDROBASE

A Dendrobase é uma base de dados de sistemas genéticos para espécies arbóreas tropicais, que integra dados ecológicos e genéticos, bem como resultados de estudos em parcelas experimentais. A idéia básica é reunir as informações existentes com a finalidade de determinar, para diferentes espécies-chave ou grupos de espécies arbóreas, os níveis críticos do estoque mínimo após a exploração florestal. Deve permitir o armazenamento de informações sobre sistema reprodutivo, endocruzamento, agentes de dispersão de pólen e sementes, fenologia reprodutiva, ecologia e genética, bem como informações sobre a abundância de espécies e distribuição espacial.

Para modelar os componentes da Dendrobase que armazenam informações sobre a dispersão de pólen e sementes, foram utilizadas a subteoria Ecologia de Insetos e a subteoria Classificação de Insetos (FOURO, 2002). Elas contribuíram para a identificação das entidades:

- *Vetores de Dispersão*, em que foram identificados os atributos: nome científico do vetor, nomes comuns, família do vetor e classe do vetor. Todos identificados em função da subteoria classificação de insetos.

- *Comportamento de Dispersores*, em que foram identificados os atributos: espécie do vetor (subteoria classificação de insetos), importância do vetor, tipo de dispersão (pólen ou sementes), distâncias (mínima, média e máxima) de dispersão de pólen e sementes. Todos identificados em função da subteoria ecologia de insetos, mas especificamente dos conceitos nicho ecológico, comportamento, instinto, resposta e estímulo.

Esta base de dados está em fase de validação dos dados inseridos. Na FIG. 4 podemos observar uma consulta direta aos dados sobre os vetores de dispersão.

The screenshot shows a web browser window with a search interface. The title bar indicates the URL is 'http://www.cnpq.br/dendrobase'. The browser's address bar shows 'http://www.cnpq.br/dendrobase'. The page content includes a navigation menu with 'Home', 'Policy', 'About', and 'Contact'. Below this is a search area with 'Menu Data Source', 'Data Search Menu', and 'Preview Reports'. The search results are displayed in a table titled 'Data Search Menu :: Vectors Genus Flow ::'.

Scientific Name of the Vector	Common Names	Other Common	Family of the Vector	Class of the Vector
<i>Acerididae</i> sp.	Graças uppers	-	Aceridae	flies
<i>Agrius convolvuli</i>	hawkmoth	-	Sphingidae	moths
<i>Alicidocera denboscii</i>	-	-	Chrysomelidae	beetles
<i>Aldape</i> sp.	-	-	Anthophoridae	bees
<i>Amegilla ardrewi</i>	trap-lining nong-tongued bee	-	Anthophoridae	bees
<i>Amegilla insularis</i>	trap-lining nong-tongued bee	-	Anthophoridae	bees
<i>Amegilla pendleburyi</i>	trap-lining nong-tongued bee	-	Anthophoridae	bees
<i>Amegilla</i> sp.	-	-	Anthophoridae	bees
<i>Amata latima</i>	-	-	Nymphalidae	butterflies
<i>Amata jacobaeae</i>	-	-	Nymphalidae	butterflies

FIGURA 4 – Dendrobase

Não foi utilizado o ADSOD *Insecta* para a construção da Dendrobase, uma vez que sua construção já havia sido iniciada desde 1998. Além disso, a mesma não se trata de uma base de dados do domínio de entomologia e sim do domínio da genética de espécies arbóreas, que necessita do conhecimento sobre a ecologia e classificação de insetos.

Mas, com a utilização da teoria do domínio de entomologia, foi identificado pelos especialistas, que a subteoria classificação de insetos pode ser utilizada da mesma forma, sem necessidade de ajustes para a classificação e identificação de espécies arbóreas tropicais.

6 CONCLUSÕES E TRABALHOS FUTUROS

O uso de ontologias para apoio ao projeto de banco de dados tem sido objeto de vários trabalhos de pesquisa. E, um dos aspectos que diferenciam nosso enfoque do trabalho de Dowell (1995), é que neste trabalho, a ontologia

é criada ou utilizada se estiver disponível, para apoiar a construção da base de dados, desde o projeto conceitual até o projeto físico, buscando proporcionar um aumento da produtividade da equipe desenvolvedora, a partir do entendimento do domínio apoiado pela existência de uma teoria do domínio.

A semelhança entre os trabalhos de Storey (1998 e 2000) e de Ullrich (2000) com o presente trabalho, está na idéia de que o projeto de bases de dados é melhor desenvolvido e demonstra melhor a semântica do domínio modelado com o apoio de ontologias, estejam elas incorporadas em ferramentas CASE ou em ADSOD.

Todos estes trabalhos identificam e enfatizam a existência de algum tipo de impacto que o uso de ontologias tem sobre a construção de modelos de dados, mas nenhum deles mostra como a construção de bases de dados pode ser apoiada pela utilização do conhecimento modelado em uma ontologia de domínio, como foi estabelecido na definição do processo para construção de bases de dados de pesquisa.

Com a disseminação e definição de ontologias de domínio, a atividade de modelagem conceitual tem, disponível, uma teoria de conteúdo (CHANDRASEKARAN, 1999), que apóia no entendimento, na estruturação e na representação do significados das *coisas* relevantes do mundo real, permitindo que o projetista gere modelos que reflitam cada vez mais natural e diretamente a concepção que os usuários têm do mundo real.

Foi apresentado o ADSOD *Insecta*, para o domínio de entomologia, construído para apoiar o desenvolvimento de bases de dados neste domínio específico. Ainda não foi possível a utilização do ambiente no apoio à construção de bases de dados especificamente no domínio de entomologia, mas o processo definido pode ser utilizado na construção de base de dados de pesquisa em qualquer domínio.

Foi construída uma teoria do domínio para entomologia e, para comprovar a interoperabilidade entre domínios diferentes, parte desta teoria já pode ser utilizada no apoio à construção de uma base de dados para o domínio de genética de espécies arbóreas tropicais, tendo sido identificado, ainda, que uma de suas subteorias deve transformar-se numa ontologia independente que fará integração com a teoria do domínio de entomologia e uma teoria que precisa ser construída para o domínio de genética de espécies arbóreas tropicais.

São considerados como objetivos primários da engenharia de *software*, o

aprimoramento constante da qualidade do produto de *software* que é gerado e o aumento considerável da produtividade dos engenheiros de *software*, sempre com a abrangência direcionada à eficácia e eficiência do produto. Foi apresentada, neste trabalho, uma abordagem que busca proporcionar o alcance destes objetivos na construção de bases de dados de pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABNT. *Tecnologia de informação – processos de ciclo de vida de software*: NBR ISO/IEC 12207. Rio de Janeiro, 1998.
- BATINI, C.; CERI, S.; NAVATHE, S.B. *Conceptual database design: an entity-relationship Approach*. [S.l.: s.n.], 1992.
- CHANDRASEKARAN, B.; JOSEPHSON, J.R.; BENJAMINS, V.R. What are ontologies, and why do we need them?. *IEEE Intelligent Systems*, p. 20-26, 1999.
- CHEN, P.P. The Entity-Relationship model - toward a unified view of data. *ACM Transactions on Database Systems*, v. 1, n. 1, p. 9-36, 1976.
- DOWELL, M.L.; STEPHENS, L.M.; BONNELL, R.D. Using a domain knowledge ontology as a semantic gateway among databases. In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING, IJCAI 1995, Montreal. Montreal, 1995.
- FOURO, A.M.M. *Apoio à construção de bases de dados de pesquisa em ambientes de desenvolvimento de software orientados a domínio*. Dissertação. COPPE/UFRJ. Rio de Janeiro, 2002.
- GUARINO, N. Formal ontology and information system. In: GUARINO, N. (ed) *Formal Ontology in Information System*. [S.l.]; IOS Press, 1998. p. 3-15.
- KOLP, M.; ZIMÁNYI, E. Enhanced ER to relational mapping and interrelational normalization. *Information and Software Technology*, n. 42, p. 1057-73, 2000.
- OLIVEIRA, K.M. *Modelo para construção de ambientes de desenvolvimento de software orientados a domínio*. Tese. COPPE/UFRJ. Rio de Janeiro, 1999.
- OLIVEIRA, K.M.; GALOTTA, C.; ROCHA, A.R.C. Construção de ambientes de desenvolvimento de software orientados a domínio, In: JORNADAS IBEROAMERICANAS DE INGENIERÍA DE REQUISITOS Y

- AMBIENTES DE SOFTWARE – IDEAS 2000, Cancun. Cancun, 2000. _____ . Defining and building domain-oriented software development environment. In: INTERNATIONAL CONFERENCE SOFTWARE & SYSTEMS ENGINEERING AND THEIR APPLICATIONS – ICSSEA, 12th . 1999, Paris. Paris, 1999c.
- OLIVEIRA, K.M.; ROCHA, A.R.C.; TRAVASSOS, G.H. Using domain-knowledge in software development environments, In: PROCEEDINGS OF THE 11TH INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING, 1999. Kaiserslauter. Kaiserslauter, 1999b, p. 180-187,
- PARSONS, J.; WAND, Y. Using objects for systems analysis. *Communications of the ACM*, v. 40, n. 12, p. 104-110, 1997.
- ROCHA, A.R.C.; AGUIAR, T.C.; SOUZA, J.M. Taba: a heuristic workstation for software development. In: COMPEURO 90, Tel Aviv. Tel Aviv, 1990.
- SILVA, C.G.M. da. *Netuno: um ambiente de desenvolvimento de software orientado a domínio*. Dissertação. COPPE/UFRJ. Rio de Janeiro, 2000..
- STOREY, V.C., Understanding and representing relationship semantics in database design In: PROCEEDINGS OF THE 5 TH INTERNATIONAL CONFERENCE ON APPLICATIONS OF NATURAL LANGUAGE TO INFORMATION SYSTEMS, NLDB' 2000. Versailles. Versailles, 2000.
- STOREY, V.C.; DEY, D.; ULLRICH, H.; SUNDARESAN, S. An ontology-based expert system for database design. *Data and Knowledge Engineering*, v. 28, n.1, p.31-46, 1998.
- ULLRICH, H.; PURAO, S.; STOREY, V.C. An ontology for classifying the semantics of relationships in database design. In: PROCEEDINGS OF THE 5 TH INTERNATIONAL CONFERENCE ON APPLICATIONS OF NATURAL LANGUAGE TO INFORMATION SYSTEMS, NLDB' 2000. Versailles. Versailles, 2000.
- WAND, Y. Ontology as a foundation for meta-modelling and method engineering. *Information and Software Technology*, n. 38, p. 281-287, 1996.
- WAND, Y; STOREY, V.C.; WEBER, R. An ontological analysis of the relationship construct in conceptual modeling. *ACM Transaction on Database Systems*, v. 24, n. 4, p. 494-528, 1999.