

UMA PLATAFORMA ORIENTADA À OBJETOS PARA DESENVOLVIMENTO DE SISTEMAS ESPECIALISTAS DE DIAGNÓSTICO DE DOENÇAS VIA WEB

SILVIA MARIA FONSECA SILVEIRA MASSRUHÁ¹
JULIANO PASTORELLI DUTRA²
SÉRGIO APARECIDO BRAGA CRUZ³
SANDRA SANDRI⁴
JACQUES WAINER⁵

RESUMO: No escopo deste trabalho é apresentada uma infra-estrutura orientada à objetos, que suporta via web uma abordagem integrada para diagnóstico, investigação e tratamento de desordens. Nesta nova abordagem, o conhecimento é basicamente modelado através de associações causais entre objetos e a inferência é abdutiva. A validação do modelo teórico foi realizada em fitopatologia, abordando especificamente a diagnose, investigação e tratamento de doenças de milho. Entretanto, devido a sua generalidade, espera-se que os desenvolvimentos teóricos obtidos possam ser aplicados para resolver problemas de diagnóstico tanto na área vegetal quanto na área animal.

PALAVRAS-CHAVE: inferência abdutiva, lógica nebulosa, orientação à objetos, sistemas especialistas, doenças de plantas.

AN OBJECT-ORIENTED FRAMEWORK FOR DEVELOPMENT OF DISEASES DIAGNOSIS EXPERT SYSTEMS BY WEB

ABSTRACT: This work presents an object-oriented approach to support by web an integrated framework for diagnosis, investigation and treatment of disorders (diseases, computer faults, etc.). In this new framework, the knowledge is basically modeled through causal associations between objects and inference is abductive. The validation of the theoretical model has been performed in phytopathology, specifically, at the diagnosis, investigation and treatment of the corn plants diseases. However, due to its generality, it is expected that the theoretical developments obtained in this work can successfully employed to solve diagnosis problems in animal and vegetal area.

KEY-WORDS: abductive inference, fuzzy sets, object-oriented, expert systems, knowledge, plants diseases.

1. INTRODUÇÃO

Na Embrapa Informática Agropecuária foi desenvolvido, em parceria com a Embrapa Milho e Sorgo, um sistema para diagnóstico de doenças de milho, que está atualmente em uso (<http://diagnose.cnptia.embrapa.br>). Os grandes desafios da construção destes sistemas estão relacionados com o entendimento, modelagem e representação do conhecimento. No sistema para diagnóstico de doenças do milho foram utilizadas técnicas manuais para aquisição do

¹ Doutora em Computação Aplicada, Pesquisadora da Embrapa Informática Agropecuária, E-mail: silvia@cnptia.embrapa.br

² Bacharel em Ciência da Computação, Consultor Técnico, E-mail: juliano@brainweb.com.br

³ Mestre em Engenharia Elétrica, Pesquisador da Embrapa Informática Agropecuária, E-mail: sergio@cnptia.embrapa.br

⁴ Doutora em Ciência da Computação, Pesquisadora do Instituto Nacional de Pesquisas Espaciais – INPE e Pesquisadora Visitante do Instituto de Inteligência Artificial de Barcelona - IIIA, E-mail: sandri@iia.csic.es

⁵ Doutor em Ciência da Computação, Professor e Pesquisador do Instituto de Computação - Unicamp, E-mail: wainer@ic.unicamp.br

conhecimento baseadas em entrevistas, em acompanhamento ou modelos. Quanto à representação do conhecimento foram utilizados regras de produção que são baseados nos mecanismos de inferência dedutiva (efeitos -> causa). Pode-se argumentar que o problema principal com este tipo de abordagem é que os fenômenos ocorrem na natureza de forma contrária a este raciocínio.

Na década de 90, o conhecimento passou a ser modelado na sua forma natural: causas -> efeitos, com o raciocínio do tipo abduutivo (Peng & Reggia, 1990). Além do diagnóstico correto, o tratamento efetivo é um outro fator que interfere no sucesso de um processo de raciocínio clínico. Em complementação a estas tarefas, quando a investigação é bem feita, obtém-se maior eficácia e eficiência no diagnóstico, além de reduzir os custos e riscos.

Em MASSRUHÁ et al. (2005) foi apresentada uma abordagem integrada para diagnóstico, investigação e tratamento de desordens. Nesta nova abordagem, denominada Teoria das Coberturas Nebulosas (TCN), o conhecimento é basicamente modelado através de associações causais e a inferência é abduitiva (desordens->manifestações). Conceitos de Parsimonious Covering Theory (PCT) (PENG & REGIA, 1990), lógica nebulosa e teoria de decisão são também integrados, de maneira a tratar os vários aspectos inerentes aos processos envolvidos em raciocínio clínico.

No contexto deste trabalho, está descrita a metodologia bem como os resultados obtidos para portar o modelo acadêmico de diagnóstico, investigação e tratamento desenvolvido em Prolog, para uma nova arquitetura baseada em tecnologia J2EE (Java 2 Platform, Enterprise Edition). Na seção 2 é apresentado o objetivo deste trabalho de pesquisa. Nas seções 3 e 4 estão descritas a metodologia utilizada bem como os resultados obtidos neste trabalho. Finalmente, na seção 5 estão descritas as conclusões.

2. OBJETIVO

O objetivo deste trabalho é apresentar um *framework* orientado à objetos que suporte uma abordagem integrada para diagnóstico, investigação e tratamento de desordens. Neste trabalho, está descrita a metodologia bem como os resultados obtidos para portar o modelo acadêmico de diagnóstico, investigação e tratamento desenvolvido em Prolog, para uma nova arquitetura baseada em tecnologia J2EE. As bases de conhecimento de doenças de milho e feijão foram utilizadas como estudo de caso do presente trabalho. Entretanto, o objetivo final deste trabalho de pesquisa é compará-lo com as outras arquiteturas desenvolvidas na Embrapa Informática Agropecuária visando validar a flexibilidade e a facilidade de expandir o *framework* proposto para outras bases de conhecimento.

3. MATERIAL E MÉTODOS

Nesta seção está descrita a metodologia utilizada no desenvolvimento deste trabalho. Este trabalho foi desenvolvido em três etapas principais: Modelagem da arquitetura, especificação do ambiente de desenvolvimento e testes da aplicação.

Na primeira etapa, foi necessário investigar a Teoria das Coberturas Nebulosas – TCN proposta em MASSRUHÁ et al. (2005). A arquitetura proposta para suportar o modelo TCN é composta por 3 módulos principais: diagnóstico, investigação e tratamentos. Em todos os 3 módulos, o conhecimento envolvido é do tipo causal e todo o raciocínio é abduutivo. Na tarefa de diagnóstico, as associações que modelam o conhecimento são do tipo *desordem -> manifestações*. Na tarefa de investigação, as associações são do tipo *exames -> manifestações*. Finalmente, na tarefa de tratamento, as associações são do tipo *tratamentos -> desordens*. Como o conhecimento sobre o caso é dado pelas manifestações presentes no caso, o raciocínio é naturalmente abduutivo.

De forma complementar a análise da TCN, também foram analisados os códigos fontes em Prolog utilizados na implementação do modelo. Antes de iniciar o processo de desenvolvimento do software, todos os códigos Prolog foram traduzidos manualmente para a tecnologia e lógica JAVA, pois os softwares que se propunham a automatizar este processo foram bem insuficientes. Os programas em Prolog que compõem a TCN são claramente divididos em três partes: base de conhecimento, mecanismo de inferência e especificação do caso particular. O mecanismo de inferência abduutivo, apresentado em MASSRUHÁ et al. (2005), foi implementado através dos algoritmos: *bipartido_nebuloso*, *exames_nebulosos* e *tratamentos_nebulosos*.

Após a revisão da arquitetura da aplicação proposta em MASSRUHÁ et al. (2005), a nova arquitetura foi desenhada no modelo três camadas, conforme as mais rigorosas especificações do J2EE. Isto permite que se tenha uma simplificação no processo de manutenção e expansões futuras do sistema. Na Figura 1 está apresentada a abordagem orientada a objetos proposta neste trabalho.

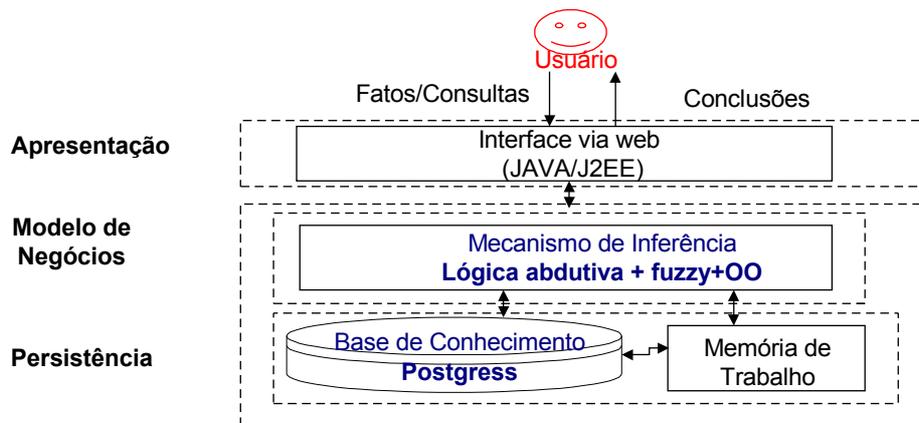


Figura 1 : Arquitetura IV – JAVA(J2EE)

O modelo J2EE em três camadas contempla: **Camada de Apresentação:** responsável pelo layout, navegação e tratamento das ações; **Camada de Negócios:** responsável por todas as lógicas e regras que definem a aplicação; **Camada de Persistência:** responsável pelo armazenamento e consulta ao Banco de Dados.

O segundo passo, foi montar uma infra-estrutura do projeto (arquitetura) baseada em definições de padrões de projeto e frameworks. Nesta etapa, foram utilizados os mais importantes frameworks (em código livre), para tornar mais produtivo o processo de desenvolvimento, a saber: **Struts:** padroniza o controle das ações vindas da camada web e sua apresentação (<http://struts.apache.org/>); **Spring:** injeção de dependências, instanciando objetos antecipadamente (<http://www.springframework.org/>); **Hibernate:** facilita a persistência e cria uma abstração para maior independência da base de dados (<http://www.hibernate.org/>).

Para que o projeto pudesse ser implementado em tecnologia J2EE foi necessário inicialmente modelar o problema sob uma perspectiva de Orientação a Objetos (OO) (RUMBAUGH, 1999), bem como desenhar toda a Base de Dados, que no sistema original era apenas uma base simplificada de conhecimento textual em Prolog, contendo apenas a chave de identificação (id) e nome da entidade.

Antes do início das atividades de desenvolvimento, segunda etapa, foram propostas e introduzidas algumas alterações no ambiente de desenvolvimento do sistema, dentre elas a adoção da ferramenta de desenvolvimento Eclipse e a definição do sistema gerenciador de banco de dados pelo PostgreSQL. Na terceira etapa, para testar o núcleo do sistema, foi necessário realizar testes modulares *bottom-up*. No início foram estudadas as entradas e saídas de métodos auxiliares, tais como: *nebulosos*, *intervalos*, etc. Todo o princípio dos testes era o de executar a mesma chamada em Prolog e Java e comparar a saída, que deveria ser idêntica.

Na seqüência foram realizados também testes similares com os métodos específicos com cada uma das três etapas do processo: diagnóstico, investigação e tratamento.

4. RESULTADOS E DISCUSSÕES

Para fins didático, antes de detalhar os resultados obtidos neste trabalho, na Tabela 1 é apresentada uma síntese das abordagens anteriores ao presente trabalho adotadas no âmbito da Embrapa Informática Agropecuária. (I) A primeira versão do sistema especialista para diagnóstico de doenças do milho (v1.0) foi baseado em regras onde o mecanismo de inferência é dedutivo e sua base teórica é a lógica de primeira ordem. Esta versão do sistema foi desenvolvida em WEBLS que é um shell para implementar sistemas especialistas do tipo Q&R (questões e respostas) para Windows (MASSRUHÁ et al., 2000). (II) Uma nova arquitetura para construção de sistemas especialistas via web proposta em CRUZ et al (2002) que é muito similar a arquitetura I. Nesta arquitetura II foi desenvolvido o ServCLIPS que utiliza o CLIPS como máquina de inferência e, conseqüentemente, a linguagem de programação do CLIPS para desenvolvimento dos sistemas especialistas. (III) Em Massruhá et al. (2005) foi apresentada uma abordagem integrada para diagnóstico, investigação e tratamento de desordens (MASSRUHÁ et. al 2004, 2005). Nesta nova abordagem (III), implementada em Prolog, o conhecimento é basicamente modelado através de associações causais e a inferência é abduativa (desordens->manifestações), de maneira a tratar os vários aspectos e incertezas inerentes aos processos envolvidos em raciocínio clínico. (IV) Dentre os resultados obtidos, observou-se que a infra-estrutura atual (IV), que é baseada no modelo da arquitetura III, é muito mais robusta, eficiente e flexível. Um exemplo de flexibilidade que foi introduzida é a de permitir que exatamente o mesmo sistema (em uma única instância) seja capaz de distinguir entre diferentes problemas da agricultura (milho, soja, algodão, etc) e da pecuária (gado, ovelha, etc..). Para tal, somente é necessário que o próprio especialista cadastre outras bases de conhecimento (com doenças, manifestações, exames, etc) para os respectivos problemas da agricultura e da pecuária.

	Representação do conhecimento	Mecanismo de Inferência	Base teórica	Ferramentas de implementação	Internet	Desenvolvedor	Usuário
I	Regras de produção	Dedução	Lógica de predicados de primeira ordem	WeBLS – CGI Prolog	Sim	Engenheiro do conhecimento	Produtor e especialistas
II	Regras de produção	Dedução	Lógica de predicados de primeira ordem	SevCLIPS	Sim	Engenheiro do conhecimento	Produtor e especialistas
III	Regras de produção	Abdução + Lógica nebulosa	TCP + lógica nebulosa	AMZI Prolog	Não	Engenheiro do conhecimento	Produtor e especialistas
IV	OO	Abdução + lógica nebulosa	TCP + lógica nebulosa	JAVA (J2EE)	Sim	Especialista	Produtor e especialistas

Tabela 1 :Comparação entre as 4 arquiteturas propostas.

Outra flexibilidade é a possibilidade de integração com outros sistemas, tal como o Agritempo (www.agritempo.gov.br) que fornece variáveis como temperatura e umidade, simplificando o uso e aumentando a precisão do sistema. Boa parte do sistema foi desenhado de forma a permitir configuração, sem a necessidade de se refazer código o que permitirá um uso mais amplo do sistema (para várias culturas simultaneamente), com fases, severidades e comportamento temporal bastante distintos. Esta parametrização foi o cerne deste trabalho de pesquisa, pois assim, é possível utilizar o sistema não apenas para a cultura de milho (inicialmente proposta), mas para qualquer outra, sem a necessidade de ter que executar diferentes instâncias do mesmo sistema. Cabe ressaltar ainda que o modelo proposto na

arquitetura III (PROLOG) foi mapeado para arquitetura IV (JAVA) mantendo as propriedades inerentes ao modelo.

5. CONCLUSÕES

O resultado final alcançado foi satisfatório tanto do ponto de vista de arquitetura da aplicação quanto do ponto de vista da aplicação final. Em relação a nova arquitetura pode-se destacar que: (a) mostrou-se mais flexível, robusta e eficiente que o Prolog, sob o ponto de vista de engenharia de software, para a implementação do modelo TCN; (b) conforme previsto em MASSRUHÁ et al. (2005), a geração automática da base de conhecimento, através do módulo especialista, demanda muito menos tempo e atenção; (c) a integração dos três módulos que compõem a TCN (*bipartido_nebuloso*, *exames_nebulosos* e *tratamentos_nebulosos*) em uma única infra-estrutura tornou o sistema mais eficiente; (d) a interface mais amigável dos três módulos que compõem a TCN (*bipartido_nebuloso*, *exames_nebulosos* e *tratamentos_nebulosos*) tornou o sistema viável sob o ponto de vista de usabilidade; (e) confirmou-se a hipótese da simplificação no processo de manutenção e evolução do sistema. Cabe ressaltar ainda que a arquitetura proposta permite trabalhar o processo de diagnóstico, investigação e tratamento de quaisquer doenças na área vegetal e animal. Especificamente, neste trabalho, a validação contemplou desordens de milho e feijão. Entretanto, espera-se que os desenvolvimentos teóricos obtidos neste trabalho possam ser aplicados para outros problemas de diagnóstico dinâmico. Um exemplo na área de diagnóstico e controle de falhas de satélite pode ser encontrado em MASSRUHÁ (2003). Esta generalização é muito importante para a Embrapa, pois a quantidade de conhecimento da Instituição é farto, o que tornará possível explorar o mesmo mecanismo de apoio ao produtor em diferentes áreas do conhecimento rural.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- CRUZ, S. A. B.; LEITE, M. A. A.; MOURA, M. F.; SOUZA, E. ServCLIPS: a tool to develop www expert systems. In: WORLD CONGRESS OF COMPUTER IN AGRICULTURE AND NATURAL RESOURCES, 2002, Iguazu Falls. **Proceedings...** Iguazu Falls: ASAE, 2001. p. 636-641.
- MASSRUHÁ, S. M. F. S.; CRUZ, S.A.B. da; SOUZA, E.de. Diagnose virtual: um sistema para diagnóstico de doenças do milho via web. In: CONGRESSO DA SBI-AGRO, 2. = AGROSOFT 99, 19 e 20 de novembro de 1999, Campinas. **Anais ...** [S.l.]: s.n., 2000.
- MASSRUHÁ, S. M. F. S. **Uma teoria de coberturas nebulosas para diagnóstico, investigação e tratamento**. 2003. 251 f. Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos.
- MASSRUHÁ, S. M. F. S., SANDRI S. A. ; WAINER,J., 2004. Ordering manifestations for investigation in incomplete diagnosis. In: 10th International Conference IPMU 2004. **Proceedings...** Perugia, Italy: 2004. pp.1153-1160
- MASSRUHÁ, S. M. F. S. ; SANDRI, S. ; WAINER, J. ; MORANDI, M. An integrated framework for clinical problem solving in agriculture. In: Efitá /WCCA 2005 Joint Congress on IT in Agriculture. **Proceedings...** Vila Real, Portugal: Efitá /WCCA 2005, 2005a. p. 1400-1407.
- PENG, Y., REGGIA, J. A. **Abductive inference models for diagnostic problem-solving**. New York: Springer Verlag, 1990. 285 p.
- RUMBAUGH, J.; JACOBSON, I.; BOOCH, G. **The Unified modeling language reference manual**. New York: Addison Wesley Longman Inc, 1999, 550p.
- SUN MICROSYSTEMS, Inc. **Java 2 Platform, Enterprise Edition (J2EE)** Disponível em: <<http://java.sun.com/j2ee/index.jsp>>. Acesso em: 10 jun. 2005.