USO DE ONTOLOGIAS PARA A BUSCA POR SOFTWARE AGROPECUÁRIO

Autores

Frederico de Miranda Coelho
Email: coelho@agrosoft.softex.br
Vínculo: Universidade Federal de Juiz de Fora - Curso de Informática - Bolsista de CNPq - Núcleo de Pesquisa em Qualidade de Software
Endereço: ICE – Campus Universitário – Juiz de Fora/MG CEP.: 36036-330
Telefone: 32 -229-3301

Fernanda Claudia Alves Campos
Email: fernanda@agrosoft.softex.br
Vínculo: Universidade Federal de Juiz de Fora – Departamento da Ciência da Computação - Núcleo de Pesquisa em Qualidade de Software
Endereço: ICE – Campus Universitário – Juiz de Fora/MG CEP.: 36036-330
Telefone: 32 -229-3311

Regina Maria Maciel Braga
Email: regina@cos.ufrj.br
Vínculo: Universidade Federal de Juiz de Fora – Colégio Técnico Universitário - Núcleo de Pesquisa em Qualidade de Software
Endereço: Rua Bernardo Mascaranha, 1283 Bairro Fabrica
Telefone: 32-215-1947

Resumo

Este trabalho relata o projeto de especificação de uma ontologia para o domínio agropecuário que identifica perfeitamente os softwares agropecuários disponíveis no mercado, através de uma nomenclatura comum utilizada pela comunidade agropecuária. O resultado desta classificação será utilizado para melhorar os sistemas de busca dos guias e catálogos do setor e identificar os diferentes produtos e atributos para avaliação considerando suas funções específicas. Estamos trabalhando com sete categorias, dezenove setores e vinte e três sub-setores.

Abstract

This paper describes a project that aims to specify an ontology for the agricultural domain that will identify all the commercial software products, according to the community common vocabulary. The results can be used to improve the search systems of guides and catalogs and, also, to identify the different products and their attributes to be evaluated considering their specific functions. We are working with seven categories, nineteen sectors and twenty three sub-sectors.

Palavras chaves

Software Agropecuário, Ontologia, Qualidade de Software.

1 – INTRODUÇÃO

A disponibilidade de produtos de software para a agropecuária é bem diversa. Atualmente existem softwares no mercado que atendem aos mais diversos segmentos da agricultura e da pecuária. No entanto, ao longo dos anos, pudemos identificar a dificuldade na busca e identificação de produtos que pudessem atender às necessidades dos usuários finais.

Existem alguns catálogos e guias de classificação dos softwares agropecuários, mas estes são incompletos e, geralmente, utilizam sistemas de busca inadequados. O usuário que procura um software que atenda às suas necessidades em termos de funcionalidades, plataforma operacional, áreas de cobertura, entre outras, sente dificuldades em encontrar um produto ou produtos adequados e com isso acaba por não adquiri-los. Isso ocorre por falta de uma classificação precisa ou por uma classificação baseada em palavras chave que muitas vezes não são consideradas na consulta realizada pelo usuário por desconhecimento destas.

Observando este problema e analisando os resultados alcançados com o uso de ontologias no entendimento e delimitação de um dado domínio de aplicação (Braga et al., 1999; Falbo, 1998; Guarino, 1998) e para atender a uma demanda dos usuários e desenvolvedores, está sendo desenvolvido um projeto para a especificação de uma classificação para o setor de software agropecuário, utilizando conceitos e termos usualmente referenciados pelos usuários deste tipo de software. Para isso contamos com o apoio do Núcleo Softex Agrosoft Juiz de Fora-Viçosa-Lavras e Embrapa-CNPGL.

O objetivo é a especificação de uma ontologia, caracterizada como uma rede semântica, onde termos (conceitos, funcionalidades e relacionamentos) do domínio são explicitados para facilitar a busca por software. Assim, os produtos de software disponíveis no mercado são classificados de acordo com esta rede. Desta forma, um determinado produto pode ser classificado utilizando vários conceitos, funcionalidades, entre outros.

Um determinado usuário que queira adquirir conhecimentos sobre a área, pode utilizar a ontologia, uma vez que para cada um dos termos e relacionamentos é disponibilizada uma descrição dos mesmos, juntamente com os termos relacionados. Neste aspecto, a ontologia poderia ser comparada com um guia em uma rede hipermídia, onde o usuário vai sendo guiado pelas informações e navegando através de seus relacionamentos.

2 – CONCEITOS BÁSICOS

As pesquisas mais recentes em Inteligência Artificial (Guarino, 1998) apontam o uso de ontologias de domínio como um dos mecanismos promissores para se descrever de forma precisa os conceitos relevantes a um dado domínio de aplicação. O autor descreve o termo ontologia, no contexto de sistemas de informação, como sendo um vocabulário específico usado para descrever uma certa realidade. Este vocabulário é descrito através de uma hierarquia de conceitos e relacionamentos entre estes conceitos.

A vantagem de se usar ontologias neste contexto é que elas permitem a reutilização e o compartilhamento de um vocabulário comum entre softwares das mais diversas

origens, permitindo assim uma classificação mais precisa de aplicações (softwares) do domínio.

As abordagens, disponíveis para recuperação de informação de forma geral, mesmo as disponíveis via Web, podem ser consideradas tradicionais. Geralmente estas buscas são baseadas em texto livre; que realizam uma análise na página HTML disponibilizada para classificação; ou buscas baseadas em um vocabulário controlado.

O uso de um vocabulário controlado é baseado no uso de um conjunto de palavras chave que são utilizadas como termos para a indexação. Estas palavras chave são derivadas e definidas por especialistas e são projetadas para descrever da melhor maneira possível conceitos relevantes ao domínio. Neste contexto, a criação sistemática e controlada deste vocabulário poderia ser considerada como similar à criação de uma ontologia do domínio.

Segundo (Prieto-Díaz, 1991), o uso de um vocabulário controlado para a classificação de software é mais apelativo do que a análise de texto livre, uma vez que a disponibilidade de texto livre que classifica um software é muito pouca. Além disso, quando um software é descrito, geralmente esta descrição é sucinta e baseada em termos técnicos, o que dificulta uma classificação mais apurada, onde conceitos do domínio de aplicação do software sejam levados em consideração. Em um vocabulário controlado, os termos representativos do domínio são organizados em um esquema de classificação, que provê uma rede de relacionamentos pré-definidos, introduzindo assim mais informação semântica.

Existem alguns catálogos e guias de classificação dos softwares agropecuários, entre eles o Guia Agrosoft (Guia Agrosoft, 2000) que é disponibilizado na Web e utiliza a busca por categorias, sete ao todo, ou por palavras chave constante da descrição do produto e o catálogo FARMSOFT ´94 (FARMSOFT, 1994) que é dividido em cinco grandes áreas, onde o software deve se classificar em uma e somente uma dessas áreas, em setores, onde encontram-se tanto setores mais gerais quanto específicos juntos, em funções gerais e usuários, constantes de uma pequena lista com poucas opções.

A abordagem proposta utiliza um mecanismo de classificação mais adequado e abragente, uma vez que é baseado em vários aspectos do software (facetas), que ainda são baseados em uma ontologia previamente definida por especialistas do domínio. Para cada ligação do software com termos da ontologia são atribuídos pesos, que denotam uma maior ou menor relação do software com um dado termo ontológico.

3 – ONTOLOGIA PARA SOFTWARE AGROPECUÁRIO: UMA PROPOSTA

A abordagem deste trabalho pode ser considerada como um esquema de vocabulário controlado, onde este vocabulário é descrito através de uma ontologia criada para a área de software agropecuário. Com isso, garantiu-se a corretude e precisão da classificação em relação aos termos do domínio. Foram levados em consideração diferentes aspectos ou abordagens pelas quais o software pode ser classificado. Criou-se um conjunto de facetas, similar ao trabalho de (Prieto-Diáz, 1991), que baseados na opinião de especialistas do domínio, acredita-se que descrevem

perfeitamente os diferentes “prismas” pelos quais um software agropecuário possa ser classificado.

Para a especificação desta ontologia do domínio foram utilizadas técnicas de engenharia de domínio, particularmente o método de engenharia de domínio Odyssey-DE (Braga, 1999). O modelo, que especifica os termos ontológicos, formando assim uma rede semântica de termos do domínio, é denominado de modelo de *features*. Esse modelo captura as características (*features*) gerais das aplicações desenvolvidas em um domínio, explicitando as similaridades e diferenças, classificadas de forma hierárquica no modelo.

Esta ontologia permite a classificação de um software em mais de uma categoria e o resultado desta busca deverá listar todos os softwares que possuem tais características ou similaridades a estas. Outra opção para o usuário será fazer uma busca hierárquica, onde poderá ir passo a passo até encontrar o que deseja. Já se tem sete categorias, figura 1.

Categorias						
Administração	Comércio	Controle de Processos	Produção Animal	Produção Vegetal	Treinamento	Outros

Figura 1 – Categorias existentes

Os setores em estudo totalizam dezoito:

Agroecologia, Análise da Água, Analise do Solo, Arboricultura, Comércio de Agrícola, Comércio Animais, Controle de Doenças, Controle de Pragas, Cooperativas, Cultivo, Estufa, Laticínios, Maquinário, Mecanismo de Cooperativas e Fornecedores, Nutrição Animal, Produção Geral de Plantas, Propriedades Rurais e Vegetais.

Nesta hierarquia tem-se vinte e sete sub-setores, trinta áreas de atuação. As funcionalidades são vinte e cinco, assim descritas:

Administração (Função), Agrometeorologia, Alimentação, Ambiente, Analise de Energia, Automatização, Avaliação de Adoção de Tecnologia, Contabilidade, Defensivos Agrícolas, Diagnostico Agroecologico, Economia, Estudo, Fertilizantes, Gerenciamento, Marketing, Monitoramento, Planejamento, Planos de Negócios, Previsões, Processos de Produção e Armazenagem, Produção Animal (Função), Programas Comparativos, Recursos Genéticos, Regulamentação, Sociologia Rural.

As sub-funcionalidades são trinta e seis, são sete funcionalidades específicas e quatorze tipos de usuário, assim listados:

Agricultores, Associações de Produtores, Comerciantes, Contadores, Cooperativas, Ecologistas, Economistas, Empresas de Gerenciamento, Laticínios, Pecuaristas, Pesquisadores, Serviços Gerais, Veterinários e Outros.

Completando essa proposta têm-se cinco *features* de hardware e três de software.

O metaesquema da ontologia está fundamentado no esquema de classificação, pois além de procurar por palavras chave e procurar parcialmente por *strings*, o esquema de classificação pode servir como uma estrutura hierárquica para navegação pela base de dados. Para a Ontologia do Setor de Software Agropecuário foram criadas facetas, com objetivo de permitir uma classificação do software baseada em um conjunto mais amplo de categorias, e para cada faceta foram detalhados os conceitos classificadores - *features*, adquirindo assim uma precisão mais apurada e flexibilidade na classificação.

Na proposta deste estudo existem duas grandes facetas, uma delas é onde se encontram as grandes características do software – Categorias e a outra é onde estão as possíveis funcionalidades mais gerais do software – Funcionalidades. Já as características mais especificas encontram-se em uma ordem decrescente, primeiro em Setores, depois em Sub-Setores e por último em Áreas de Atuação. O mesmo acontece com as Funcionalidades, onde se especificam em Sub-Funcionalidades e depois em Funcionalidades Específicas. Nas demais facetas Usuários, Hardware e Software encontram-se os conceitos do tipo de usuário que o software pode satisfazer e o hardware e os softwares necessários.

Para cada *feature* é utilizada uma *template* estruturada que tem por objetivo descrever detalhadamente os conceitos do domínio. Esta estrutura é denominada padrão de domínio, similar aos padrões de documentação para o *framework* HotDraw (Johnson, 1992) e padrões para documentação de componentes de software descritos em (Ferreira et al., 1996).

A *template* descrevendo em detalhes a estrutura usada é apresentada na figura 2 e um exemplo de instanciação da *template*, descrevendo um conceito no domínio de software agropecuário, é apresentado na figura 3. Pode-se observar que no campo exemplos de instância do conceito, cada software é listado juntamente com um peso da *feature*, no caso Bovino Leiteiro, em relação à classificação do software, este peso é levando em consideração na busca realizada.

Nome: termo que caracteriza a essência do conceito. (ou funcionalidade abstrata)
Sinônimo: outros termos que tenham mesmo significado.
Tipo de conceito (Categorias, Setores,Sub-Setores, Área de Atuação, Funcionalidade,Sub-Funcionalidade, Funcionalidade Especifica, Usuário, Hardware e Software):em qual tipo o conceito se encaixa.

Fontes: lista de fontes de onde o conceito foi identificado.
Pode ser documentação, especialistas do domínio.

Descrição: descrição detalhada do conceito

Figura 2 – Template de um padrão de domínio

Nome: Bovino Leiteiro
Sinônimo: leite, gado leiteiro, gado de leite, rebanho leiteiro, vaca leiteira.
Tipo de conceito: (Categorias, Setores,Sub-Setores, Área de Atuação, Funcionalidade,Sub-Funcionalidade, Funcionalidade Especifica, Usuário, Hardware e Software):Área de Atuação.

Fontes: FARMSOFT 94; Computers in Agriculture; Guia Agrosoft
Descrição: Informações específicas sobre a criação de gado leiteiro.
Restrição:

Figura 3 – Instanciação de um padrão de domínio

Na descrição da ontologia, foi utilizado uma notação gráfica, semelhante ao modelo de classes *the Unified Modeling Language* - UML (Fowler, 1997). A única questão é que foi restringido os tipos de relacionamentos. Somente utilizou-se os relacionamentos do tipo herança e associação. Além disso, uma *feature* é representada por uma classe em UML. Utilizou-se também o conceito de pacotes de UML (***Package***) como uma forma de organizar as *features* em grupos.

Para maior esclarecimento do estudo em questão, na figura 4 foi detalhado a *package* Animais e os relacionamentos dessa *package* com as demais *features* que não pertencem a package em questão na figura 5. Nesta package encontra-se a feature, Produção Animal, que tem como objetivo representar o que há de comum em qualquer criação de animal. As demais features herdam as características desta feature genérica. Features como a dos Bovinos, que possui o que é comum em qualquer criação de bovinos e que não é comum para as demais criações, tem suas próprias especializações, a de Bovino Leiteiro e a de Bovino de Corte. Tanto a

primeira quanto a segunda possuem o que é necessário para a criação de bovinos leiteiros ou de corte respectivamente e que ainda não foi herdado de suas features superiores.

As demais classes são conceituadas de forma semelhante às descritas acima, claro que cada uma com o seu próprio enfoque, o qual é bem perceptível pelo nome das *features*.

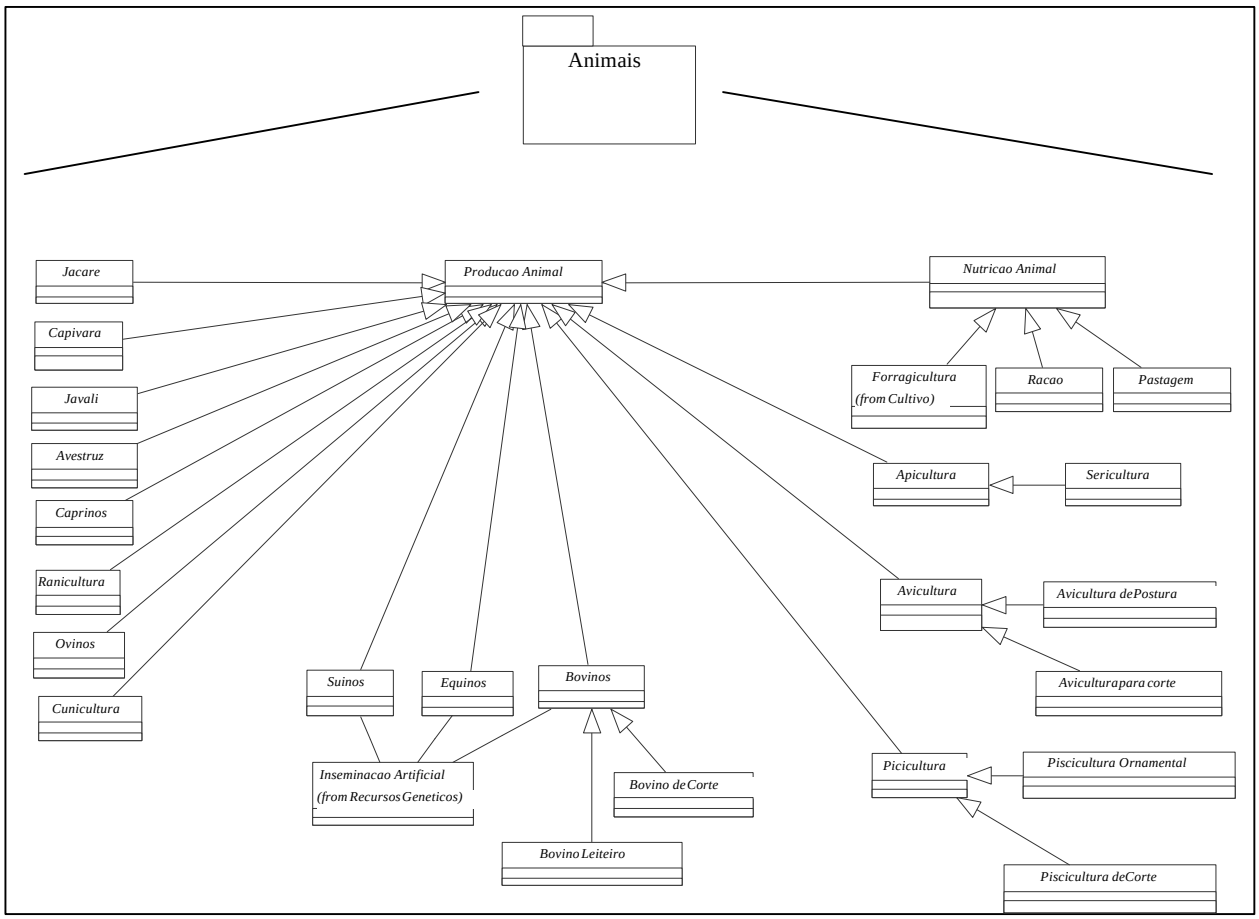

Figura 4 – Package Animais

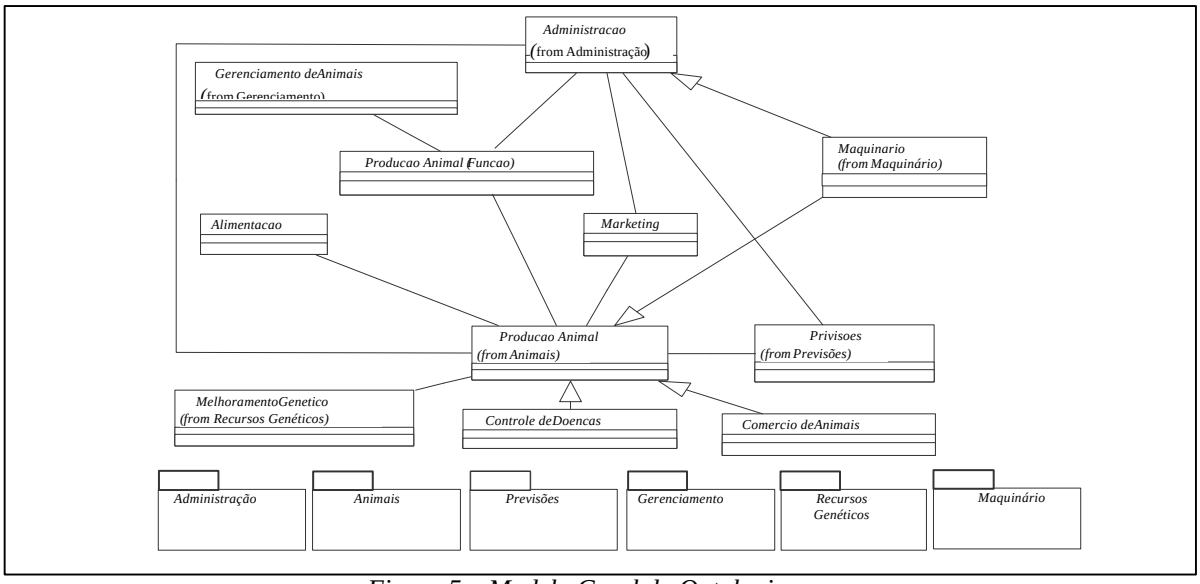

Figura 5 – Modelo Geral da Ontologia

4 – CONSIDERAÇÕES FINAIS

Foi apresentado neste artigo uma proposta de classificação de software agropecuário baseada na especificação de uma ontologia do domínio. Esta ontologia permite a classificação de um software em mais de uma categoria e o usuário poderá pesquisar o software pelas características (*features*) ou pelos sinônimos destas, sendo que o resultado desta busca deverá listar todos os softwares que possuem tais características ou similaridades a estas, baseados em pesos atribuídos a cada um dos termos ontológicos utilizados na busca, em relação aos software listados.

Esta abordagem tem como objetivo buscar solucionar os problemas encontrados nos catálogos e guias atuais e possibilitar uma maior facilidade e eficiência de busca por parte do usuário. O objetivo proposto pelo trabalho foi alcançado com, com a finalização da fase de especificação da ontologia. A próxima etapa, já em fase de desenvolvimento, é criar um servidor Web que permita a busca de software agropecuário baseado nas ontologias.

Como objetivo final, buscamos a especificação de um mecanismo de busca na Web, que através do uso da ontologia para software agropecuário, possa identificar possíveis produtos que atendam as necessidades de seus usuário com um nível de acerto superior às outras abordagens até então utilizadas.

5 – REFERÊNCIAS BIBLIOGRÁFICAS

- Guia Agrosoft, 2000 em www.agrosoft.com/guia/index.htm.
- Braga, R.M.; Werner, C.M.L.; Odyssey-DE: Um Processo para Desenvolvimento de Componentes Reutilizáveis, X CITS , Curitiba, Maio 1999.
- Falbo, Ricardo., *Integração do Conhecimento em um Ambiente de Desenvolvimento de Software*, D. Sc., thesis, COPPE, Universidade Federal do Rio de Janeiro, Dez, 1998.

- FARMSOFT'94 – International Agricultural Software Catalogue, Faculté des Sciences Agrnomiques de Gemblox , Belgique, 1994
- Ferreira, Monica, Werner, Claudia - Documenting Reusable components with patterns, ICSR4, Orlando, 1996
- Fowler M. - UML Distilled: Applying Teh Standard Object Modeling Language, Addison-Wesley, 1997
- Guarino, N., *Formal Ontology and Information Systems*, In: Guarino, N. (ed.), Proc. Of the 1st International Conference, Trento, Italy, IOS Press, 1998.
- Johnson, R; - Documenting Frameworks using patterns, OOPSLA'92
- Prieto-Díaz, R.; Implementing Faceted Classification for Software Reuse, communications of the ACM 34(5), ACM, May, 1991, pp 88-97.