

SBS-Agro: Sistema de Busca Utilizando Ontologias e Retorno do Usuário

Alexandre Procaci da Silva¹ Jean Wanderlei Alves¹
Regina Braga^{1,2} Fernanda Campos^{1,3}
{npqs@dcc.ufjf.br}

1 - Núcleo de Pesquisa em Qualidade de Software

2 – Departamento de Informática/CTU

3 – Departamento de Ciência da Computação/ICE

Universidade Federal de Juiz de Fora

Juiz de Fora, MG, Brasil, CEP: 36015-400

Abstract

In the last years, the availability of agricultural software is increasing. In this context, it is important the specification of search mechanisms on the web that allow the identification of right products for a given reality. Although there are some specific agricultural search mechanisms on the web, these mechanisms do not fulfill user's expectations. The main goal of this work is the specification of a search mechanism for the agricultural domain, using ontologies and relevance feedback techniques. We believe that using these techniques, our search mechanism can retrieve more precise results.

Keywords: Software Engineering, Databases, Agricultural Software, Ontology, Relevance Feedback, Information Retrieval, Search Mechanisms for the web.

Resumo

Nos últimos anos, a disponibilidade de software agropecuário vem aumentando significativamente. Diante desta realidade, é importante a especificação de mecanismos de busca na web que permitam aos usuários deste setor identificar os produtos adequados para a sua realidade. Apesar de já existirem mecanismos de busca específicos para a agropecuária, os mesmos não atendem de maneira satisfatória às necessidades dos usuários. Neste contexto, este trabalho tem como objetivo descrever a especificação de um mecanismo de busca por software agropecuário utilizando os conceitos de ontologia e retorno do usuário. Através da aplicação destas técnicas, acreditamos que os resultados do mecanismo de busca sejam mais precisos e voltados para as reais necessidades de seus usuários.

Palavras-chave: Engenharia de Software, Banco de Dados, Software agropecuário, ontologias, retorno do usuário, recuperação de informação, mecanismos de busca na web.

1 Introdução

A *web* é uma imensa base de dados, uma grande coleção de textos, livros, artigos, revistas, jornais, catálogos de produtos e de serviços, músicas, sons, imagens, fotografias, softwares, vídeos, filmes e peças publicitárias. Porém, nela não existe um critério padronizado ou pré-definido para a organização destas informações, que ficam disponíveis em páginas acessadas através de um endereço. Para que o usuário tenha acesso à informação que ele necessita, é preciso que ele tenha o endereço da página, ou utilize um sistema de busca, e tenha como resposta vários *links* para páginas que contenham algo relacionado ao assunto. Porém, os sistemas de busca que tratam de assuntos gerais geralmente retornam um grande número de *links* que não terão nenhuma utilidade para o usuário. Isto pode tornar-se ainda pior quando procuramos informações sobre software, principalmente, quando precisamos comparar vários produtos.

No contexto da agropecuária, a comunidade de usuários deste tipo de software enfrenta as mesmas dificuldades ao utilizar catálogos especializados, pois a grande maioria dos catálogos disponíveis [19],[8], geralmente utiliza mecanismos de busca por palavra-chave, que não são muito eficientes, por serem de difícil acesso ou não são compatíveis com o vocabulário utilizado pelos usuários, apresentando resultados pouco satisfatórios.

Considerando esse problema e analisando os resultados alcançados com o uso de ontologias no entendimento e delimitação de um dado domínio de aplicação [5], [21], [13], [20], [18] e para atender a uma demanda dos usuários e desenvolvedores de software agropecuário, foi desenvolvida uma ontologia [9], [7] para o setor de software agropecuário. Esta ontologia tem o propósito de padronizar os termos da área e ser utilizada em um sistema de busca, o Sistema de Busca por Software Agropecuário (SBS-AGRO) disponibilizado na *Web*.

Desta forma, apresentamos neste trabalho um sistema de busca para o setor de software agropecuário, utilizando as características da ontologia proposta e técnicas de retorno do usuário [1]. Com este mecanismo, pretendemos diminuir o volume de retorno de informações irrelevantes nas buscas realizadas e atender às necessidades de um maior número de usuários e desenvolvedores do setor. Esta diminuição do volume de informações irrelevantes para o usuário é conseguida, entre outros aspectos a serem detalhados ao longo do artigo, através da classificação dos softwares em relação aos termos da ontologia disponível. Estes termos estão divididos em diversos níveis e relacionados entre si hierarquicamente. Cada termo possui vários atributos que visam caracteriza-los e tornar seu significado claro para usuários, especialistas do domínio e desenvolvedores.

Este artigo se divide em cinco seções, além desta introdução. A seção dois descreve alguns mecanismos de busca de software agropecuário disponíveis na *web*, observando suas principais características. A terceira seção detalha a ontologia desenvolvida para o setor de software agropecuário. Na seção quatro analisamos os requisitos necessários para o funcionamento de um mecanismo de busca. A seção cinco detalha o mecanismo de busca SBS-AGRO, mostrando suas principais características e como pretendemos obter resultados mais precisos nas buscas realizadas pelos usuários. E, finalmente, na seção seis são apresentadas as considerações finais sobre a realização deste trabalho.

2 Catálogos e Guias para o Setor Agropecuário

Atualmente, existe uma grande quantidade de produtos de software para o setor agropecuário. Estes softwares atingem as mais diversas áreas e segmentos tanto da agricultura quanto da pecuária [19]. Entretanto, os usuários interessados em obter algum destes aplicativos podem encontrar dificuldades para selecionar o software que melhor os atenda, uma vez que a maioria dos catálogos e guias de classificação de software agropecuário utiliza mecanismos de busca não muito adequados do ponto de vista do retorno de informações realmente de interesse do usuário, tornando-se assim ineficientes.

Do ponto de vista dos desenvolvedores de software, nos catálogos existentes, os que desejam disponibilizar seu produto de software, geralmente, encontram poucas opções para a classificação dos mesmos. Como na maioria das vezes a busca é realizada através de palavras chave que se encontram na descrição dos softwares, o desenvolvedor precisa mostrar todas as áreas abrangidas pelo seu software em um texto geralmente composto por poucas linhas. Neste caso, o usuário final do sistema de busca pode encontrar dificuldades na busca, caso ele utilize um vocabulário diferente do que foi utilizado na descrição do software, pelo desenvolvedor. Além disso, baseando a busca somente nas informações fornecidas pelos desenvolvedores, o resultado pode não refletir a realidade do software, uma vez que o desenvolvedor do software tem interesses comerciais no mesmo e pode preencher estes formulários de forma tendenciosa. Para

exemplificar os catálogos de busca foram selecionados três guias de softwares do setor agropecuário, disponíveis na web, que serão descritos a seguir.

2.1 Catálogo de Software Agropecuário

A EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) no ano de 1996, iniciou um esforço na construção e disponibilização de uma base de dados contendo parte das informações existentes em suas unidades descentralizadas, utilizando a *web*. Tal ação foi coordenada e implementada pelo Centro Nacional de Pesquisa Tecnológica em Informática para a Agricultura (CNPTIA) [8] em parceria com outras unidades da EMBRAPA. O Catálogo de Software Agropecuário (www.cnptia.embrapa.br) dispõe de cerca de 200 softwares cadastrados, produzidos por mais de 100 fabricantes. O público-alvo desta base é, ao mesmo tempo, o proprietário rural, que procura no mercado um produto de software, as instituições produtoras de software e os estudiosos do mercado de informática.

O Catálogo mantido pelo CNPTIA pode ser consultado de duas maneiras distintas: por área de aplicação e por palavra-chave. Na busca por área de aplicação, os produtos foram classificados, de acordo com sua aplicação, em setor e sub-setor. Nesta opção, o usuário seleciona o setor e o sub-setor de seu interesse e obtém como resultado uma lista com todos os softwares relacionados aos itens escolhidos. A pesquisa por palavra-chave busca identificar produtos a partir de palavras constantes do campo descrição. O conteúdo deste campo é de total responsabilidade dos fabricantes, sendo, em alguns casos, muito reduzido.

2.2 Guia Agrosoft

O Guia Agrosoft do Núcleo Softex Agrosoft (www.agrosoft.com.br) [19] armazena um conjunto de informações sobre softwares agropecuários e seus respectivos fornecedores. Neste guia, os dados do software são cadastrados e atualizados pelo desenvolvedor tornando este processo muito mais dinâmico. Este guia, pode ser consultado de duas maneiras distintas: pesquisa por categoria ou pesquisa livre. Na pesquisa por categoria, os produtos são classificados, de acordo com sua aplicação, em oito categorias: Administração, Criações, Culturas, Mapeamento, Meio Ambiente, Universitários e Outras, e consiste na seleção de uma das categorias disponíveis, que resulta em uma lista dos softwares relacionados à categoria escolhida. Devido ao pequeno número de categorias pode-se ter alguma dificuldade neste tipo de pesquisa, por exemplo, ao selecionarmos a categoria “Criações” temos como resultado a listagem de 74 softwares, que incluem softwares que tratam de criação de suínos, bovinos, eqüinos, e vários outros. Sendo assim, um usuário interessado em um software para criação de eqüinos ao pesquisar por categorias, terá como resultado um grande número de softwares que não são relacionados ao tema que lhe interessa. A pesquisa livre consiste em uma busca por palavra-chave que procura identificar produtos a partir de palavras constantes em sua descrição. Neste tipo de busca podem haver distorções de acordo com o vocabulário utilizado pelo usuário, já que ele não considera o uso de sinônimos. Ao pesquisar por palavras que tratam de um mesmo assunto podemos ter resultados bem diferentes.

2.3 Boibão.com

No catálogo Boibão.com [2] (www.boibao.com.br) as informações foram agrupadas em registros. Cada registro contém diversos campos, que armazenam as informações das empresas, entidades, estabelecimentos e profissionais. Pode-se realizar busca específica por nome, razão social, e-mail, estado ou realizar a busca por palavra-chave. Na busca por palavra-chave o sistema irá procurar em todos os campos onde houver a ocorrência da palavra digitada. Esta busca é prioritária e torna sem efeito qualquer palavra digitada nos campos de busca específica. Em relação à acentuação, o Boibão foi desenvolvido para operar reconhecendo os caracteres de acentuação existentes na língua portuguesa. Para o sistema "ração" é diferente de "raçao". Como resultado da busca este sistema retorna para o usuário uma tabela contendo atributos das empresas do setor agropecuário.

2.4 Considerações sobre mecanismos de busca por software agropecuário existentes

Observando os resultados obtidos em buscas realizadas nos catálogos descritos nesta seção, podemos perceber as dificuldades encontradas pelo usuário ao procurar o software de seu interesse. Estes catálogos, geralmente, não apresentam resultados eficazes, principalmente, quando observamos os resultados obtidos ao utilizarmos a opção de busca por palavras chave. Nesta opção, o Catálogo de Software Agropecuário e o Guia Agrosoft utilizam um mecanismo semelhante, que consiste em identificar produtos a partir de palavras constantes em sua descrição. Este tipo de busca pode apresentar resultados bem diferentes uma vez que depende do vocabulário utilizado pelo desenvolvedor ao descrever o software. O Guia Agrosoft apresenta uma outra opção, a busca por categorias. No entanto, devido ao pequeno número de categorias, elas tratam de assuntos muito amplos, por isso além deste tipo de busca resultar em uma grande lista de softwares, a maioria dos softwares listados geralmente não trata do assunto de interesse do usuário.

O Catalogo de Software Agropecuário apresenta uma opção parecida com a busca por categorias do Guia Agrosoft, que neste caso é chamada de busca por área de aplicação, onde o usuário tem como resultado uma lista com os softwares relacionados ao setor e sub-setores selecionados. Com este mecanismo, temos um resultado melhor, já que o número de setores e sub-setores do Catalogo de Software Agropecuário é bem maior, o que possibilita o usuário selecionar um assunto mais específico e obter um resultado um pouco mais preciso. Apesar disso, os setores e sub-setores deste mecanismo não cobrem uma grande parte do domínio do setor de software agropecuário, como o SBS-AGRO, e nem utiliza uma ontologia especialmente especificada para o domínio, utilizando metodologia especialmente desenvolvida para tal propósito [3]. Nota-se também que não há uma descrição dos setores e sub-setores, o que pode causar equívocos tanto no cadastramento dos softwares como na seleção feita pelo usuário.

3 Ontologia para o Setor de Software Agropecuário

Guarino [18] descreve o termo ontologia, no contexto de sistema de informações, como sendo um vocabulário específico usado para descrever uma certa realidade. O conceito de ontologia tem tido importância crescente na pesquisa atual, principalmente quando o assunto é integração de fontes heterogêneas [12], [15]. O motivo pelo qual este conceito tem sido amplamente estudado é que muitas vezes ao integrar-se diversas fontes de informação, o sentido real dos conceitos utilizados não correspondem, ou até mesmo não são corretamente identificados. Dessa forma, procurou-se definir uma maneira de modelar tais conceitos, para cada domínio ou aplicação como vocabulários representativos dos objetos presentes em tais fontes e os relacionamentos entre eles.

O desenvolvimento de ferramentas para integração de fontes heterogêneas e extração de informações está tendo importância relevante nos últimos anos devido à diversidade de informações disponibilizadas no meio eletrônico. Estas informações encontram-se, na maioria das vezes, nos mais diferentes formatos, ocasionando certa dificuldade no que diz respeito ao pleno acesso a tais dados. Além disso, existe a dificuldade em encontrar a informação que obedeça ao sentido real das necessidades do usuário.

Neste contexto, o uso de ontologias no setor agropecuário permite a reutilização e o compartilhamento de um vocabulário comum entre desenvolvedores e usuários de softwares das mais diversas origens, permitindo assim uma classificação mais precisa dos mais diversos tipos de software relacionados ao domínio. Para a especificação desta ontologia de domínio foram utilizadas técnicas de engenharia de domínio, particularmente utilizou-se o método de engenharia de domínio Odyssey-DE [4], cujo objetivo principal é facilitar a reutilização de software em um dado domínio de aplicação. O modelo que especifica os termos ontológicos, formando assim uma rede semântica de termos do domínio é denominado, no contexto do Odyssey-DE em particular e na área de engenharia de domínio em geral, de modelo de *features*¹.

A abordagem do trabalho desenvolvido pode ser considerada como um esquema de vocabulário controlado, onde este vocabulário foi descrito através de uma ontologia [9]. A ontologia criada trata especificamente de termos ligados à área de software agropecuário. Em seu desenvolvimento procurou-se manter a máxima correção e precisão na escolha dos termos do domínio baseando-se na opinião de especialistas do domínio e posterior validação da ontologia por outro grupo distinto de especialistas [9]. Foram levados em consideração diversos aspectos ou abordagens pelas quais o software possa ser classificado.

Preocupando-se com a necessidade de evolução e manutenção, a ontologia (Figura 1) foi criada de maneira que novos termos possam ser adicionados de acordo com o surgimento de softwares em áreas não cobertas no plano original. Ao relacionar o software com os termos da ontologia, os avaliadores poderão cadastrar novos termos desde que alguma característica relevante do software não seja coberta pelos termos existentes. O novo termo deverá pertencer a um dos tipos já existentes (categorias, setores, funcionalidades, entre outros) e poderá ser encaixado nas “árvores” hierárquicas já existentes para que ele faça parte de uma sequência lógica, relacionando-se com os outros termos da ontologia.

Com o objetivo de permitir uma classificação do software baseada em um conjunto mais amplo de termos, foram criadas facetas e para cada faceta foram detalhados os conceitos classificadores, adquirindo assim uma precisão mais apurada e flexibilidade na classificação. Na classificação desenvolvida [9], existem duas grandes facetas: **Categorias**, onde se encontram as grandes características do software e **Funcionalidades**, onde estão as possíveis funcionalidades mais

¹ Um modelo de *features* [10] é um modelo que captura as características (*features*) gerais das aplicações desenvolvidas em um domínio. Assim, as similaridades e diferenças entre as aplicações são explicitadas através das *features*, classificadas de forma hierárquica no modelo.

gerais do software. Para cada termo é utilizada uma *template* que tem por objetivo descrever detalhadamente os conceitos do domínio (Figura 2).

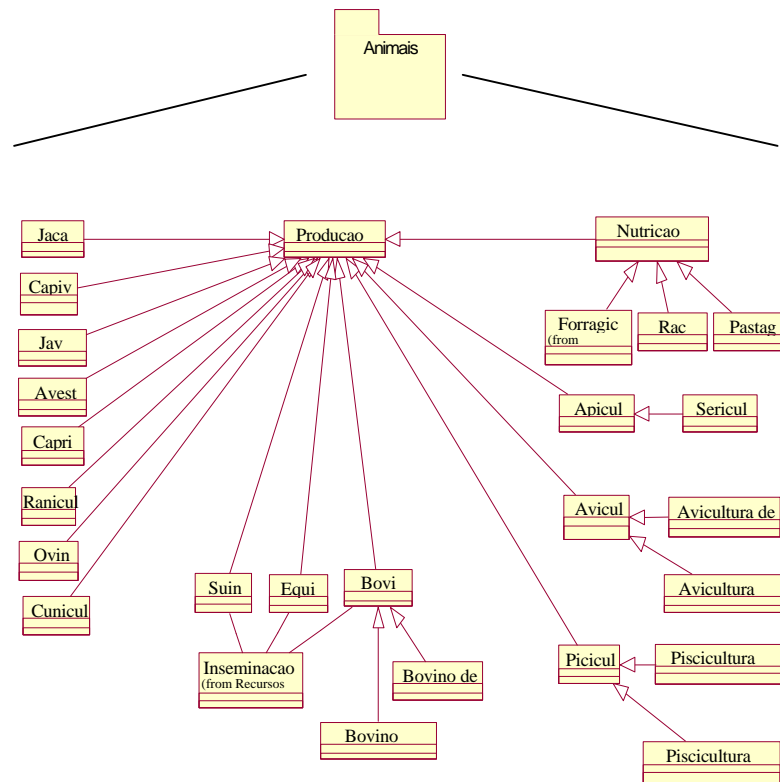


Figura 1 – Ontologia para o domínio de software agropecuário (visão parcial)

Nome: nome que caracteriza a essência do termo.
Tipo de conceito (Categorias, Setores, Sub-Setores, Áreas de Atuação, Funcionalidades, Sub-Funcionalidades, Funcionalidades Específicas): em qual tipo o termo se encaixa.
Fontes: lista de fontes onde o termo foi identificado (pode ser documentação e/ou especialistas do domínio).
Descrição: Descrição detalhada do termo.
Sinônimos: Lista de sinônimos conhecidos para o termo
Restrição: restrições que podem ocorrer quando o conceito for usado.
Softwares relacionados ao termo: produtos de software já relacionados ao termo, conjuntamente com o grau de relacionamento definido pelos avaliadores (consultar seção 4 deste artigo)

Figura 2 – Template de termo ontológico (adaptado de [9])

4 Classificação dos Softwares na Ontologia

Observando as deficiências dos catálogos analisados na seção 2, optou-se por especificar um mecanismo de busca utilizando a ontologia descrita. Com este mecanismo pretende-se melhorar os resultados obtidos nas buscas por palavras-chave, além de apresentar a proposta de uma nova opção para o usuário, chamada de busca em árvore. Analisando os requisitos necessários para o funcionamento de um mecanismo de busca, foi criado um banco de dados, onde são armazenados todos os dados utilizados no mecanismo de busca, ou seja, cadastramento dos termos da ontologia, incluindo os dados da *template* e os dados relativos aos produtos de software. O processo de cadastramento e atualização dos softwares foi desenvolvido de forma que o desenvolvedor possa realizar a atualização dos dados do seu software diretamente pela *Web*. Dessa forma podemos tornar esse processo bastante dinâmico, possibilitando que suas informações sejam atualizadas sempre que surgir uma nova versão ou for constatado algum erro nos dados do software. É também preciso que o software seja classificado segundo os termos da ontologia. Para estabelecer o grau de relacionamento entre o software e cada um dos termos são necessárias avaliadores que tenham um bom conhecimento do software e das áreas por ele abrangidas. Assim, estes avaliadores examinam os softwares e os classificam, sendo que

cada software pode ser classificado por vários avaliadores. O cadastro dos softwares é realizado através de um formulário disponibilizado na *Web* constituídos dos seguintes campos:

| Campo | Descrição |
|---------------|---|
| Nome | Nome do Software. |
| Versão | Versão do Software que esta sendo cadastrado. |
| Desenvolvedor | Empresa que desenvolveu o software |
| Contato | Meio de contato com desenvolvedor, neste campo poderá ser preenchido com telefone, e-mail, homepage ou qualquer outro meio de contato que o usuário possa utilizar. |

Quadro 1- Dados do Cadastro de um Software

Para que possamos fazer a busca é necessário que os softwares estejam relacionados aos termos da ontologia. Logo, faremos uso da experiência dos avaliadores que deverão conhecer o software e também o domínio por ele abrangido ou ter alguma experiência prática utilizando-o. Baseados nessas informações, identificamos três tipos de avaliadores:

Desenvolvedores: Desenvolvedores de produtos de software agropecuário. Este tipo de usuário, além de ter a possibilidade de cadastrar seus produtos, também tem a possibilidade de classificar seu produto de acordo com a ontologia disponível. Esta classificação somente se refere aos produtos por ele cadastrados e segue os mesmos parâmetros para atribuição de valores dos outros usuários, porém o peso atribuído a esta classificação é menor do que a classificação realizada pelos outros usuários.

Especialistas: Usuários especialistas no domínio que o software se aplica, tais como pesquisadores da área agropecuária, agrônomos, veterinários, zootecnistas, entre outros. O especialista deve ser capaz de avaliar se o software atenderá os usuários no que esta sendo proposto pelo desenvolvedor e se as informações contidas nele são coerentes. Para estes usuários é disponibilizada uma interface na *web*, acessada através de senha, que permite ao usuário classificar, através de atribuição de valores, o nível de ligação semântica que determinado produto deve ter com determinados termos do domínio. Para facilitar o trabalho do especialista, todas as vezes que um novo produto é cadastrado para ser disponibilizado através do mecanismo de busca, ele é notificado para que sejam atribuídos os pesos para as ligações. Uma classificação prévia de quais termos da ontologia teriam maior ligação com o produto, é apresentada a este usuário. Esta classificação prévia é baseada na descrição do produto feita pelo desenvolvedor. A classificação deste tipo de usuário é considerada a mais relevante para o sistema quando da realização da busca.

Usuários: usuário final, é o usuário que utiliza o mecanismo de busca para procurar produtos de software agropecuário que atendam suas expectativas. Este usuário, utilizando a ontologia do domínio disponibilizada através da *web*, seleciona os termos mais relacionados à busca, de acordo com as facetas de interesse.

Para que o avaliador classifique o software, primeiramente ele deve cadastrar-se, o que consiste no preenchimento um formulário, disponibilizado também na *Web*, com os seguintes dados (quadro 2).

| Campo | Descrição |
|--------------|--|
| Nome | Campo com o nome do avaliador; |
| Instituição | Empresa/instituição na qual o avaliador trabalha; |
| Contato | Campo preenchido com alguma forma de contato com o avaliador com por exemplo telefone, e-mail, etc |
| Tipo | Tipo no qual o avaliador se enquadra: especialista, desenvolvedor ou usuário. |

Quadro 2 - Dados do Cadastro de um Avaliador

Observando as características de cada tipo de avaliador foi decidido que a classificação feita por cada um deles deveria ter um grau de importância diferente (quadro 3).

| Tipo de avaliador | Grau de relevância |
|--------------------------|---------------------------|
| Desenvolvedor | 2 |

| | |
|--------------|---|
| Usuário | 3 |
| Especialista | 5 |

Quadro 3 - Grau de relevância da avaliação conforme o tipo de avaliador

O grau de importância é um peso a ser atribuído para a classificação realizada pelo avaliador, que será levado em consideração no cálculo da média do nível de relacionamento dos softwares com os termos. Na classificação, o avaliador seleciona o nível de relacionamento do software com cada termo da ontologia, começando pelas categorias, até as funcionalidades específicas. No processo de classificação são utilizados cinco níveis de relacionamento, que estão listados abaixo.

| Nível de Relacionamento | Peso |
|-------------------------|------|
| Totalmente Relacionado | 4 |
| Altamente Relacionado | 3 |
| Relacionado | 2 |
| Pouco Relacionado | 1 |
| Não Relacionado | 0 |

Quadro 4 – Relacionamento dos softwares com os termos da ontologia.

5 O mecanismo de busca SBS-AGRO

O SBS-AGRO, associado à ontologia descrita em [9], procura traduzir as necessidades tanto dos produtores quanto dos compradores de software agropecuário em um mecanismo capaz de reunir os diversos vocábulos encontrados no setor para uma busca por palavra-chave mais eficaz, e o uso dos termos de uma ontologia, para uma busca através da navegação pelos termos da ontologia.

Para que este sistema funcione de forma organizada e satisfatória, foram desenvolvidos subsistemas que se direcionam às diferentes funções do mecanismo como um todo. Estes subsistemas são:

- **Software:** que trata das operações relacionadas aos softwares que serão os resultados das pesquisas no mecanismo de busca;
- **Taxonomia,** que trata das operações relacionadas à ontologia utilizada para se classificar os softwares cadastrados;
- **Avaliação,** que trata das operações de avaliação dos softwares em relação à taxonomia por diferentes avaliadores.

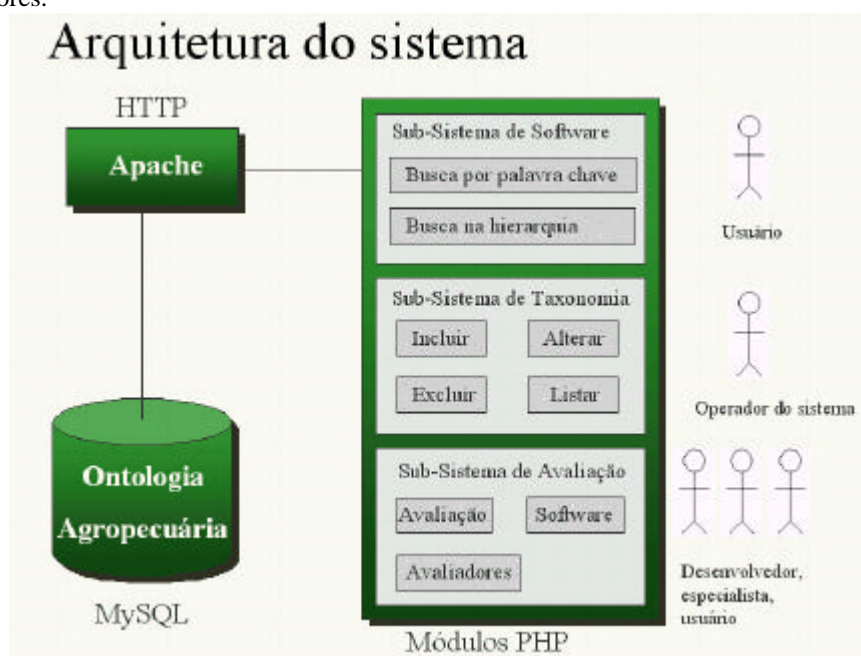


Figura 3– Arquitetura do SBS-AGRO

Na Figura 3 é apresentado um modelo básico da arquitetura do sistema. Foi criada uma base de dados que comportasse as informações a serem armazenadas incluindo o desenvolvimento de aplicações para entrada de dados. Para todas essas bases foram geradas aplicações para consulta usando recursos da *web*. Para a criação da base de dados foi utilizado o MySQL(www.mysql.org) como servidor de banco de dados. Para acessar os dados armazenados no servidor fizemos uso da linguagem de *script* PHP (www.php.org), embutida no HTML.

5.1 Modelos de Busca no SBS-AGRO

5.1.1. Busca por Palavras-chave

A pesquisa por palavra-chave (Figura 4) busca identificar os termos da ontologia relacionados às palavras digitadas pelo usuário na interface do mecanismo de busca, a partir da ocorrência dessas palavras no campo **sinônimos** da *template* (Figura 2) dos termos. Após identificados os termos, verificamos quais os softwares estão relacionados a estes termos. Estes softwares serão organizados em ordem decrescente de acordo nível de relacionamento com os termos, e então serão transmitidos ao usuário como resultado. O nível de relacionamento entre os softwares e os termos é definido através da fórmula apresentada no quadro 4.

$$NR = [(P1 * R1) + (P2 * R2) + \dots + (Pn * Rn)] / na$$

Onde:

na – é o número de avaliadores que classificaram o software separadamente;

NR – é o nível de relacionamento entre o software e o termo selecionado;

Pn – é o nível de relacionamento entre o software e o termo (quadro 4) selecionado determinado na avaliação **n**;

Rn – é o grau de relevância da avaliação **n** observado o tipo de avaliador responsável por ela;

Quadro 5– Fórmula que calcula o nível de relacionamento entre os softwares e os termos da ontologia.

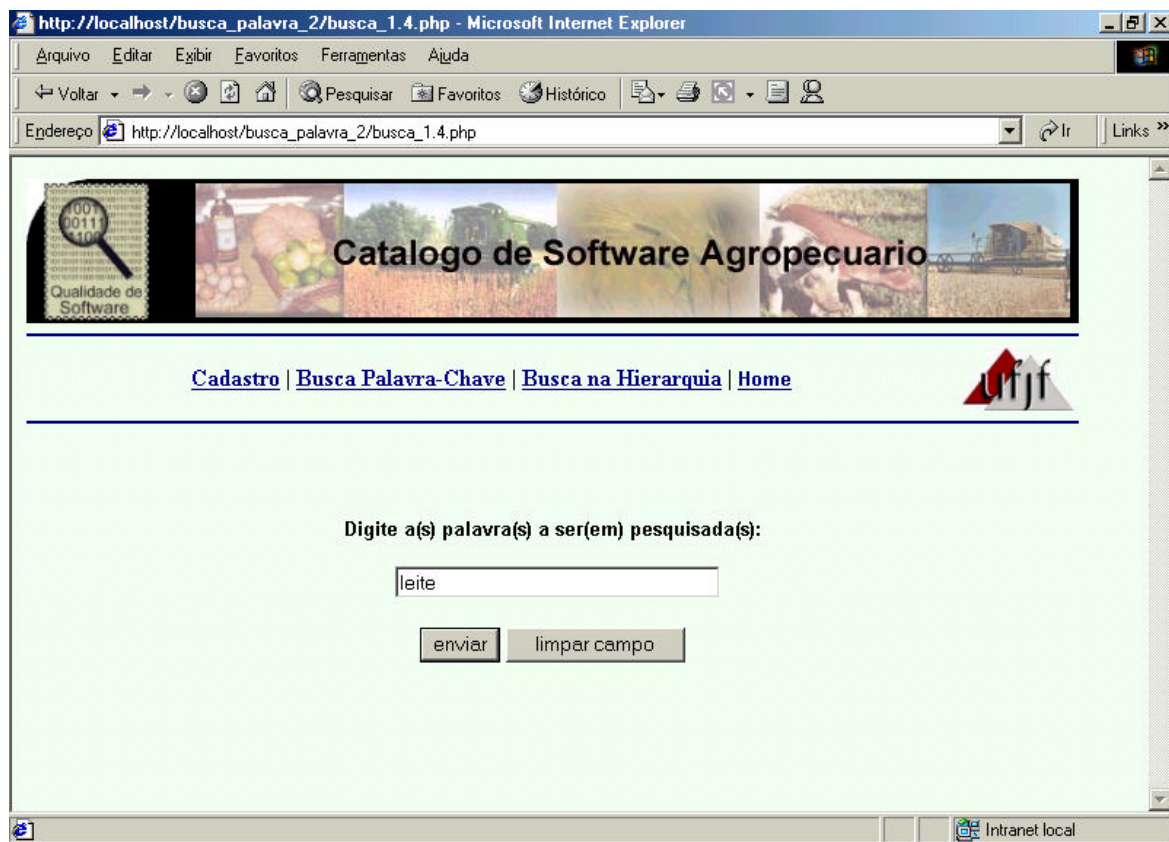


Figura 4 – Busca por palavras chave.

Por exemplo, ao digitarmos “leite” como palavra-chave (Figura 4), o mecanismo de busca pesquisará no campo sinônimos de todos os termos da ontologia, a ocorrência desta palavra e encontrará o termo “bovino leiteiro”, logo, serão identificados todos os softwares relacionados a este termo. Supondo que sejam encontrados dois softwares, chamados

Agro1 e Agro2 e que estes apresentem como nível de relacionamento (NR) os valores 2 e 4, respectivamente. Assim o usuário terá como resultado uma lista com os dois softwares, sendo que o software Agro2 será exibido primeiro pois está mais relacionado ao termo (Figura 5). Com este exemplo, notamos que ao utilizar a busca por palavras-chave, o usuário poderá utilizar um vocabulário diferente do utilizado pelo desenvolvedor, já este mecanismo considera os sinônimos dos termos utilizados na classificação do software, e não as palavras encontradas em sua descrição.

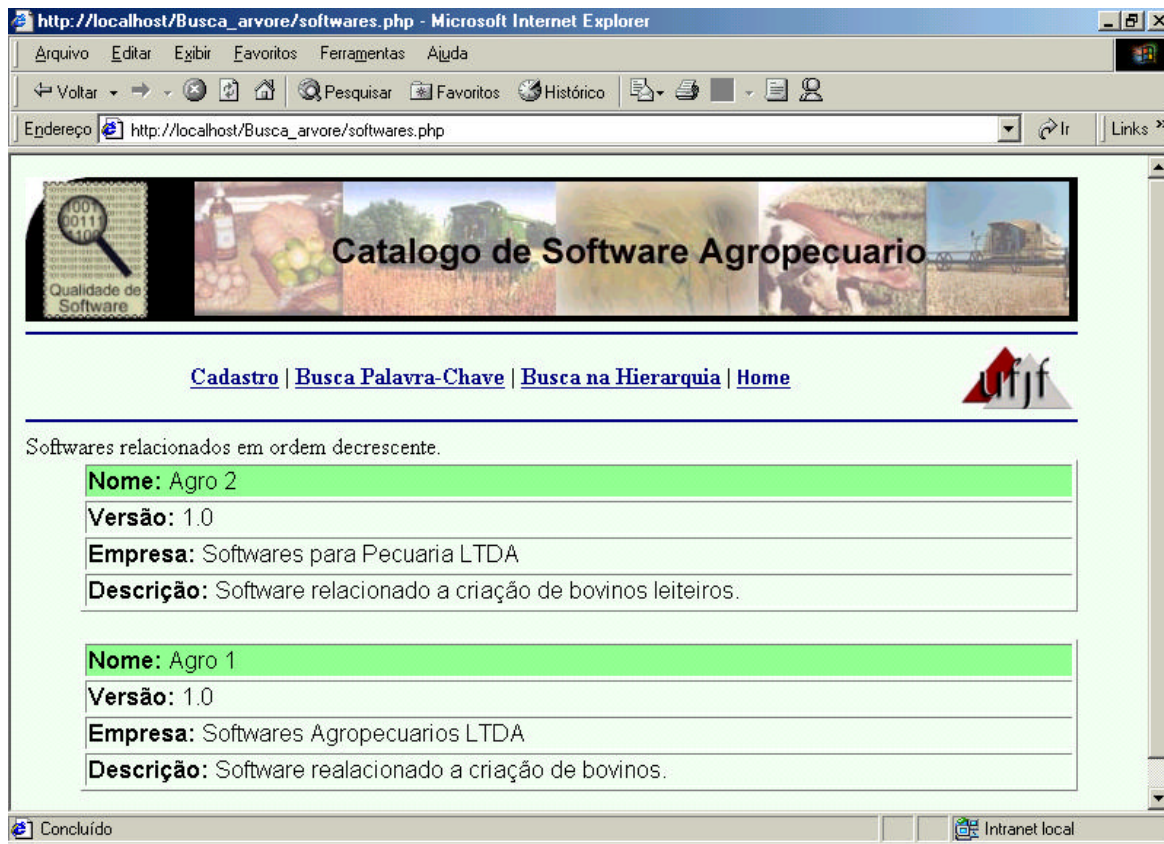


Figura 5– Resultado da busca por palavras-chave.

5.1.2 – Busca em árvore

Na busca em árvore o usuário “navega” através dos níveis hierárquicos da ontologia, selecionando em cada nível, os termos de seu interesse. Com o conjunto de termos selecionados o mecanismo fornecerá como resultado uma lista de softwares relacionados a estes termos. Ao escolher a busca em árvore, o usuário é direcionado para uma página que contém todas as categorias, nesta página ele irá selecionar as categorias relacionadas ao software desejado e logo depois, poderá selecionar os setores, sub-setores, e assim por diante. Na opção das funcionalidades, o mecanismo listará todos os termos deste tipo, mostrando em seguida as sub-funcionalidades relacionadas às funcionalidades selecionadas. Após a seleção de todos os termos o mecanismo de busca exibirá todos os softwares relacionados aos termos escolhidos, em ordem decrescente em relação ao nível de relacionamento.

Por exemplo, ao procurarmos um software relacionado à criação de bovinos leiteiros, escolhemos primeiramente as categorias (Figura 6), onde selecionamos a categoria “produção animal”, passando para os setores podemos selecionar “comércio de animais”, nos sub-setores selecionamos “bovinos” e nas áreas de atuação selecionamos “bovino leiteiro”, neste caso podemos selecionar “mostrar softwares”, pois os termos escolhidos identificam o software desejado, porém poderíamos continuar selecionando as funcionalidades, sub-funcionalidades e funcionalidades específicas. Finalizando a busca o mecanismo seleciona todos os softwares relacionados com os termos escolhidos e exibe uma lista (a exemplo da Figura 5) com todos eles ordenados pelo nível de relacionamento.

Neste exemplo, notamos que a ontologia é bastante específica permitindo que o usuário selecione detalhadamente as características do software desejado. Além de apresentar como resultado, uma lista de softwares, ordenada pelo grau de

relacionamento entre software e os termos selecionados, fazendo que, na maioria das vezes, o software mais adequado para o usuário esteja entre os primeiros da lista.

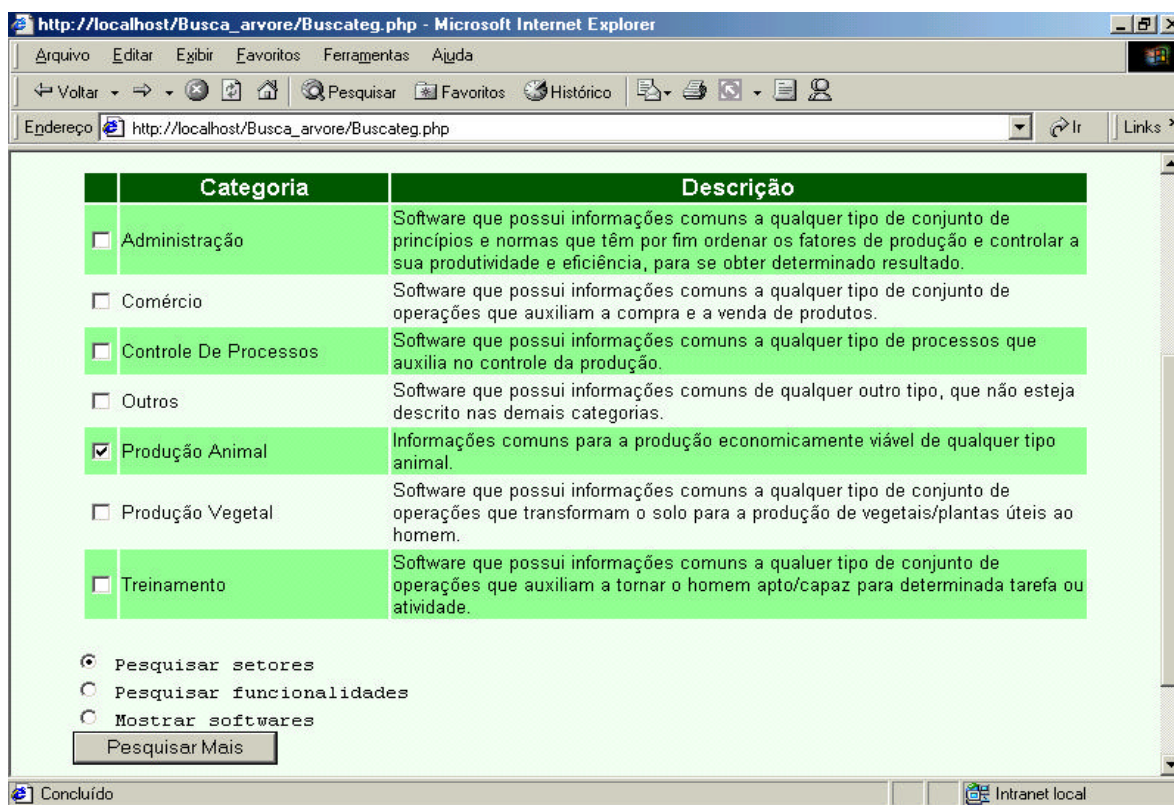


Figura 6 – Busca em árvore, seleção das categorias.

6 Considerações Finais

Atualmente, pesquisadores, professores e pessoas interessadas no setor agropecuário buscam diminuir a resistência desta comunidade na adoção de novas tecnologias. Existem, na *web*, vários *web sites* onde são apresentados diversos softwares voltados para o setor de software agropecuário. Com o objetivo de facilitar a obtenção destes softwares por parte dos agropecuaristas, alguns destes disponibilizam sistemas de busca dedicados a este setor. Porém, a maior parte destes mecanismos de busca não é muito eficiente, exigindo um esforço maior do usuário e apresentando uma grande quantidade de resultados irrelevantes. Um sistema de busca para softwares agropecuários mais eficiente pode contribuir bastante para o processo de informatização do setor agropecuário, reduzindo consideravelmente o esforço do usuário na procura de um produto que o atenda.

Este trabalho apresentou o sistema de busca SBS-AGRO que utiliza ontologia e retorno do usuário (*relevance feedback*) [1] para a classificação de softwares agropecuários. A utilização destas tecnologias tem como principal objetivo diminuir o volume de retorno de informações irrelevantes nas buscas realizadas e atender as necessidades de um maior número de usuários e desenvolvedores. A vantagem de se utilizar ontologias neste contexto é que elas permitem a reutilização e o compartilhamento de um vocabulário comum para a classificação de softwares de diferentes empresas. Este trabalho foi desenvolvido de forma que a classificação dos softwares possa ser incrementada à medida que os usuários fizerem sua avaliação dos softwares, deste modo poderemos obter resultados cada vez mais precisos.

Com o intuito de verificar a adequabilidade e precisão do SBS-AGRO em relação as buscas realizadas, o mesmo foi validado utilizando amostras do Guia Agrosoft (www.agrosoft.com) e simulações de usuários finais, utilizando para a classificação dos produtos os resultados obtidos na avaliação destes produtos para o Selo Agrosoft de Qualidade [6] e processo de avaliação da qualidade que alguns produtos se submeteram em parceria com o Núcleo de Pesquisa em Qualidade de Software (www.npqs.dcc.ufjf.br). Com base nos resultados obtidos, o Guia Agrosoft será reformulado para incorporação da proposta apresentada.

Referências Bibliográficas

- [1] Baesa-Yates, R., Ribeiro-Neto, B., 1999, Modern Information Retrieval. Addison-Wesley
- [2] Boibão.com disponível em: www.boibao.com.br/ acessado em 24 de fevereiro de 2002
- [3] Braga, R. M., 1999; Werner, C.M.L.; Odyssey-DE: Um Processo para Desenvolvimento de Componentes Reutilizáveis, X CITS , Curitiba, Maio 1999.
- [4] Braga, R.M., 2000; Busca e Recuperação de Componentes em Ambientes de Reutilização de Software, Tese de Doutorado, COPPE/Sistemas/UFRJ.
- [5] Braga, R., Werner, C.; Mattoso, M.,2001: “The Use of Mediation and Ontology Technologies for Software Component Information Retrieval”, Proceedings of ACM Symposium on Software Reusability (SSR’01), Toronto, May, pp.19-28.
- [6] Campos, F., Braga, R., Coelho, F. Quality Process to Improve Agricultural Software Products. FEESMA 2000m Germany, 2000.
- [7] Campos, F.; Braga, R.; Alves, J.; Procaci, A., 2002: Um mecanismo de busca por software agropecuário usando ontologias e Retorno do Usuário, Simpósio Brasileiro de Informática Aplicada a Agropecuária, Foz do Iguaçu, março, 2002.
- [8] Catálogo de Software Agropecuário em: www.cnptia.embrapa.br/ acessado em 24 de fevereiro de 2002
- [9] Coelho, F. M., Setor de Software Agropecuário: Proposta de uma Ontologia. Universidade Federal de Juiz de Fora, julho, 2000. disponível em www.npqs.dcc.ufjf.br.
- [10] Cohen, S, 1994, Feature-Oriented Domain Analysis: Domain Modeling, Tutorial Notes; 3rd Int. Conference on Software Reuse, Rio de Janeiro, Novembro.
- [11]Cruz, S. A. B.,Gonçalves, L. (2002), Projeto Base de Dados da EMBRAPA, Disponível: Site EMBRAPA (2002), URL: www.cnptia.embrapa.br/bases Consultado em 24 de fevereiro de 2002.
- [12] Dorneles, C. F. 2000. Extração de Dados Semi-estruturados com base em uma ontologia.
- [13] Falbo, R., 1998. Integração do Conhecimento em um Ambiente de Desenvolvimento de Software, D. Sc., thesis, COPPE, Universidade Federal do Rio de Janeiro, Dez, 1998.
- [14] Galotta, C. et al., 1999. Ambientes de Desenvolvimento de Software Orientados a Domínio. D. Sc., thesis, COPPE, Universidade Federal do rio de Janeiro, Abril, 1999.
- [15] Giuizzardi, G., Falbo, R. A, Pereira Filho, J. G. Using Objects and Patterns to Implement Domain Ontologies. Simpósio brasileiro de Engenharia de Software, Rio de Janeiro, Out. 2001.
- [16] Gonzaga, F. S. & Birckan, G., PHP e MySQL, Florianopolis – SC, outubro de 2000.
- [17] Gruber, T. R., 2001. *What is an ontology?*, documento online <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
- [18] Guarino, N., *Formal Ontology and Information Systems*, In: Guarino, N. (ed.), Proc. Of the 1st International Conference, Trento, Italy, IOS Press, 1998.
- [19] Guia Agrosoft em www.agrosoft.com.br, acessado em 24 de fevereiro de 2002
- [20] Nieto, E. M., OBSERVER: *An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies*, Tesis Doctoral, Departamento de Informatica e Ingenieria de Sistema, Universidade de Zaragoza, Nov, 1998.
- [21] Oliveira, K. M., Modelo para Construção de Ambientes de Desenvolvimento de Software Orientados a Domínio, D. Sc., thesis, COPPE, Universidade Federal do Rio de Janeiro, Out, 1999.
- [22] Stumme G., 2001, Workshop on “Ontologies”, 18 de setembro de 2001, Viena, Austria, documento online <http://onto2001.aifb.uni-karlsruhe.de/>, 2002.

Agradecimentos: Agrademos a colaboração do Projeto OdysseyShare da COPPE/Sistemas/UFRJ, da equipe do Softex Agrosoft e Embrapa-CNPGL.