

CATÁLOGO DE RECURSOS EM INFORMAÇÃO AGROPECUÁRIA NA INTERNET: DESENVOLVIMENTO DE WEBSITE

Laurimar Gonçalves Vendrusculo¹, Marcia Izabel Fugisawa Souza²

1. Resumo

O presente trabalho sintetiza o processo de desenvolvimento do site do Catálogo de Recursos em Informação Agropecuária (CRIA) na Internet. Enfoques relativos ao *design* do site, organização e armazenamento da informação são descritos. O site CRIA pretende contribuir para a qualificação, a disseminação e o acesso a recursos de informação demandada, pelos diversos segmentos do agronegócio, utilizando a rede Internet como infra-estrutura de comunicação.

Termos para indexação: repositório de informação; informação agropecuária; aplicativo na internet; website.

Index terms: information repository; agricultural information; internet's applicative website.

2. Introdução

O aumento exponencial dos usuários na Internet nos últimos anos e a crescente complexidade dos serviços oferecidos pela World Wide Web (WWW) se apresentam como oportunidade e desafio à difusão da informação. A Web é uma forte concentradora de interesses, o que resulta na maximização de esforços de ações conjuntas.

Os mecanismos de busca existentes fornecem meios de buscas "horizontais" e genéricas. Consultas feitas a estes mecanismos podem frustrar a expectativa de usuários que necessitem encontrar informações especializadas em determinadas áreas (Lea-Cox, 1998). Desenvolver serviços de qualificação de links especializados e recursos de informação apresenta-se, então, como demanda prioritária a vários segmentos. Segundo Lewis (1998), a agregação de valores aos recursos de informação permite não somente a herança dos dados mas o estabelecimento de novos relacionamentos entre as partes.

A Embrapa Informática Agropecuária iniciou o processo de desenvolvimento do site de Catálogo de Recursos em Informação Agropecuária (CRIA), atualmente disponível em <http://www.cria.org.br>, com o qual pretende contribuir para a qualificação, a disseminação e o acesso à informação demandada pelos segmentos do agronegócio, agroindustrial e florestal. A rede Internet foi utilizada como infra-estrutura básica e o diagrama básico do site é apresentado pela Fig. 1.

O propósito deste site é a criação de repositórios de informação qualificada que contemplem: documentos/publicações com texto na íntegra; coleção de links especializados e genéricos; informações sobre mercado agrícola, dentre outras. Pretende-se que este site ofereça serviço de informação qualificada, que assegure agilidade e facilidade no acesso e na recuperação, visando satisfazer as necessidades de informação agropecuária do cliente, garantindo economia de tempo e racionalização de custos, modernizando o processo de transferência de tecnologia com a disponibilização de informações precisas e atualizadas.

¹ Pesquisadora da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
(laurimar@cnptia.embrapa.br)

² MSc, Técnico de Nível Superior, Embrapa Informática Agropecuária. (marcia@cnptia.embrapa.br)

Os temas abordados pelo CRIA são:

1. **Ação governamental** - inclui informações sobre legislação agropecuária, políticas agrícolas e programas nacionais implementados pelo Governos Federal e Estadual e seus agentes.
2. **Agroinformática** - inclui informações sobre instituições, pesquisadores, listas de discussão, software de demonstração, *freeware* e *shareware*, catálogos de software, revistas online etc.
3. **Conjuntura** - contém informações relacionadas a safras, tendências de mercado, cotação de preços e indicadores conjunturais agropecuários.
4. **Tecnologias** - engloba informações técnico-científicas e informativas além de prover um ambiente para trabalho colaborativo entre pesquisadores e especialistas.
5. **Agenda de eventos agropecuários** - apresenta classificados de diferentes tipos de eventos (feiras, leilões, simpósios, congressos etc).
6. **Meteorologia** - reúne links de instituições que veiculam informações sobre o assunto.
7. **Instituições** - contempla reunião de links qualificados de sites de instituições que atuam na agropecuária e agroindústria.

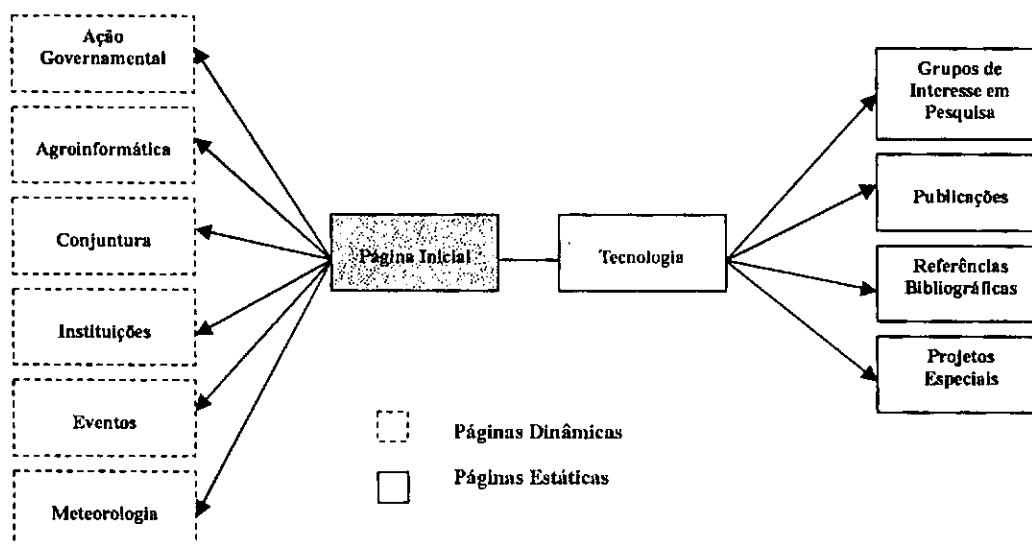


FIG. 1. Diagrama do Site CRIA.

3. Design do Site

O CRIA é um site multi-institucional com enfoque em informação, por isto foi projetado segundo Siegel (1996), seguindo os preceitos dos sites chamados de *terceira geração*. Utilizou-se os seguintes atributos no processo de *design* e desenvolvimento:

- uso da metáfora de um catálogo tradicional de produtos - pretende-se que esse modelo, usado no projeto visual, guie e integre o usuário ao site;
- indexação através de ferramenta de consulta - utilizou-se o *Enterprise Server* para consulta a documentos estáticos e consultas através da linguagem SQL às bases de dados;
- tratamento de imagens visando redução do tamanho e complexidade; definição de conjunto reduzido de cores para melhor caracterização de um site institucional;
- diagramação do texto utilizando recursos como: GIF transparente e larguras absolutas para tabelas;
- adoção de *frames* (molduras) com a finalidade de orientar o usuário;
- redução da rolagem nas páginas iniciais do site;
- limitação do tamanho das páginas (tamanho máximo de 70 kb), com objetivo de acelerar a carga dos arquivos;
- utilização de meta *tags* na página inicial através das *tags: description, keywords e distribution*, as quais facilitam a indexação do site por robôs e/ou mecanismos de busca externos.

4. Armazenamento e Organização da Informação

As informações apresentadas pelo site estão armazenadas, atualmente, de três maneiras:

- através de bases de dados;
- em documentos no formato PDF (Portable Document File) da Adobe System; e
- páginas HTML estáticas.

4.1. Bases de Dados Atuais

As bases existentes foram desenvolvidas utilizando o sistema de gerenciamento de banco de dados (SGBD), Microsoft Access 97. São elas:

- Base de Links - Armazena as informações contidas nos tópicos: ação governamental, agroinformática, conjuntura, instituições e meteorologia. Os links disponíveis nestes tópicos são qualificados segundo os atributos mostrados pela Fig. 2.

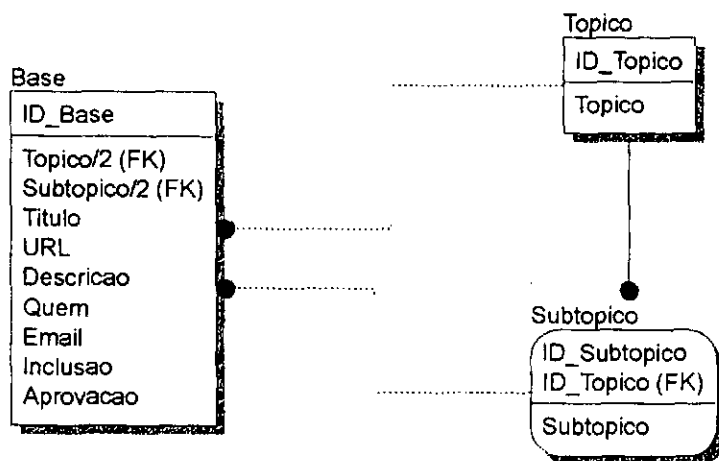


FIG. 2 Esquema de dados associado a base de Links.

- Base de Eventos - organiza um repositório de informações relacionadas aos eventos agropecuários nas suas diversas modalidades. Note-se que na Fig. 3 os atributos Aprovado e Expirado controlam a disponibilização dos eventos no site.

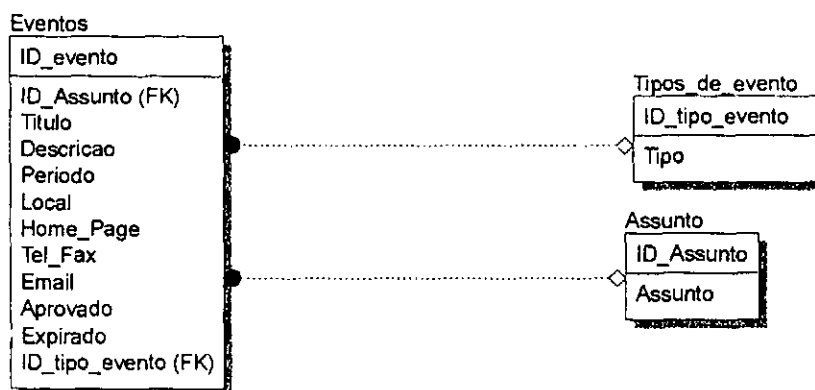


FIG. 3. Esquema de dados associado a base de Eventos.

4.2. Documentos PDF

Reúne uma coletânea de documentos e publicações técnicas, com textos na íntegra, disponíveis no item Publicação do tópico Tecnologia. Em sua maioria, os textos são publicações técnicas da Embrapa e foram selecionados obedecendo a critérios de qualidade e pertinência. Os autores são consultados sobre a disponibilização de seus documentos na Web.

Optou-se por gerar arquivos no formato Portable Document File (.pdf) em função da crescente tendência de disponibilização de arquivos deste tipo na Internet e do uso gratuito do seu respectivo visualizador (Acrobat Reader)³.

Os instrumentos utilizados no projeto de digitalização de documentos neste projeto são:

- *Scanners*: Umax Modelo: UC1260;

³ Ferramenta da Adobe System Incorporated.

•Microcomputador 486 DX2 - 16 Mb de RAM- 800 Mb de disco rígido;

•Software Adobe Acrobat 3.0

As etapas do processo de digitalização constituem-se de:

- a) seleção do material a ser digitalizado;
- b) parametrização do Software *Adobe Acrobat Exchange*³;
- c) processo de digitalização;
- d) conversão e tratamento de texto.

a) Seleção de Material

São selecionadas publicações para digitalização, levando-se em conta a sua qualidade e a pertinência do tema abordado. Esses documentos podem ser comunicados, relatórios e instruções técnicas de cunho basicamente informativo.

Os textos em preto e branco são separados daqueles coloridos e com fotos, objetivando a otimização do processo de digitalização em função da maior simplicidade do primeiro tipo.

b) Parametrização

O pacote utilizado para aquisição das imagens e filtragem do texto é o *Adobe Acrobat 3.01 for Windows*. Procede-se a parametrização no módulo *Adobe Exchange* em função dos textos selecionados. Os principais parâmetros adotados neste processo são destacados na Tabela 1.

TABELA 1. Parâmetros utilizados no Módulo *Adobe Exchange*.

Parâmetros	Significado	Valores usados
<i>Mode</i>	presença ou ausência de cores	<i>Lineart</i> – Para textos em preto e branco 256 cores. Para textos coloridos e com imagens
<i>Original</i>	tipo de papel	<i>Reflective</i>
<i>Input</i>	qualidade da imagem adquirida	200 PPI 's
<i>Scale to</i>	proporção do tamanho da página obtida	100%

³ Ferramenta da Adobe System Incorporated.

c) Processo de digitalização

Os textos são digitalizados, inicialmente, folha a folha, diretamente do *scanner*, através do Módulo *Adobe Acrobat Exchange*.

d) Conversão e tratamento de texto

O produto inicial do processo de digitalização são as imagens; em seguida, faz-se o reconhecimento/conversão das imagens *bitmap* em caracteres textuais de forma a poder tratá-las, melhorando sua qualidade e tornando-as pesquisáveis. Esta fase se deu através do *Adobe Acrobat Exchange* o qual submete a imagem ao filtro OCR (*Optical Character Recognition*).

Encerrado o processo de conversão o *Acrobat* aponta as palavras, ainda como imagem, as quais não consegue identificar. A correção é feita manualmente, item a item. O texto pode requerer também algumas mudanças no tipo ou tamanho das letras. Trata-se da fase com maior dispêndio de tempo.

Como resultados obtém-se arquivos em formato .pdf, com tamanho e fidelidade aceitáveis, principalmente por conservar a diagramação original do texto.

5. Aplicações Online para Consulta a Bases de Dados

O sistema que possibilita a disponibilização dos dados armazenados em bases constitui-se de aplicações utilizando comandos HTML, JAVA applets e a ferramenta Utilis 1.0⁴ (Fig.4). A metodologia usada para o desenvolvimento de aplicações de consulta a bases de dados é descrita por Moura & Oliveira (1998).

Sucintamente, são descritos os passos executados pelo CGI Utilis, no caso do processo de consulta a banco de dados:

1. recebe o documento denominado metaformulário, o qual é formado de comandos HTML e *tags* especiais. As *tags* formam uma linguagem específica interpretada pelo CGI.
2. interpreta os comandos de sua linguagem de programação (*tags*) presentes no metaformulário; ativa o sistema gerenciador de banco de dados através do driver ODBC;
3. ativa o sistema gerenciador de banco de dados através do driver ODBC;
4. repassa ao SGBD as questões SQL que constam do metaformulário. O SGBD executa as atividades de seleção ou inserção nas bases;
5. acessa os resultados de consulta repassados pelo SGBD; e
6. formata o resultado na forma de outro documento HTML, com os comandos interpretados que se encontram no metaformulário.

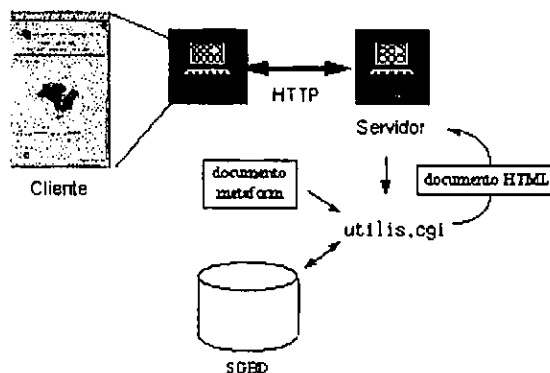


FIG. 4 . Passos realizados no tratamento de um formulário HTML para consulta a uma base de dados usando programa CGI Utilis.

⁴Tecnologia desenvolvida pela Embrapa Informática Agropecuária, anteriormente denominada de *mtform.cgi*.

O desenvolvimento das bases foi realizado com o auxílio do software ERWin⁶, próprio para criação de modelos. O software *Website 1.1e*⁷ que simula um servidor HTTP (Hypert Text Transfer Protocol) para a plataforma PC e os drivers ODBC (*Open DataBase Connectivity*) são responsáveis pela padronização da comunicação entre sistemas diversos e o SGBD utilizado Microsoft Access 97.

6. Resultados e Trabalhos Futuros

O sistema/site CRIA é uma tecnologia de informação baseada no paradigma Cliente (Browser) e Servidor (Servidor HTTP, Servidores de Banco de Dados, Programas CGI), que utiliza a Internet como meio de comunicação.

O Site vem sendo continuamente atualizado e contextualizado, tanto em relação ao seu conteúdo, quanto à organização da sua estrutura.

Como ações futuras prevê-se a conversão das bases implementadas para um sistema gerenciador de banco de dados mais robusto, possivelmente a tecnologia do SGBD OpenIngres, visando assegurar melhor controle de acesso e segurança.

Pretende-se que o site seja um ambiente de interatividade e participação de grupos de interesses distintos de usuários (pesquisadores, especialistas, estudantes, produtores, extensionistas e outros interessados na área). Em função disto, planeja-se disponibilizar ferramentas que permitam o trabalho colaborativo.

Está em andamento a adaptação e validação da metodologia Dublin Core (Dublin ..., 1998), adotada para descrição dos recursos de informação disponíveis no site do CRIA.

A organização apresentada pelo site preocupa-se com a agilidade no acesso aos recursos de informação pelos usuários do segmento agropecuário.

7. Agradecimentos

As autoras agradecem a contribuição do Consultor Hermes Pacheco e de Moacir Pedroso Júnior - idealizadores do modelo conceitual deste projeto -, aos pesquisadores e técnicos da Embrapa Informática Agropecuária em especial a Alvaro Seixas Neto e Tercia Zavaglia Torres pela constante colaboração, bem como aos estagiários pela atuação no desenvolvimento deste sistema: Stefan Arthur Pastorek, Geane Cristina de Melo e Marcelo Stefanelli Santos

8. Referências Bibliográficas

DUBLIN Core Metadata Initiative, 1998. URL: <http://purl.oclc.org/dc> Consultado em 16 de jun. 1998.

LEA-COX, J. D.; MICHAELS, D. T. Development of an electronic information and communication resource for the nursery industry. In: ZAZUETA, F.S.; XIN, J., ed. *Computers in agriculture 1998: proceedings of the 7th international conference*. St. Joseph: ASAE, 1998. p.339-344

MOURA, M.F.; OLIVEIRA, C.A.S. *Aplicativo de consulta aos dados meteorológicos do CNPH via internet: estimativas de médias e gráficos obtidos "online"*. Campinas: EMBRAPA-CNPTIA, 1998. 60p. (EMBRAPA-CNPTIA. Relatório Técnico, 5).

SIEGEL, D. *Criando sites arrasadores na web: a arte da terceira geração em design de sites*. São Paulo: Hayden Books/Quark Editora, 1996. 269p.

LEWIS, T. The new economics of information. *IEEE Internet Computing*, v.2, n.5, p.93-94, Sept-Oct. 1998.

⁶ Erwin[®], produzido pela *Logic Works*[®], subsidiária da Platinum Inc.

⁷ WebSite, produzido pela O'Reilly.

IMPRESSO



*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério de Agricultura e do Abastecimento
Rua Dr. André Tosello, s/nº Caixa Postal 6041 - Barão Geraldo
13083-970 - Campinas, SP
Fone (019) 289-9800 Fax (019) 289-9594
E-mail: cnptia@cnptia.embrapa.br
<http://www.cnptia.embrapa.br>*