

AS BASES DE DADOS OPERADAS PELA EMBRAPA*

WILSON MIASHIRO

EMBRAPA – Departamento de Difusão de Tecnologia
 Edifício Venâncio 2000 – 4º andar
 70000 Brasília, DF

O Sistema de Informação Técnico-Científico (SITCE) da EMBRAPA opera, a nível nacional, as bases de dados de publicações, de bibliografias, de projetos de pesquisa e de tecnologias e recomendações. Discorre sobre as principais bases do exterior, nos Estados Unidos, Inglaterra e França, operadas pela EMBRAPA através do Serviço de Disseminação Seletiva de Informação. Descreve a estruturação dos perfis, o modo de sua permanente atualização, o controle de produção nas diversas fases na recuperação da informação e aponta a possibilidade de transferência de *software* utilizado no sistema.

Vamos apresentar, rapidamente, aqui, as bases que estão em operação na EMBRAPA pelo SITCE (Sistema de Informação Técnico-Científico da EMBRAPA). O SITCE é uma infra-estrutura de informação que foi adotada por uma instituição de pesquisa agropecuária, uma instituição de administração descentralizada.

A EMBRAPA tem, atualmente, em sistema cooperativo, 3.500 pesquisadores, com cerca de 60 centros de unidades e 80 bibliotecas. É a rede que compõe o sistema da EMBRAPA.

Dentro desta apresentação vamos separar as bases de dados geradas pela EMBRAPA, as bases de dados nacionais e as bases de dados do exterior, que nós estamos operando.

Dentre as bases de dados nacionais poderemos destacar quatro. Essas bases foram geradas por sistemas desenvolvidos aqui mesmo. Nenhum desses sistemas é pacote do exterior.

Utilizamos, para cadastramento, os sistemas-padrão, desenvolvidos de forma convencional, e a parte de acesso *on-line* é feita através de um pacote da IBM, o STAIRS; só a parte de recuperação é *on-line*.

* Texto transcrito de fita, não revisado pelo autor.

As bases de dados operadas pela EMBRAPA

A primeira base, a Base de Publicações EMBRAPA/Empresas Estaduais, tem todos os documentos gerados pelo sistema cooperativo de pesquisa agropecuário, e dentro dessa base nós temos, hoje, material produzido de 80 até 82, para acesso *on-line*. Todo o material de 83 também já está cadastrado. Devemos incluir nessa base material produzido de 74 até 79, que são cerca de 4.700 referências. Nós temos aí, então, um total de onze mil referências, cobrindo o período de 74 até 83.

A segunda base de dados, a Base de Bibliografias, é composta por todas as bibliografias que foram levantadas e geradas pela rede. São cerca de 5.000 bibliografias que, hoje, também estão disponíveis por esse sistema.

Uma terceira base de dados, a Base de Dados de Projetos de Pesquisa, envolve cerca de quatro mil projetos de pesquisa em andamento, na EMBRAPA.

A estrutura de pesquisa da EMBRAPA é formada por programas nacionais de pesquisa. Esses programas são definidos de acordo com prioridades de Governo, e debaixo de cada programa nacional de pesquisa nós temos um grupo de projetos.

Para os 40 programas nacionais nós temos cerca de 4.000 projetos de pesquisa em andamento, e nessa base nós temos todas as informações acerca desses projetos de pesquisa, com um resumo — pesquisadores envolvidos, recursos aplicados nos projetos, abrangência geográfica e prazo previsto de início e término dos projetos.

Existem também outras informações que estão agregadas ao controle dos projetos de pesquisa, que são os relatórios de andamento, mas isso é uma outra base de dados, que é de acesso reservado aos chefes de unidades da empresa, e não um acesso aberto ao público.

A quarta base de dados, a Base de Tecnologias e Recomendações — isso são resultados de pesquisas já disponíveis — tem cerca de mil tecnologias que foram geradas pelos projetos de pesquisa, e essas tecnologias estão disponíveis nessa base de dados. Nós temos aí a descrição da tecnologia, a aplicação dela, a que público se aplica, os benefícios potenciais de cada uma dessas tecnologias geradas.

Essas são as principais bases, a nível nacional.

As bases de dados do exterior que estão sendo operadas pela EMBRAPA, através do serviço de Dissiminação Seletiva da Informação, são a Agrícola, da Biblioteca Nacional de Agricultura dos Estados Unidos, e CAB, da Inglaterra. São duas bases gerais da área agrícola. A terceira, Food Science Technology Abstracts (FSTA), é uma base específica com cerca de 1.500 referências por mês, enquanto a Agrícola e a CAB têm cerca de 12 mil referências por mês. BIOSIS e Chemical Abstracts, têm cerca de 30 mil referências por mês; e do SCISERARCH, que é produzida pelo ISI, nós recebemos cerca de 40 mil referências por mês. Esta é uma base multidisciplinar, que traz informações bastante importantes, porque traz exatamente as informações que estão nos limites de aplicação da área agrícola para outras áreas. Geralmente não se encontram essas referências em bases específicas da área agrícola. A Pascal, uma base da França que está em caráter experimental, ainda está em processo de conversão. É a única base que não está em operação normal. A Tele-

gen, a Invagan e a Energy, as três produzidas pelas Intelligence, dos Estados Unidos, são bases bastante específicas, pequenas, com cerca de mil referências por mês.

De todas essas bases, o Serviço de Disseminação Seletiva da Informação dispõe de todo o acervo retrospectivo, que cobre mais ou menos o período de 70, 72 até 84.

Há um formulário de cadastramento para o SDI, onde os usuários, pesquisadores de sistemas, informam a parte de endereços, tipo de busca, título do perfil e o assunto descritivo em que se deseja fazer a pesquisa; no verso desse formulário há a indicação de pesos e palavras-chave ou códigos. O preenchimento pode ser feito utilizando-se os manuais de cada uma das bases de dados.

A estrutura de um perfil que é utilizado pelo sistema é simples. Esse sistema permite que sejam feitas expressões booleanas, para busca; permite ainda a indicação dos campos de busca. Pode-se fazer a pesquisa no texto total ou escolher os campos de autor, título, língua; pode-se fazer indicação de peso para cada palavra-chave, orientar a busca. O peso combinado com a booleana permite uma melhor recuperação, às vezes.

Os perfis são estruturados em grupos. No caso, temos o grupo 1 com um único termo, o grupo 2 com uma série de termos, e no grupo 3 a restrição por línguas. Isso permite truncagem à direita e à esquerda.

O sistema IBT é dividido em cinco módulos. Este também tem cinco módulos. Não é coincidência, pois todos os sistemas do SDI têm esses módulos, basicamente. Há o módulo de atualização de perfis, que é o primeiro. A parte de atualização pode ser feita diretamente, via terminal. A entrada de um perfil gera então dois arquivos: um arquivo de endereçamento e um arquivo de palavras-chave. Esse arquivo de palavras-chave é o que vai ser utilizado pelo módulo de busca.

Existe um segundo módulo, o módulo de conversão. Todas as bases de dados operadas pelo sistema passam por esse módulo. Ou seja, o sistema é capaz de operar, no caso, 10 bases de dados, 50, 60; depende do número de programas de conversão.

Depois dessa conversão, a fita convertida é ativada juntamente com um grupo de palavras-chave e o número de perfis, e tem-se como resultado as referências recuperadas, que entram num quarto módulo, que é um módulo de emissão das listagens, e a gravação num quinto módulo, de um histórico dessas buscas.

Este é o controle administrativo do SDI.

O sistema ainda permite o controle da produção do sistema, através de quinze relatórios: um relatório dando a imagem de todo o arquivo de perfis, uma crítica dos perfis, notificação de cadastramento (cada usuário, após o cadastramento, recebe a informação de que já está cadastrado no sistema e quais foram os códigos e as palavras-chave utilizados nesse perfil), perfis eliminados, registros convertidos (esses relatórios saem na conversão das bases de dados, nos informam para cada base de dados quantos registros se tem mensalmente), um resumo de cada busca,

tificação do número de referências que ele recuperou e se ele ultrapassou, em alguns casos, a 200, 300 ou 400 referências na recuperação; a emissão propriamente dita das referências, no caso do SDI, sai em fichas, e as referências para buscas retrospectivas saem em formulário-padrão, não em fichas), uma relação total de perfis cadastrados a crítica na retroalimentação (o sistema tem uma retroalimentação); depois de cada pacote desses há um formulário de retroalimentação do sistema. Esse relatório permite, então, o acompanhamento da retroalimentação, da atualização da retroalimentação. Essa estatística de atualização, retroalimentação, indica, para todas as unidades, quais foram os perfis que não responderam ao formulário de retroalimentação. Há ainda um alerta por unidade (isto é, os perfis de uma determinada unidade que não responderam à retroalimentação), uma relação completa das buscas realizadas (podemos ter um histórico para cada perfil, de todas as buscas realizadas no período de 1, 2 ou 3 anos), e um relatório de buscas realizadas por bases de dados.

Esses são os relatórios emitidos pelo sistema de SDI, para auxiliar na gerência do sistema.

O SDI tem hoje 2.460 perfis cadastrados nessas bases de dados, cobrindo 116 instituições (na grande maioria instituições da própria EMBRAPA). Temos ainda várias universidades, e cerca de 250 perfis no Cone Sul (Argentina, Paraguai, Uruguai e Bolívia pertencem ao convênio Caribe Cone Sul). Além disso há algumas instituições privadas que, de vez em quando, solicitam recuperação nessas bases.

No primeiro dia, na abertura deste Encontro, o Prof. Murilo Cunha nos falou de algumas preocupações em relação a acesso a bases de dados, e uma das preocupações que ele levantou é quanto ao pagamento ou não desses serviços. Nenhum desses usuários está pagando, hoje, esse serviço. Nós temos várias universidades cadastradas e esse serviço não está sendo cobrado; isso nós podemos garantir para todas essas bases de dados. Todas as instituições oficiais, digo, governamentais, podem ter acesso a essas bases de dados. Só vamos estudar com cuidado os casos das instituições privadas, mas as outras instituições têm todas essas informações disponíveis, a qualquer momento, na EMBRAPA.

Estamos operando o sistema há mais ou menos sete anos. Começamos com uma única base de dados, a Agrícola, o antigo Cain, e fomos progredindo até atingir esse número de bases de dados.

Esse sistema já foi transferido para alguns outros países, e esse *software* está disponível para qualquer instituição do Brasil que se interessar por ele. Podemos transferir os programas, tudo o que tivermos lá, para instituições a nível de governo do Brasil, o oferecemos também todas as buscas, abertas a qualquer instituição.

O serviço do SDI vem desenvolvendo um reforço, há pouco tempo, em termos de pessoal, atuando nesse serviço e hoje podemos garantir que o máximo, em tempo de resposta, é de cerca de oito dias, a partir do recebimento de uma solicitação

de busca na EMBRAPA até a emissão, por correio, do resultado da pesquisa. Com tranqüilidade podemos garantir que em uns oito dias teremos feito o serviço todo.

Abstract

Databases operated by EMBRAPA.

SITCE (EMBRAPA's System of Technical and Scientific Information) operates at a national level several data bases related to publications, bibliographies, research projects, new technologies and results. Comments the major foreign data bases (United States, England and France) operated by EMBRAPA through its SDI service. Describes the structuration of profiles and the process of its permanent actualization, its system of production control in the different phases of information retrieval and suggests the possibility of transferring its software.