

APLICAÇÃO DA BIBLIOMETRIA NA CONSTRUÇÃO DE INDICADORES SOBRE A PRODUÇÃO CIENTÍFICA DA EMBRAPA

ROBERTO DE C. PENTEADO FILHO¹, LEANDRO I. L. DE FARIA², JOB L. G. VIEIRA³, MARIA H. KURIHARA⁴, ANTONIO F. D. AVILA⁵, LUC QUONIAM⁶

1 Assessoria de Comunicação Social, Empresa Brasileira de Pesquisa Agropecuária - Embrapa
Pq.E.B. - Av. W 3 Norte - Final, 90770-901, Brasília, DF, Brasil
Email: roberto.penteado@embrapa.br

2 Núcleo de Informação Tecnológica em Materiais – NIT/UFSCar;
Departamento de Ciência da Informação, Universidade Federal de São Carlos - UFSCAR
Rod. Washington Luís, km 235, 13565-905, São Carlos, SP, Brasil
Email: leandro@nit.ufscar.br

3 UPIS – Faculdades Integradas SEPS 712/912 - Cj A, Brasília, DF, Brasil;
Secretaria de Administração Estratégica, Empresa Brasileira de Pesquisa Agropecuária – Embrapa
Pq.E.B. - Av. W 3 Norte - Final, 90770-901, Brasília, DF, Brasil
Email: job.vieira@embrapa.br

4 Departamento de Tecnologia da Informação, Empresa Brasileira de Pesquisa Agropecuária - Embrapa
Pq.E.B. - Av. W 3 Norte - Final, 90770-901, Brasília, DF, Brasil
Email: kurihara@sede.embrapa.br

5 Secretaria de Administração Estratégica, Empresa Brasileira de Pesquisa Agropecuária - Embrapa
Pq.E.B. - Av. W 3 Norte - Final, 90770-901, Brasília, DF, Brasil
Email: flavio.avila@embrapa.br

6 CRRM/LePont, Universidade de Toulon et du Var - BP 132 83957 - La Garde Cedex, França, e
Centro Franco-Brasileiro de Documentação Técnica e Científica - Cendotec
Av. Prof. Dr. Lineu Prestes, 2242 - Cidade Universitária, 05508-000, São Paulo, SP, Brasil
Email: quoniam@cendotec.org.br

RESUMO

A base de dados - Acervo Documental da Embrapa - contém cerca de 340.000 registros de material bibliográfico, disponíveis para consulta na Internet. Esta iniciativa visa construir indicadores sobre a produção científica dos pesquisadores da Empresa a partir dos diversos campos da base. O trabalho envolve a definição dos indicadores, a análise e definição dos campos do Acervo Documental que são repositório das informações necessárias, definição de estratégias de tratamento e normalização e o uso dos softwares de análise bibliométrica para gerar os indicadores. Pretende-se demonstrar que com o uso de tecnologias de *Textmining* para análise bibliométrica, pode-se produzir série de indicadores da produção técnico-científica de uma organização, a partir de uma base de acervo documental. Também será utilizada a tecnologia de *Datawarehousing* visando facilitar a disponibilização, a utilização de técnicas analíticas avançadas e a integração corporativa destes dados em um ambiente *Oracle*.

Palavras Chave: bibliometria, text mining, data warehouse, indicadores de C&T, inteligência competitiva.

Classificação: Aplicações empresariais, governamentais, na ciência e tecnologia e outras

ABSTRACT

BUILDING SCIENTIFIC PRODUCTION INDICATORS AT EMBRAPA USING BIBLIOMETRY

The database Acervo Documental da Embrapa, also accessible online, provides some 340,000 bibliographic references. This work is an effort to build indicators on the scientific production of Embrapa's research scientists from the different data fields of this database. The project includes the definition of these indicators, an analysis and the definition of the fields in the Acervo Documental da Embrapa, which are the repositories for the necessary information, the definition of normalization and treatment strategies, and the use of text mining technologies for bibliometrical analysis in the production of these indicators. The paper aims to demonstrate how the application of these softwares, such as *Infotrans*, *Dataview* and *Matrisme*, can generate a series of indicators in an organization's scientific-technical production. A perspective of *Datawarehouse* in the integration of these data for the Oracle environment is also considered.

Keywords: bibliometry, SC&T indicators, text mining, data warehouse, competitive intelligence.

INTRODUÇÃO

A Empresa Brasileira de Pesquisa Agropecuária - Embrapa, vinculada ao Ministério da Agricultura, Pecuária e Abastecimento, foi criada em 26 de abril de 1973. Sua missão é viabilizar soluções para o desenvolvimento sustentável do agronegócio brasileiro por meio de geração, adaptação e transferência de conhecimentos e tecnologias, em benefício da sociedade. A Embrapa atua por intermédio de 37 Centros de Pesquisa, três Serviços e, na sua Sede, 15 Unidades Centrais, estando presente em quase todos os Estados da Federação, nas mais diferentes condições ecológicas.

Para chegar a ser uma das maiores instituições de pesquisa do mundo tropical, a Empresa investiu sobretudo no treinamento de recursos humanos, possuindo, hoje, 8.530 empregados, dos quais 2.045 são pesquisadores, 47% com mestrado e 49% com doutorado, operando um orçamento da ordem de R\$ 660 milhões anuais. Está sob a sua coordenação o Sistema Nacional de Pesquisa Agropecuária-SNPA, constituído por instituições públicas federais, estaduais, universidades, empresas privadas e fundações, que, de forma cooperada, executam pesquisas nas diferentes áreas geográficas e campos do conhecimento científico.

A Embrapa adota, desde 1994, o denominado Sistema Embrapa de Planejamento (SEP). O SEP está evoluindo para o Sistema Embrapa de Gestão (SEG), por meio da adoção do sistema competitivo de projetos. O novo SEG se estrutura em cinco Macroprogramas: 1) Grandes desafios nacionais; 2) Competitividade e sustentabilidade setorial; 3) Desenvolvimento tecnológico incremental; 4) Comunicação e transferência e 5) Processos gerenciais.

O AGRONEGÓCIO DO FUTURO

O Agronegócio brasileiro tem sido, na última década, o principal sustentáculo da boa saúde da economia nacional. Isto graças aos elevados superávits comerciais por ele gerados, atualmente próximos dos US 20 bilhões anuais. A Embrapa aposta num futuro promissor para o agronegócio brasileiro. A competitividade deste último se definirá, cada vez mais, por sua capacidade de incorporar, de forma contínua e sustentada, inovações tecnológicas que permitam atender às crescentes demandas do mercado interno e desafiar os subsídios dos competidores e a tendência histórica de preços decrescentes no mercado internacional de produtos agrícolas.

Assim, a Embrapa está implementando um processo sistemático de análise e ajuste de sua Agenda Prioritária de P&D de forma a, cada vez mais, concentrar esforços no desenvolvimento de inovações que potencializem a competitividade e a sustentabilidade do setor agropecuário e a viabilização de negócios de pequeno porte, que contribuam para o aumento da renda, do emprego e da inserção social, a redução de riscos e o aumento de precisão, a aceleração dos processos de informação e decisão e a agregação de qualidade e valor aos produtos do agronegócio brasileiro.

OBJETIVOS ESTRATÉGICOS DA EMBRAPA

São os seguintes os Objetivos de Gestão da Empresa, segundo o III Plano Diretor da Embrapa¹, necessários para que cumpra bem sua missão:

- Viabilizar soluções tecnológicas para o desenvolvimento de um agronegócio competitivo em uma economia global;
- Viabilizar soluções tecnológicas para o agronegócio, que promovam a sustentabilidade das atividades econômicas com o equilíbrio ambiental;
- Viabilizar soluções tecnológicas que contribuam para diminuir os desequilíbrios sociais;
- Viabilizar soluções tecnológicas para fornecimento de matérias-primas e alimentos que promovam a saúde e a melhoria do nível nutricional e da qualidade de vida da população.

INFORMAÇÕES GERENCIAIS

A Embrapa dispõe de bases de dados gerenciais, entre outras, o Sistema Plano Anual de Trabalho (Sispat), o Sistema de Avaliação das Unidades (SAU) e o Sistema de Avaliação e Premiação por Resultados (Sapre) que administram boa parte dos processos, planos, atividades e resultados.

As informações do Sispat são utilizadas para o processo de negociação, acompanhamento e avaliação das metas do Sapre, assim como para a elaboração de relatórios gerenciais para o Ministério da Agricultura e do Abastecimento e Tribunal de Contas da União, entre outros, além de servir de base para o processo de monitoramento institucional e planejamento estratégico da Empresa. Esta, também utiliza os indicadores de desempenho das Unidades de Trabalho para avaliar e premiar empregados e equipes de trabalho segundo as metas de produtividade da empresa como um todo.

Entre os indicadores apurados pelo Sispat/SAU/Sapre estão:

1. Produção Técnico-Científica: 1.1-Artigo em Periódico Indexado; 1.2-Capítulo em Livro Técnico-Científico; 1.3-Artigo em Anais de Congresso / Nota Técnica; 1.4-Resumo em Anais de Congresso; 1.5-Orientação de Tese de Pós-Graduação

2. Produção de Publicações Técnicas: 2.1-Circular Técnica; 2.2-Comunicado Técnico; 2.3-Boletim de Pesquisa e Desenvolvimento; 2.4-Documentos; 2.5-Organização/Edição de Livros ou Sistema de Produção; 2.6-Artigo de Divulgação na Mídia

3. Desenvolvimento de Tecnologias, Produtos e Processos: 3.1-Cultivar Gerada / Lançada e Evento Elite; 3.2-Cultivar Testada / Recomendada; 3.3-Prática / Processo Agropecuário; 3.4-Raça-Tipo; 3.5-Insumo Agropecuário; 3.6-Processo Agroindustrial; 3.7-Metodologia Científica; 3.8-Máquina, Equipamento ou Instalação; 3.9-Estirpes; 3.10-Monitoramento / Zoneamento; 3.11-Software; 3.12-Organização da Informação para Base de Dados

INFORMAÇÕES BIBLIOGRÁFICAS

Conhecer em detalhe sua produção científica e tecnológica é fundamental para uma organização de pesquisa, tanto para subsidiar seu direcionamento estratégico como para avaliar se as ações que estão sendo tomadas quanto à pesquisa científica estão de acordo com o planejado em seus programas de pesquisa e desenvolvimento.

A Embrapa conta com um acervo descentralizado, disponível nas bibliotecas de suas várias unidades. Incorporada à Base de Dados da Pesquisa Agropecuária, a Empresa produz a base Acervo Documental (disponível em CD-ROM ou através da Web no site <http://www.cnptia.embrapa.br/servicos/bdpa/frame_bdpa.html>) que reúne cerca de 340.000 referências bibliográficas e contempla praticamente a totalidade dos trabalhos científicos e tecnológicos de seus pesquisadores, além de todas as publicações presentes nas bibliotecas da Empresa. O Acervo Documental pode ser utilizado para a recuperação de dados sobre as publicações da Embrapa, com vistas a subsidiar o processo de análise e construção de indicadores. No entanto, o volume de informação disponível e a finalidade da base, criada para recuperar publicações, tornam-se desafios para o processo de análise de sua produção técnico-científica.

Segundo Faria², a disponibilização de bases de dados eletrônicas aumentou a oferta de informação sobre praticamente qualquer assunto (Dou³, Porter⁴, Wormell⁵). Mas, quando há muita informação, como é o caso do Acervo Documental da Embrapa, o processo de análise e síntese da informação fica mais complexo, exigindo o uso de metodologias adequadas. O fato

de as bases de dados oferecem informação abundante, estruturada e eletrônica, motivou o surgimento de metodologias de análise da informação auxiliada por computadores, visando à elaboração de informação de alto valor agregado para o tomador de decisão (Quoniam⁶).

Este trabalho visa à realização de uma análise da produção científica da Embrapa para verificar o volume de produção científica e tecnológica anual, os temas de pesquisa abordados pela Empresa e suas unidades e a evolução desses temas ao longo de sua história. Também se buscará a construção de indicadores sobre a produção científica dos pesquisadores da Embrapa e de seus Centros de Pesquisa.

METODOLOGIA

Serão utilizados os softwares *Infotrans* (da IuK GmbH - Alemanha), para reformatação dos dados, *Dataview* e *Matrisme* (do CRRM/LePont - Universidade de Aix-Marseille III e Universidade de Toulon et du Var - França), para tratamento bibliométrico dos dados, MS Excel para elaboração de gráficos e tabelas e Oracle-Discover para armazenamento e disponibilização de dados aos tomadores de decisão da Embrapa. O Acervo Documental da Embrapa é a base utilizada.

Preliminarmente foram apurados três indicadores gerais: Produção Científica da Embrapa por Ano, Produção Científica da Embrapa por Pesquisador e Produção Científica da Embrapa por Centro de Pesquisa. Os campos do Acervo trabalhados foram: autoria, origem material, afiliação, ano de publicação e título.

Por meio de uma consultoria realizada pela bibliotecária Vera Lúcia de Campos Octaviano, foram identificadas as principais inconsistências dos registros, nos campos Autoria, Título. Por exemplo: entradas como "João José Silva", "Silva, João José", "Silva, J. J.", "Silva, JJ" e outras variantes são padronizados para "Silva, J.J.". Os campos Origem Material e Afiliação com Unidades de Pesquisa grafadas em diferentes formas são padronizados. Os campos selecionados são processados no software *Infotrans*, que permite a criação de regras genéricas de padronização para todos os nomes, títulos e palavras a serem reformatados. Outro recurso importante do software é o da eliminação automática de todas as duplicações de registros.

Finalizada esta etapa, estando a base pronta para a análise bibliométrica, os campos extraídos (autoria, origem material, afiliação, ano de publicação e título) são tratados no *Dataview*, com a geração de matrizes, Ano de publicação-Autor, Ano de publicação-Afiliação, Autor-Título, Afiliação-Título, Autor-Afiliação. Estas matrizes foram exportadas

para o *Matrisme*, visando a criação de redes relacionais de autores e temas, e para o MS Excel, para a elaboração de gráficos e tabelas.

A partir destas tabelas, cria-se o modelo relacional e os dados são carregados para o banco de dados Oracle. Em seguida, cria-se o modelo dimensional de dados no ambiente de data warehouse Oracle-Discover, iniciando-se assim o processo de extração, transformação e carga (ETL), importando estes dados para o modelo de dados dimensional. Os dados foram disponibilizados para acesso pelos pesquisadores através da criação de consultas básicas, porém dada a facilidade de construção de novas consultas que este ambiente proporciona ao usuário final, abre-se variadas perspectivas de consultas *ad hoc* assim como a aplicação de análises estatísticas e geração de planilhas e gráficos multidimensionais de maior amplitude.

DISCUSSÃO

A importância e valor dos indicadores se pauta na possibilidade de quantificar "coisas" intangíveis como a geração de conhecimentos. Alguns exemplos de indicadores de científicos são: número de publicações por ano da empresa, número de pesquisadores ativos, evolução dos principais temas de pesquisa ano a ano, recursos financeiros aplicados e bolsas concedidas.

O Ministério da Ciência e Tecnologia⁷ (MCT) publica na Internet, no site: <http://www.mct.gov.br/estat/ascavpp/Default.htm>, indicadores de C&T referentes a: recursos aplicados, recursos humanos, bolsas de formação e pesquisa, fomento, produção científica e patentes. Na página de produção científica, por exemplo, são apurados sete indicadores: Número de artigos publicados em periódicos científicos internacionais indexados, diversas comparações 1981-2000; Número de artigos publicados em periódicos científicos internacionais indexados, diversas comparações, por áreas do conhecimento, 1981-2000; Número de artigos científicos e técnicos publicados - principais países - 1981-2000; Vinte países com maior crescimento no número de artigos publicados em periódicos científicos internacionais indexados, 1995-2000; Produção bibliográfica no triênio 1997-1999; Produção técnica no triênio 1997-1999 e Orientações concluídas no triênio 1997-1999.

Rostaing⁸ ampliou a definição de Prichard⁹ para "bibliometria" como a aplicação de métodos matemáticos e estatísticos a publicações para incluir também conjuntos de elementos extraídos dessas publicações, como, por exemplo, referências bibliográficas.

Para Van Raan¹⁰ (apud. Faria¹, p.34), "a elaboração de indicadores de Ciência e Tecnologia é uma das principais aplicações da bibliometria". Mugnaini e Quoniam¹¹

sublinham que a bibliometria pode ser vista como um instrumento para agregação de valor à gama de indicadores quantitativos que os governos vêm gerando para análise da produção científica.

Faria¹ (p. 9) acrescenta que "os indicadores são uma forma de sintetizar e agregar valor à informação". Ele sintetiza o processo de tratamento automatizado da informação por meio da figura a seguir:

FIGURA 1: Processo de tratamento automatizado da informação



Faria cita ainda Quoniam¹² e Moge¹³ sublinhando que, raramente os dados recuperados das bases podem ser diretamente utilizados pelos softwares de análise bibliométrica. Isto acontece porque tais bases foram desenvolvidas para identificação e localização de documentos e não para análise. Assim, sempre é preciso melhorar os dados e até inserir dados externos ou criar novos dados para estudá-los em profundidade.

O Acervo Documental da Embrapa, apesar de reunir praticamente a totalidade dos trabalhos dos pesquisadores da Empresa, não foi desenvolvido para permitir a análise dos dados e sim para agilizar a identificação e localização de documentos. Seu uso para análise automatizada exige, portanto, uma longa etapa de preparação e reformatação dos dados, sem o qual não teríamos qualidade das informações.

Van Raan¹⁰ afirma que tal preparação precisa também ajustar-se às condições em que o tratamento é realizado. Para estudos focalizados e de duração relativamente curta, como alguns dos trabalhos de inteligência competitiva, a preparação dos dados normalmente é desenvolvida especificamente para esta tarefa. Para estudos sistemáticos e de duração mais longa, como a produção de indicadores para a avaliação da produção científica, como é o caso neste trabalho, a preparação dos dados torna-se um processo contínuo e sistemático, que visa

à criação de um sistema de informação perene, aprimorado ao longo do tempo com a experiência resultante de sucessivas aplicações (apud. Faria¹, p. 22).

Leeds¹⁴ cita quatro etapas no tratamento automatizado de dados: **identificação das necessidades de informação; preparação dos dados; tratamento bibliométrico e análise dos resultados**. Para ele, a **preparação dos dados** consome cerca de 60% de todo o esforço empreendido no tratamento automatizado de dados. Antes da preparação, cerca de 20% do tempo total seria dedicado à **identificação das necessidades de informação**. Após a preparação, 10% do tempo vai para o **tratamento bibliométrico** e outros 10% para a **análise dos resultados e assimilação do conhecimento**.

A próxima etapa neste ciclo, segundo a metodologia desenvolvida pelo laboratório CRRM/LePont, da Universidade de Aix-Marseille III e da Universidade de Toulon et du Var, na França, é o tratamento destes dados em tecnologias de análise bibliométrica, o *Dataview* e o *Matrisme*. Os softwares têm como princípio básico a construção de grupos por propagação e a representação das relações entre os elementos constituintes de um grupo por meio de uma rede (Boutin¹⁵, Quoniam¹⁶).

O processo contínuo de transformação de dados brutos em *conhecimento estratégico*, tem crescido exponencialmente nos dias atuais. A própria necessidade das empresas em obter conhecimento da sua cadeia de produção tem levado a um uso intensivo de tecnologias da informação disponibilizadas pela indústria de "software" em todo o mundo.

No entanto, como já foi ressaltado neste trabalho, na maioria dos casos, os dados armazenados em bases de dados nas organizações não estão prontos para o processo de análise informacional, portanto, dificultam o processo de transformação em conhecimento. Com o auxílio de ferramentas de *Text Mining* e *Data Warehouse*, os dados e informações armazenadas em banco de dados podem ser consultados e facilmente entendidos pelo usuário.

O propósito de uso da tecnologia de *Data Warehouse* é proporcionar a organização destes dados, reformata-los e armazená-los de forma eficiente e ainda, facilitar a recuperação da informação para a tomada de decisão através de uma interface amigável e de fácil acesso. Os softwares *Infotrans* e *Dataview*, que fazem o trabalho de limpeza dos dados e geração de dados quantitativos, antecedem o trabalho de *Datawarehousing* extraindo dados numéricos de dados textuais. Por outro lado, o uso da tecnologia de *Datawarehousing* evoca a necessidade de se ter um repositório central e integrado de informações que poderão, mais tarde, comportar e integrar novas estruturas de dados.

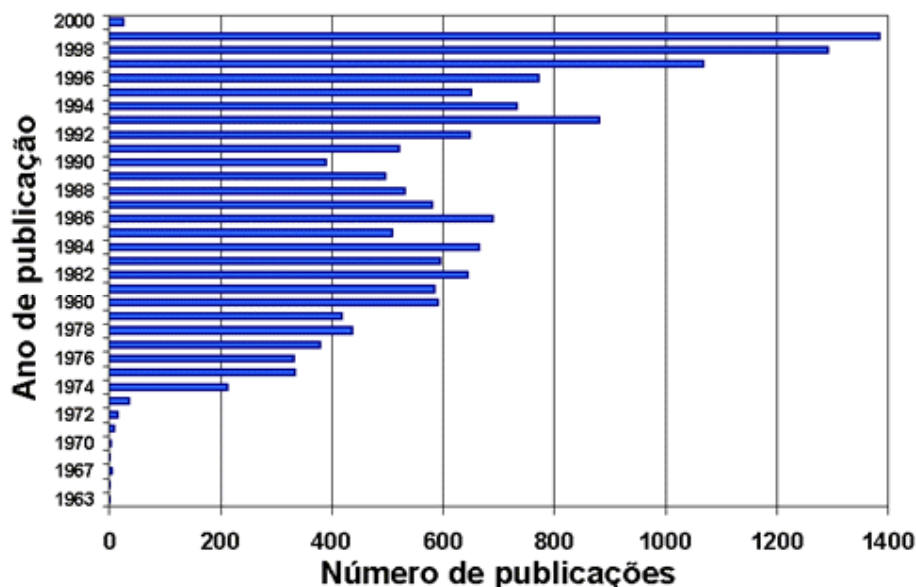
De acordo com Inmon e Hackathorn¹⁷ a maioria das ferramentas do mercado de software oferecem alguma maneira de filtrar dados para garantia de qualidade, durante os processo de extração e transformação. Entretanto, ferramentas específicas para limpeza e tratamento de dados, como o *Infotrans*, *Dataview* e *Matrisme* oferecem mecanismos bem mais sofisticados e conseqüentemente maior credibilidade da informação. Por exemplo, no caso da base de dados Sispat, os indicadores são gerados a partir de informações levantadas pelos pesquisadores e técnicos das Unidades da Embrapa. Ao levantar estes dados de forma automática, a partir da análise bibliométrica, evita-se a subjetividade do trabalho humano.

RESULTADOS

Como resultado da reformatação e do tratamento bibliométrico dos dados extraídos do Acervo Documental foram automaticamente produzidos, em fase piloto, três indicadores científicos: rankings, tabelas e figuras da produção científica da Embrapa.

O primeiro, um indicador de frequência básico como a evolução do número de publicações, representada na figura 2. Pode-se observar que, a partir de 1997, a produção de publicações por pesquisadores da Embrapa mudou de patamar, ultrapassando a marca de 1100 por ano e cresce significativamente desde então.

FIGURA 2: Produção científica da Embrapa - Ano de publicação



Fonte: Acervo Documental Embrapa, até 2000

O segundo indicador básico produzido é o da produção científica dos pesquisadores da Embrapa (figura 3). Após a reformatação inicial dos dados, considerando-se a produção nos 29 anos de criação da Empresa, um resultado preliminar verificado é que quatro pesquisadores: Santos, H.P.S.; Campos, O.F.; Souza, C.N.A.; Ignaczak, J.C., produziram entre 257 e 150 trabalhos.

FIGURA 3: Produção científica da Embrapa - pesquisadores

Antes da reformatação		Após a reformatação	
1	Pesquisadores	1	Pesquisadores
2	EMERAPA (BRASILIA, DF)	2	SANTOS, H.P.S.
3	EMERAPA CENTRO DE PESQUISA AGROPECUARIA DOS CERRADOS	3	CAMPOS, O.F.
4	EMERAPA	4	SOUZA, C.N.A.
5	EMERAPA SERVIÇO NACIONAL DE FORTALECIMENTO E CONSERVAÇÃO	5	IGNACZAK, J.C.
6	EMERAPA CENTRO NACIONAL DE PESQUISA DE TRIGO (PASSO FULADO)	6	ARIAS, G.
7	EMERAPA CENTRO NACIONAL DE PESQUISA DE SOJA (LONDRINA)	7	MOREIRA, J.C.S.
8	EMERAPA DEPARTAMENTO DE INFORMAÇÃO E DOCUMENTAÇÃO	8	CUNHA, G.R.
9	SANTOS, H.P. DOS	9	ALVIM, M.J.
10	ALVIM, M.	10	BAIER, A.C.
11	ROSA, A. DO N.	11	ROSA, A.N.
12	EUCLIDES FILHO, K.	12	AMEROSI, I.
13	EMERAPA DEPARTAMENTO DE DIFUSÃO DE TECNOLOGIA, BRASILIA, DF	13	MINELLA, E.
14	EMERAPA (BRASILIA, DF)	14	LHAMBY, J.C.B.
15	TEODORO R.L.	15	EUCLIDESFILHO, K.
16	POTT, A.	16	BOTREL, M.A.
17	VALERIO, J.R.	17	BONATO, E.R.
18	EMERAPA DEPARTAMENTO DE DIRETRIZES E METODOS DE PLANEJAMENTO	18	TEODORO, R.L.
19	BIANCHINI, I.	19	VERNEQUE, R.S.
20	VILELA, D.	20	POTT, A.
21	MARTINEZ M.L.	21	PICININI, E.C.
22	ARIAS, G.	22	LEMUS, A.M.

Fonte: Acervo Documental Embrapa, até 2000

Finalmente, o terceiro indicador básico diz respeito à produção científica dos Centros de Pesquisa da Embrapa (figura 4). Apesar de serem resultados preliminares, que poderão ser revistos, as Unidades Descentralizadas mais produtivas nestes 29 anos da Embrapa foram o Centro Nacional de Pesquisa de Trigo, o Centro Nacional de Pesquisa de Gado de Leite e o Centro Nacional de Pesquisa de Gado de Corte, que produziram entre 2.459 e 1.735 trabalhos.

FIGURA 4: Produção científica da Embrapa - Unidades

Antes da reformatação

1	Instituição	Publicações
2	CAMPO VAZIO	6269
3	CNPGL	1827
4	EMBRAPA. CENTRO NACIONAL DE PESQUISA DE GADO DE CORTE (CAM)	1183
5	EMBRAPA	776
6	EMBRAPA. CENTRO DE PESQUISA AGROPECUARIA DO PANTANAL (COI)	288
7	EMBRAPA. CENTRO DE PESQUISA AGROPECUARIA DO PANTANAL (COI)	259
8	EMBRAPA - CNPT, PASSO FUNDO, RS.	247
9	CNPQ	196
10	EMBRAPA. CENTRO NACIONAL DE PESQUISA DE GADO DE CORTE (CAM)	187
11	EMBRAPA SEMIARIDO, CP 23, CEP 56.300.970 PETROLINA, PE.	183
12	EMBRAPA AMAZONIA OCIDENTAL.	156
13	1-2 EMBRAPA - CNPT, PASSO FUNDO, RS.	156
14	CNPGL.	142
15	EMBRAPA TRIGO, PASSO FUNDO, RS.	136
16	EMBRAPA. CENTRO NACIONAL DE PESQUISA DE ARROZ E FELIJO (GO)	133
17	EMBRAPA.DID	116
18	EMBRAPA.DDT	105
19	1-2 EMBRAPA TRIGO, PASSO FUNDO, RS.	103
20	EMBRAPA - UEPAE DE SAO CARLOS, SAO CARLOS, SP.	102
21	EMBRAPA PECUARIA SUDESTE, SAO CARLOS, SP.	100
22	EMBRAPA. CENTRO NACIONAL DE PESQUISA DE GADO DE CORTE (CAM)	94

Após a reformatação

1	Instituição	Publicações
2	EMBRAPA	2589
3	EMBRAPA - CNPT	2459
4	EMBRAPA - CNPGL	2172
5	EMBRAPA - CNPGL	1735
6	EMBRAPA - CPAP	664
7	EMBRAPA - CPAC	634
8	EMBRAPA - CPATSA	568
9	EMBRAPA - CNPSO	432
10	CNPQ	342
11	EMBRAPA - SERVICO NACIONAL	329
12	CONSERVACAO DE SOLOS	311
13	EMBRAPA - CNPA	284
14	EMBRAPA - CPATU	263
15	EMBRAPA - CPAA	243
16	EMBRAPA - DEPARTAMENTO D	223
17	EMBRAPA - CNPAF	218
18	EMBRAPA - CPPSE	210
19	EMBRAPA - CPAFRO	208
20	EMBRAPA - UEPAE DOURADOS	197
21	DOCUMENTACAO	195
22	EMBRAPA - CNPMT	167

Fonte: Acervo Documental Embrapa, até 2000

A criação de indicadores de C&T da pesquisa agropecuária, a partir da base Acervo Documental da Embrapa, fará necessária uma reformulação de métodos e rotinas de trabalho, visando adequá-los a um processo perene de análise automatizada, conforme referido em Van Raan¹⁰.

A partir do tratamento bibliométrico da base Acervo Documental, além dos indicadores básicos como Número de Trabalhos Publicados por Ano, Número de Trabalhos Publicados por Unidade da Embrapa e o de Número de Trabalhos Publicados por Pesquisador, podem ser apurados e substituídos outros indicadores do SAU/Sapre como:

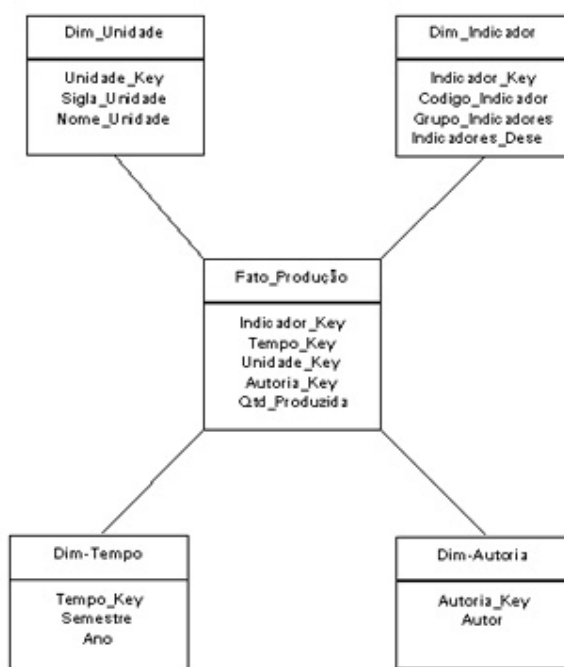
1. Produção Técnico-Científica: 1.1-Artigo em Periódico Indexado; 1.2-Capítulo em Livro Técnico-Científico; 1.3-Artigo em Anais de Congresso / Nota Técnica; 1.4-Resumo em Anais de Congresso.

2. Produção de Publicações Técnicas: 2.1-Circular Técnica; 2.2-Comunicado Técnico; 2.3-Boletim de Pesquisa e Desenvolvimento; 2.4-Documentos; 2.5-Organização/Edição de Livros ou Sistema de Produção

No entanto, para permitir a utilização desses indicadores no Sispat, SAU e Sapre, a atualização do Acervo Documental, realizada duas vezes por ano, nos meses de abril e novembro, quando são acrescentadas e reunidas todas as novas inserções, deverá ser coordenada com o calendário gerencial, para que sejam incluídos, pelas bibliotecas, todos os trabalhos produzidos no período a ser analisado.

A partir do modelo relacional definido para estes indicadores, os dados são carregados para o banco de dados *Oracle*. Em seguida, cria-se o modelo dimensional de dados no ambiente de *data warehouse Oracle-Discover* e inicia-se o processo de *ETL* (extração, transformação e carga) importando estes dados para o modelo de dados dimensional ilustrado na figura 5.

FIGURA 5: Modelo Dimensional da Produção Técnico-Científica da Embrapa



Após o tratamento qualitativo da base de dados, com o *Dataview* e o *Matrisme*, serão levantados os indicadores de produtividade existentes entre e criada uma rede dinâmica de relacionamentos, permitindo a posterior descoberta de conhecimento para apoiar o processo de decisão e gestão Empresa.

CONCLUSÃO

A bibliometria é uma ferramenta importante para o processo de produção de informações gerenciais, em particular, a geração de indicadores a partir de bases de dados bibliográficos.

A metodologia empregada permite trabalhar com dados seguros e mais objetivos, oriundos de uma base bibliográfica, além de evitar a duplicação de esforços com a

alimentação de diferentes bases de dados na Empresa com a mesma informação. Produz, ainda, além de tabelas, resultados gráficos que permitem uma melhor visualização de tendências.

Outro resultado importante obtido com este trabalho é a limpeza da base Acervo Documental de inconsistências comuns como as entradas duplicadas de autores e títulos.

Com o trabalho continuado no Acervo Documental mais indicadores poderão ser apurados e melhores e mais profundas análises da produção científica da Embrapa são a prever no futuro, integrando outras fontes de informação da própria Empresa.

Pode-se observar uma complementaridade, com o uso combinado destas tecnologias. Enquanto uma parte de dados brutos gera dados quantitativos (Textmining), a outra (Datawarehousing) pode oferecer, além de um repositório central para armazenamento dos dados, a facilidade de acesso e a possibilidade de uso de outras técnicas analíticas para os tomadores de decisão.

REFERÊNCIAS

1 EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. **III Plano Diretor da Embrapa: Realinhamento Estratégico 1999-2003**. Secretaria de Administração Estratégica. Brasília. 1998.

2 FARIA, L.I.L. **Prospecção tecnológica em materiais: aumento da eficiência do tratamento bibliométrico - uma aplicação na análise de tratamento de superfícies resistentes ao desgaste**. Université de Droit, d'Economie et des Sciences d'Aix Marseille, 176p, Tese (Doutorado) - Faculté des Sciences et Techniques de Saint Jerome, 2001.

3 DOU, H. Maitriser la prolifération de l'information grâce à la bibliometrie. **Technologies Internationales**, n.9, p.42-44, nov. 1994.

4 PORTER, A.L. **Text Mining for Technology Foresight**, 2000. Disponível em: <<http://tpac.gcatt.gatech.edu/toa/foresight-outline.html>>. Acesso em 10 jul. 2000.

5 WORMELL, I. Informetria: explorando bases de dados como instrumentos de análise. **Ciência da Informação**, Brasília, DF, v.27, n.2, p.210-216, 1998.

6 QUONIAM, L. Bibliometrie sur des references bibliographiques: methodologie. In DESVALS, H.; DOU, H. (Org.). **La veille technologique**. p. 244-252, Dunod, 1992.

7 MINISTÉRIO DA CIÊNCIA E TECNOLOGIA - MCT, **Indicadores de C&T**, 2002. Disponível em: <<http://www.mct.gov.br/estat/ascavpp/Default.htm>>. Acesso em 10 jul. 2002.

8 ROSTAING, H. **Veille technologique et bibliometrie: concepts, outils, applications**. Marseille: Université d'Aix-Marseille III, 370p. Tese (Doutorado). Faculté des Sciences et Techniques de Saint Jerome, 1993.

- 9 PRITCHARD, A. Statistical bibliography or bibliometrics?. **Journal of publication**, v. 25, , p. 348-349, 1969.
- 10 VAN RAAN, A.F.J. Scientometrics: state-of-the-art. **Scientometrics**, v. 38, n. 1, p. 205-218, 1997.
- 11 MUGNAINI, R. e QUONIAM, L. **Bibliometria em países em desenvolvimento: notas para discussão**. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 10., junho 2002, Fortaleza, SBBDCI, 2002.
- 12 QUONIAM, L. et al. Bibliometric analysis of patent documents for R&D management. **Research Evaluation**, v. 3, n. 1, p. 13-18, avr. 1993.
- 13 MOGEE, M.E. **Patents and technology intelligence**, in: ASHTON, W.B.; KLAVANS, R.A. Keeping abreast of science and technology: technical intelligence for business, Battelle Press, p.560, 1997.
- 14 LEEDS, S. Data Mining: Beware of the shaft. **Direct Marketing**. Jan. 2000. Disponível em: <<http://www.tmiassoc.com/articles/shaft.htm>>. Acesso em 10 jul. 2002.
- 15 BOUTIN, E. **A cadeia de tratamento da informação do CRRM/LePont**. In: Seminário "Tecnologias para Tratamento da Informação na Embrapa". Brasília: EMBRAPA, CRRM/LePont, Cendotec, dec. 2001.
- 16 QUONIAM, L. **Datamining, teoria e prática**. In: Seminário "Tecnologias para Tratamento da Informação na Embrapa". Brasília: EMBRAPA, CRRM/LePont, Cendotec, dec. 2001.
- 17 INMON, W.H.; HACKATHORN, R.D. **Using the Data Warehouse**. Wiley - QED Publication, 1994.