

IMPLEMENTAÇÃO DO BANCO DE DADOS SCAP NO NÚCLEO DE COMPUTAÇÃO PUC/RJ

Rosane Teles Lins Castilho *

Descrição da implantação do banco de dados SCAP no Núcleo de Computação PUC/RJ: representação da informação; metodologia do trabalho de armazenamento das informações; educação do usuário; comentários sobre o desempenho do Sistema.

1. INTRODUÇÃO

O objetivo do presente trabalho é descrever a implantação do banco de dados SCAP no Núcleo de Computação da PUC/RJ.

O SCAP (Sistema Conversacional de Consultas para Artigos de Periódicos) é discutido em detalhes em outro documento.

2. REPRESENTAÇÃO DA INFORMAÇÃO

Artigos de periódicos representativos para a comunidade científica e que apresentassem todos os campos de informação necessários a uma pesquisa bibliográfica, principalmente descritores (termos-chaves) fornecidos por seus próprios autores, foram utilizados para dar início ao banco de dados SCAP.

Antes de qualquer tentativa de convencer a comunidade científica a participar do trabalho, no sentido de redigir resumos, classificar os artigos, ou outras atividades, era necessário demonstrar o desempenho, a eficiência, a utilidade do Sistema.

* Setor de Documentação. Núcleo de Computação. Pontifícia Universidade Católica/RJ.

Observada a inexistência de qualquer padronização dos elementos que compõem os campos de informação desses artigos, fez-se necessária a adoção de normas que garantissem a uniformidade da entrada desses elementos no Sistema.

Dentre as normas adotadas destacam-se aquelas referentes a:

- **Autores Individuais:** indexados por seus sobrenomes, tiveram seus pre-nomes abreviados. Exemplo: Forsythe, G.E.
- **Sociedades, Instituições, etc.:** tiveram seus nomes substituídos por siglas. Exemplo: American National Standards Institution – ANSI.
- **Títulos dos Periódicos:** foram também abreviados, segundo a norma internacional. Exemplo: Communications of the ACM – CACM.

Uma relação dos autores, instituições e siglas, títulos abreviados, é fornecida ao usuário no momento de uma consulta ao banco de dados.

- **Termos Chaves:** maior importância foi atribuída ao presente elemento, já que as perguntas dos usuários prendem-se, geralmente, ao que existe sobre determinado assunto.

Os termos-chaves fornecidos pelos autores dos artigos não se baseavam em um vocabulário previamente controlado. Em outras palavras, o autor determinava o assunto de seu artigo utilizando sua própria terminologia. Por exemplo: o termo-chave "automata" assim vinha escrito em dado artigo, em outros "automatas" ou mesmo "automaton"; o mesmo acontecia com "time-sharing" que poderia aparecer escrito dessa forma ou mesmo "time sharing" e ainda "timesharing". Cabia também ao autor decidir a quantidade de termos que iria caracterizar o assunto de seu artigo.

A inconstância do vocabulário resultou na adoção de um dos mais importantes conceitos do sistema de indexação UNITERMO, a fim de dar a esse vocabulário a uniformidade necessária: o conceito de termos "livres". Baseado neste conceito determinou-se que nenhum termo-chave permaneceria composto, seus elementos seriam separados dando origem a dois ou mais termos inteiramente livres. Esses termos permaneceriam livres até que uma frequente combinação de elementos justificasse a formação de um conceito, isto é, um único termo chave composto de dois ou mais elementos. Termos-chaves caracterizados no exemplo "automatas" foram introduzidos no Sistema na forma em que os mesmos ocorriam. Problemas de falta de especificidade decorrente da atribuição de um grande número de termos-chaves a determinados artigos não foram considerados nesta fase. Todos os termos indicados pelo autor foram devidamente aproveitados.

A busca por informações segue, basicamente, a mesma filosofia do sistema de buscas do UNITERMO: requer somente a escolha dos termos significativos ao assunto em que se está interessado. Em seguida, operações lógicas do tipo booleano são efetuadas entre esses termos, e o resultado será a obtenção das referências bibliográficas dos artigos que poderão satisfazer a necessidade de informação.

O usuário obtém uma relação dos termos chaves armazenados no Sistema através de uma listagem denominada "dicionário de termos chaves".

3. METODOLOGIA

O método adotado no preparo e armazenamento das informações no Sistema foi o seguinte:

- Orientação do Trabalho
 - Codificação
 - Perfuração
 - Verificação
 - Controle
 - Atualização do Banco de Dados (BD)
 - "Back-up"
- **Orientação do Trabalho:** compreende o fornecimento de instruções para a transcrição dos dados dos artigos em folhas de codificação; padronização das entradas de dados; controle de produção.
- **Codificação, Perfuração, Verificação:** a fim de não concentrar o trabalho em uma única atividade, as revistas foram separadas em pequenos lotes. Enquanto esperava-se a perfuração de um lote, outro lote era codificado, ou mesmo fazia-se a verificação do lote anterior. Por verificação entende-se a análise e correção dos dados provenientes da perfuração.
- **Controle** – consiste na verificação final dos dados, efetuada pelo elemento responsável pelo projeto.
- **Atualização BD:** consiste no armazenamento dos dados no Sistema, isto é, arquivamento dos dados em disco magnético.
- **"Back-up":** reprodução do arquivo armazenado em disco magnético, em fita magnética para fins de segurança.

Através do presente esquema o banco de dados foi implantado gradativamente.

4. EDUCAÇÃO DO USUÁRIO

Após o armazenamento de uma quantidade razoável de artigos, iniciamos o trabalho de propaganda do Sistema e de educação do usuário.

Um dos artifícios utilizados para fazer a propaganda do Sistema e fazer com que os usuários passassem a conhecer e desfrutar das vantagens do mesmo, foi a solicitação de preenchimento de um formulário no qual os usuários deveriam relacionar os assuntos do seu interesse. À medida em que os formulários preenchidos eram devolvidos, nos dirigíamos ao terminal onde construímos estratégias de busca para esses assuntos, de forma recuperar, uma coleção de referências bibliográficas relevantes ao assunto em questão. De posse da listagem de artigos relacionados com assuntos do seu interesse, os usuários analisavam a qualidade dos itens, e, satisfeitos com os resultados, motivavam-se a fazer, eles próprios, novas consultas ao BD do Sistema.

Outros artifícios foram utilizados para atrair e educar os futuros usuários do SCAP: comunicações internas, boletins do Núcleo, propagandas afixadas em quadro de aviso. Nenhuma delas no entanto, foi tão eficaz quanto a prática do formulário.

A aceitação do SCAP por parte dos seus usuários favoreceu o prosseguimento dos trabalhos de implantação do banco de dados. Já nesta fase temos contado com a participação do pessoal especializado, que, ciente da utilidade do Sistema, motivou-se a cooperar em um trabalho que somente reverteria em seu próprio benefício.

5. CONCLUSÕES

Testes foram aplicados ao banco de dados com o propósito não só de determinar o nível de desempenho do mesmo, mas também com o propósito de verificar o que deveria ser feito para elevar esse nível.

Os testes revelam que o fator que determinou a obtenção de um nível de recuperação de informações satisfatório, foi, sem dúvida, a linguagem utilizada para caracterizar o conteúdo dos artigos. Os "unitermos" têm apresentado resultados satisfatórios em termos da recuperação de itens relevantes a um assunto em questão. No entanto, observam-se problemas, decorrentes principalmente de:

1. ocorrência irregular de sinônimos.
2. sintaxe, devido a falsa ou incorreta coordenação dos termos.
3. excessiva quantidade de termos chaves atribuídos aos artigos, resultando em alguns casos, na recuperação de um grande número de documentos, no entanto pouco relevantes.

Os problemas do tipo 1 e 2 deverão ser minimizados por ocasião de uma revisão no dicionário de termos chaves. Um certo controle no vocabulário deverá ser efetuado, embora o conceito de "unitermos" deva ser mantido já que a experiência os revelou satisfatórios.

Problemas do tipo 3 serão também minimizados com a participação do pessoal especializado e técnicos em informação, no controle da atribuição de termos-chaves aos artigos.

Contínuas avaliações do desempenho do Sistema serão efetuadas com o intuito de aprimorá-lo e mantê-lo responsável às necessidades da comunidade científica.

Abstract

Implementation of the SCAP data base at Núcleo de Computação PUC/RJ: information representation; procedures for information storage; user's acceptance; comments on the system performance.

6. REFERÊNCIAS BIBLIOGRÁFICAS

CARVALHO, A.V. — **Um sistema conversacional de consultas para artigos de periódicos.** Rio de Janeiro, Pontifícia Universidade Católica, 1975. Tese de Mestrado em Informática.

LANCASTER, F.W. — "MEDLARS": Report on the evaluation of the operating efficiency. *American Documentation*, 20 (2) April, 1969.

— **Vocabulary control for information retrieval.** Washington, D.C., Information Resources Press, 1972.