

TENDÊNCIAS EM INFORMÁTICA PARA PROCESSAMENTO DE INFORMAÇÕES NO CENTRO DE PESQUISA E DESENVOLVIMENTO DA TELEBRÁS

EDUARDO TADAO TAKAHASHI

Centro de Pesquisa e Desenvolvimento, CPqD
Telecomunicações Brasileiras S.A., TELEBRÁS
13100 Campinas, SP

O Centro de Pesquisa e Desenvolvimento da TELEBRÁS tem dedicado considerável esforço à organização de uma infra-estrutura em informática, que viabiliza, entre outras metas, a automatização, em considerável grau, de tarefas de tratamento de informações. Este trabalho descreve sucintamente as diretrizes traçadas, as atividades já desenvolvidas e as perspectivas da área no CPqD.

1. INTRODUÇÃO

A infra-estrutura necessária para a automatização do tratamento de informações técnicas (itens bibliográficos, patentes, documentos de projeto/produto, etc.) em um centro de P & D é extremamente complexa. A grosso modo, há três diferentes facetas que compõem essa infra-estrutura, a saber:

- (a) equipamentos: que equipamentos de processamento central (grandes computadores, minis, micros) deverão ser utilizados? Que tipo de periféricos (terminais, impressoras, etc.) estarão conectados aos equipamentos centrais? Que outros equipamentos (equipamentos de microfilmagem, discos óticos, etc.) deverão completar a infra-estrutura?
- (b) *software*: qual será a arquitetura do sistema (centralizada, distribuída)? Como será o processamento (*on-line/batch*)? Que funções deverão ser incluídas no sistema?
- (c) convenções de O & M: que convenções serão adotadas para a classificação, indexação e descrição de referências? Que procedimentos implementarão essas convenções em conexão com o uso do sistema de *software*?

O planejamento inicial da infra-estrutura em informática do CPqD para o processamento de informações técnicas, considerando as três facetas acima mencionadas, defrontou-se com os seguintes problemas:

- diversidade de equipamentos, cada qual incompatível até certo ponto, com similares, e inadequados para as características ideais de uso no Brasil. Por exemplo, não há uma convenção uniforme para o processamento de "ç" e acentuação nas impressoras nacionais;
- ausência de *software* com as características desejadas, incluindo:
 - disponibilidade em programa-fonte;
 - portabilidade para equipamento de diversos fabricantes e portes (ex: DEC-10 e micro de 8 bits);
 - extensibilidade para permitir integração a aplicações paralelas;
- inexistência ou volatilidade de convenções básicas em que basear o sistema, como por exemplo:
 - vocabulário padronizado para a indexação de referências,
 - processamento de caracteres especiais da língua portuguesa.

Isto posto, o CPqD viu-se na contingência de limitar-se à situação circundante e desenvolver um sistema que refletisse essa situação ou, por outro lado, engajar-se em um trabalho mais profundo de organização de infra-estrutura básica, a par do desenvolvimento do sistema desejado.

A segunda alternativa foi a escolhida, e os capítulos posteriores detalham a implementação dessa decisão.

2. INFRA-ESTRUTURA BÁSICA

Em função da decisão tomada, considerável esforço foi investido em infra-estrutura básica, compreendendo três atividades distintas, a saber: a estruturação de um *thesaurus* próprio, a definição de bases de dados e a padronização da interface de aplicações com o usuário.

2.1 Estruturação de *thesaurus*

A padronização de um vocabulário técnico é essencial para um ambiente multidisciplinar que processa e mesmo gera informações técnicas.

Tendo já arcado com os problemas resultantes da indexação livre de documentos em um experimento anterior, decidiu-se estruturar um *thesaurus* de descritores voltado para as necessidades do CPqD. Tal *thesaurus*, já estruturado, foi baseado no SPINES da UNESCO (1), com adições baseadas no INSPEC (2), e se encontra hoje com os termos já carregados *in totum* no DEC-10 do CPqD, estando em curso ainda a tarefa final de revisão de RTs. As atividades especificamente relacionadas com o *thesaurus* do CPqD estão descritas em (3) e deverão ser publicadas proximamente.

2.2 Definição de bases de dados

Sendo o armazenamento e a recuperação eficiente de referências um requisito essencial para todas as aplicações na área, foram desenvolvidos três grupos de rotinas que permitem a criação, modificação e eliminação de referências em bases de dados:

- a) Base de dados no formato CPqD (BDFC): este conjunto de rotinas permite a organização de bases de dados hierárquicas para quaisquer tipos de itens no CPqD, incluindo primordialmente referências oriundas da biblioteca, do grupamento de patentes e do grupamento de documentação técnica. O conjunto é utilizado em diversas aplicações no DEC-10 (como veremos adiante), e deverá ser transportado para um microcomputador de 8 bits, executando CP/M, no início de 85.
- b) Bases de Dados Relacional (BDRE): este conjunto de rotinas permite a organização de bases de dados *relacionais* de uso absolutamente geral. Seu uso pretendido será a classe de aplicações *aparentada* a processamento bibliográfico, mas na direção de *query systems* e sistemas de apoio à decisão. De uma maneira geral, portanto, as aplicações de maior interesse para a Biblioteconomia deverão estar colocadas no uso de BDFCs, provendo-se para aplicações especiais a conversão automática de BDFCs para BDREs. O conjunto de rotinas de BDRE foi adaptado de (4) para o DEC-10, e executa igualmente em microcomputadores de 8 bits executando CP/M e microcomputadores de 16 bits executando MS-DOS, garantindo assim a portabilidade das aplicações nele baseadas.
- c) Base de Dados CAL (BDCA): finalmente, há um conjunto de rotinas que permitem a organização de bases de dados em formato CALCO, utilizadas especificamente para intercâmbio com o *mundo exterior*. Isto significa que deverá haver o desenvolvimento de rotinas de conversão entre BDCA e o formato interno do CPqD.

2.3 Padronização de interface com o usuário

Outro problema comum na automação desordenada de atividades é a introdução de sistemas com uma multidão de convenções para a interação com usuários.

Visando padronizar tais convenções para todas as aplicações, e ao mesmo tempo reduzindo o tempo necessário para o desenvolvimento de novos sistemas, implementou-se um *controlador de menus* de caráter geral, que opera sobre descrições de menus armazenadas como constantes, independentes do código, e de fácil alteração. Pretende-se estudar, no futuro, a possibilidade de se automatizar a própria definição de menus através de um programa conversacional que interagirá com o projetista.

3. APLICAÇÕES

Presentemente, as seguintes aplicações estão operacionais, todas baseadas na utilização da infra-estrutura apresentada no capítulo anterior:

- a) Controle de documentos: sistema que permite estruturar um produto em uma árvore de itens componentes e associar a cada item um número arbitrário de documentos técnicos. Entre outras informações, o sistema permite controlar, para cada item/documento, seu estado, localização física, autores, eventos, etc.
- b) Controle de *thesauri*: sistema que permite estruturar *thesauri* de termos agrupados em polígonos, onde cada termo terá versões em outras línguas, bt's, nt's e rt's em número arbitrário, etc.
- c) Codificação de itens: sistema que orientará o usuário na atribuição de códigos de itens e inserção dos mesmos em bases de dados. O código de item CPqD é amplo o suficiente para acomodar itens de qualquer natureza, incluindo documentos técnicos e referências em geral.
- d) CALCO: sistema que permite criar, atualizar e listar bases de dados bibliográficos no formato CALCO, incluindo opções de relatórios padronizados e de geração de fitas para intercâmbio com outras instituições.

4. PERSPECTIVAS

Há ainda em fase de análise preliminar um projeto de automação de controle de disseminação, que deverá utilizar boa parte da infra-estrutura criada. Segundo as nossas primeiras decisões, esse sistema, à semelhança do de controle de itens (em que será baseado), deverá operar em microcomputadores e no DEC-10 sobre bases de dados redundantes.

As conversões de formato para intercâmbio externo deverão igualmente ser abordadas em breve, e a integração dos sistemas logo após.

Finalmente, convém mencionar linhas de desenvolvimento paralelo que, em breve, deverão ter forte impacto sobre as atividades aqui descritas. Tais linhas incluem a pesquisa exploratória em *spelling correction*, composição de textos, *query languages*, etc. presentemente em curso.

A experiência do CPqD em processamento bibliográfico é, conforme se depreende do acima exposto, bastante profunda, em consequência direta da abordagem adotada. É portanto razoável prever que, em futuro imediato, o CPqD venha a prestar serviços à comunidade em geral, não só como centro de informações na área em que atua, como também na transferência de *know-how* em automação de processamento bibliográfico.

Abstract:

Trends in informatics for information processing in the Centre for Research and Development of the Brazilian Telecommunications Corporation (TELEBRÁS)

EDUARDO TADAO TAKAHSHI

Describes briefly the developed directives, the activities and trends for information automation in the Centre for Research and Development of the Brazilian Telecommunication Corporation (TELEBRÁS).

REFERÉNCIAS

1. UNESCO. SPINES Thesaurus. Unesco, 1976. Vols. I a III.
2. IEE. INSPEC Thesaurus. IEE, 1981.
3. CPqD. Atividades de organização de thesuri no CpQD. Documento interno, CPqD, 1982.
4. SWAN, T. Pascal Data Base Management System. Hayden, 1983.