

MISTRAL: SISTEMA DE RECUPERAÇÃO DOCUMENTÁRIA

JEAN HABRAN

Setor de Projetos e Serviços
ABC BULL S.A.
01000 São Paulo, SP

O Sistema de recuperação documentária, MISTRAL, é um sistema integrado de documentação automatizado, que inclui: busca retrospectiva, disseminação seletiva da informação, entrada em linha de documentos, indexação automática, gestão de léxico e tesouros, interface para videodisco, máquinas de microfilmagem e de fotocomposição. Permite a organização dos maiores bancos de dados na Europa e no mundo.

1. INTRODUÇÃO

O volume de dados publicados aumenta diariamente. Essa *inflação* de documentos, pesquisas, descobertas e normas, em vez de simplificar, complica sensivelmente o trabalho dos bibliotecários, juristas, pesquisadores, cientistas e engenheiros, em função da impossibilidade de assimilar todas as informações e da dificuldade de se ter acesso aos documentos pertinentes em tempo hábil.

Existem também problemas complexos para repertoriar, classificar e estocar as informações.

Para ajudar os especialistas a recuperar rapidamente a informação, sem *ruído* nem perda de tempo, foram criadas ferramentas como o MISTRAL, que permitem realizar a busca retrospectiva e a disseminação seletiva das informações.

A busca retrospectiva possibilita recuperar uma informação cadastrada através de perguntas feitas pelo próprio usuário, com possibilidades de ampliar ou restringir o campo de pesquisa. A informação pode ser materializada pela fonte primária, ou seja, o texto original, ou através de uma referência contida numa fonte secundária para permitir a localização do documento original.

A disseminação seletiva possibilita a emissão periódica de listas de referências de documentos recebidos ou disponíveis, concernentes a um determinado perfil do usuário. É a maneira de informar sistematicamente os membros da organização.

2. CONCEITOS

No caso específico do MISTRAL, existe a possibilidade de cadastrar tanto o texto original como uma referência secundária.

As informações são estruturadas num banco de dados (*databank*), em discos magnéticos, ocupando pouco espaço físico e possibilitando o acesso rápido, em segundos. No caso de texto integral, trata-se de banco de dados textual; no caso de resumo, fala-se de banco de dados bibliográficos.

O banco de dados MISTRAL pode ser dividido em até 25 bases (*database*). Os assuntos tratados em cada base podem ser classificados em até 50 domínios e cobrir até 50 períodos de tempo. Um banco de dados MISTRAL pode reperi-toriar de 4 a 32 milhões de documentos, dependendo do comprimento da referência unívoca MISTRAL que acompanha cada documento e que pode chegar a 34 caracteres alfanuméricos.

O suporte básico da informação é o documento, que pode, por sua vez, ser subdividido em até 149 campos, cada um deles contendo até 65.535 caracteres alfanuméricos, o equivalente a 10.000 palavras portuguesas. São previstos 50 níveis de confidencialidade (acesso restrito a usuários habilitados), a nível de documentos, e 50 outros a nível de campo.

3. INDEXAÇÃO DOS DOCUMENTOS

A indexação dos documentos pode ser manual, semi-automática ou automática.

No caso de indexação manual, as palavras-chave são escolhidas pelo próprio documentalista e agrupadas num campo específico.

A indexação semi-automática é geralmente realizada no texto de um resumo. Após eliminação dos separadores (espaço e sinais de pontuação) e das palavras ditas vazias (artigos, conjunções, etc...), os outros termos são comparados a um vocabulário controlado, para verificar se eles podem ou não servir de indexadores do documento.

A indexação automática permite a análise direta, num texto em formato livre, através de ferramentas lingüísticas que restabelecem o contexto das palavras, verificam a concordância dos adjetivos, reduzem os verbos à forma infinitiva e extraem a raiz dos termos. São então selecionados automaticamente os descritores que servirão para indexar o documento.

As palavras-chave, descritores ou indexadores são reagrupadas em léxicos ou listas dos termos utilizáveis para a recuperação e a disseminação dos documentos.

Se os termos precisam ser inter-relacionados, são constituídos *Thesauri* para estabelecer e manter relações dos tipos hierárquico, sinonímia, vide também, composto de, componente, modificado por, modificando, além das equivalências de termos em línguas estrangeiras.

MISTRAL autoriza 50 léxicos e *Thesauri* por base de dados.

4. CRIAÇÃO E MANUTENÇÃO DE DOCUMENTOS

A introdução e a atualização de documentos podem ser realizadas em processamento por lote ou em tempo real, através de terminais.

No caso de introdução interativa, o usuário pode escolher as funções através de *menu*, e telas autoexplicativas são visualizadas para ajudar a digitação.

Geralmente a indexação é realizada automaticamente na introdução do documento no terminal.

5. RECUPERAÇÃO

A busca retrospectiva, que pode também ser efetuada nas duas modalidades, autoriza a interrogação de maneira conversacional ou transaccional, e oferece recursos de adaptação ao nível de conhecimento do sistema por parte do usuário final. Podem ser usadas até 12 línguas diferentes para a consulta e a monitoração do usuário.

A pesquisa é realizada passo a passo, por perguntas combinando operadores booleanos, termos dos léxicos e resultados de pesquisas anteriores. É possível truncar os termos e ampliar ou diminuir o campo da pesquisa usando termos mais genéricos ou mais específicos dos *Thesauri*. Após uma pré-seleção de parte dos documentos de uma base, pode ser efetuada uma pesquisa direta sobre o texto, com utilização de fatores de proximidade, presença ou ausência de termos na mesma frase, mesmo parágrafo, etc. Desta maneira, o campo de pesquisa é *afinado* e o usuário tem acesso à informação pertinente, sem *ruido*. O resultado da busca pode ser imediatamente visualizado ou impresso. A estratégia da pesquisa é memorizada pelo sistema para facilitar a automatizar qualquer busca posterior sobre o mesmo assunto.

6. DISSEMINAÇÃO SELETIVA

Através desse recurso, todos os novos documentos incluídos no banco de dados podem ser emitidos periodicamente e enviados aos interessados. Para tal fim, os usuários, setores e departamentos da organização são previamente cadastrados e para cada um deles é definido um perfil de interesse.

7. IMPRESSÃO

Os documentos selecionados pela busca retrospectiva ou através do perfil do usuário, em caso de disseminação seletiva, podem ser classificados e personalizados a nível de *lay-out*, títulos, cabeçalhos e endereço do destinatário no momento da impressão.

No caso da busca retrospectiva, os comandos de composição de textos podem ser fornecidos interativamente e memorizados para uso posterior; a impressão

pode ser realizada em modo *on-line*, em terminal impressora, ou *off-line*, usando uma das impressoras do computador central.

8. INTEGRAÇÃO COM OUTROS SISTEMAS

MISTRAL pode também ser integrado a outros *software*, como por exemplo a SYMPHONIE, *software* de composição eletrônica de texto que permite a paginação e a edição dos léxicos, *Thesauri* e até parte do banco de dados bibliográfico, em máquina de fotocomposição.

SYMPHONIE pode ser usado tanto em ambiente de automatização de escritórios como também em jornais, universidades, etc.

MISTRAL prevê também uma interface com máquina de fotocópias, microfimagem e videodisco.

9. REFERÊNCIAS

Alguns exemplos de utilização do MISTRAL em áreas específicas são materializados pelos bancos de dados:

- . QUESTEL, que reagrupa os principais bancos de dados existentes na França e alguns americanos, para permitir a teleconsulta;
- . NORIANE, que concerne às normas francesas e normas internacionais ISO de responsabilidade do AFNOR;
- . GCM, que se refere a jurisprudência;
- . INPI, que fornece informações sobre patentes;
- . PASCAL, que reagrupa várias bases científicas multidisciplinares do CNRS/ Informascience;
- . CANCERNET, onde são cadastrados os resultados das pesquisas e da luta contra o câncer;
- . EURIS, que permite o acesso às informações da CEE (Comunidade Econômica Européia), incluindo documentos internos, legislação e comércio entre os países-membros.

10. UTILIZAÇÃO

Existem inúmeros casos de utilização do MISTRAL. Pelo fato de este *software* manipular documentos e permitir a criação de um banco de dados de textos, é uma ferramenta poderosa em automação de escritórios.

As correspondências, normas, estratégias e táticas da empresa poderão ser registradas no banco de dados, para permitir uma consulta imediata através de terminais convencionais, de terminais tipo *office automation* ou de terminais de videotexto.

Por intermédio da estruturação da referência MISTRAL ou através de pseudo palavras-chave será possível reagrupar os documentos em *pastas eletrônicas*, permi-

tindo o acesso por ordem cronológica, por assunto, por departamento ou por consulta direta sobre o texto.

A tomada de decisão também pode ser apoiada no uso do MISTRAL. É efetivamente possível constituir bancos de dados econométricos, onde, em alguns segundos, é possível recuperar e combinar parâmetros complexos. A utilização de números complementados por atributos permite a colocação dos dados nos seus respectivos contextos.

A utilização de MISTRAL em bibliotecas públicas ou de empresas autoriza o acesso à matéria pertinente em pouco tempo. Pode ser selecionado um livro, um artigo de uma revista ou apenas um parágrafo que cite uma portaria ou norma.

Não é necessário registrar a fonte primária, mas apenas uma referência que permita localizar rapidamente o livro na biblioteca.

Em empresas de engenharia, MISTRAL poderá permitir auxiliar a resposta a editais de concorrência. É efetivamente possível cadastrar os *currícula vitae* dos engenheiros por nível de formação acadêmica e projetos que chefiam, estocar as referências da empresa em obras públicas e até ter uma descrição dos processos industriais utilizados e uma avaliação dos respectivos custos. Em poucos minutos será possível responder às perguntas principais baseadas na experiência das empresas e das referências. Uma consulta adicional a bancos de dados externos de normas e patentes permitirá assegurar-se da viabilidade atual da solução.

Uma experiência interessante foi realizada na França, numa rede de hospitais. Após a saída de um paciente é criada uma ficha hospitalar que, além dos dados cadastrais, contém um resumo dos antecedentes médicos, os pré-diagnósticos, os medicamentos empregados e o diagnóstico final. Inúmeras estatísticas e constatações podem ser geradas a partir desses elementos: incidência e evolução de doenças por faixas etárias, efeitos colaterais de medicamentos, coincidência de diagnósticos, uso secundário de medicamentos em pré-diagnósticos incompletos.

Outros tipos de uso de MISTRAL podem ser encontrados nos setores militares, policiais, nas prefeituras, para cadastrar as mídias dos meios de comunicação (filmes, televisão, imprensa), para proteção ao crédito, jurisprudência, etc.

MISTRAL opera sob GCOS (*General Comprehensive Operating System*), em equipamentos ABCC BULL.

Abstract

MISTRAL – System of documentary retrieval

The system of documentary retrieval, MISTRAL, is an integrated system of automated documentation, which includes: retrospective searching, selective dissemination of information, management of lexicals and thesauri, interface for videodisk, microfilm and photocomposition machines. Allows the organization of the main data bases in Europe and in other parts of the world.