

LINCE: uma proposta de padronização para linguagens de acesso a base de dados

Instituto Brasileiro de Informação em Ciência e Tecnologia – Brasília-DF
Comissão Nacional de Energia Nuclear - CNEN - Rio de Janeiro-RJ
Centro de Informações Nucleares - CIN - Rua General Severiano - Rio de Janeiro-RJ

Resumo – A falta de padronização nos procedimentos de acesso a bases de dados levam os usuários a capacitar-se nas características próprias de cada linguagem de acesso, existente nos diversos centros hospedeiros dessas bases. Para superar esta dificuldade, o Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT e Centro de Informações Nucleares da Comissão Nacional de Energia Nuclear – CIN/CNEN vêm trabalhando no projeto da linguagem LINCE, que deverá ser implantada em diferentes computadores. É orientada a comandos e tem sua potencialidade voltada para o tratamento de dados referenciais e bibliográficos, porém, incorpora algumas facilidades para a recuperação de dados numéricos. Seus principais comandos permitem realizar buscas, exibir, ordenar ou gravar seus resultados, abrir ou fechar sessões de trabalho, usar diferentes bases e solicitar auxílio ao sistema. Nas expressões de busca são utilizados operadores lógicos, relacionais, de relacionamento em tesouro, de truncamento, mascaramento e de profundidade. Um subconjunto desta linguagem está implementado nos computadores do IBICT e encontra-se em fase de desenvolvimento do CIN/CNEN.

Palavras-chave - Linguagem de Recuperação, Padronização, Sistemas de Recuperação de Informação, Busca Bibliográfica, Linguagem Comum de Recuperação, Linguagem Orientada a Comandos – Padronização.

1 Introdução

Este documento contém uma descrição sucinta da proposta do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)* e do Centro de Informações Nucleares da Comissão Nacional de Energia Nuclear (CIN/C-

* Especialistas do IBICT envolvidos no estabelecimento da LINCE: Carlos Alberto Mamede Hernandes, Cláudia Canoglia, Eloísa Príncipe de Oliveira, Elza Maria Ferraz, Hélio Kuramoto, Ilza Leite Lopes, Jayme Leiro Vilan Filho, Leda Maria Louzada Melgaço, Maria Cristina Almeida Correa, Marlene Freitas de Santana, Marlene Machado da Silva, Sérgio Dagnino Falcão, Sílvia de Oliveira Barcellos e Sônia Regina Burnier

NEN)** de definição de uma linguagem comum de recuperação em linha, para dados referenciais e bibliográficos.

Trata-se de uma linguagem orientada a comandos, genérica, projetada para ser independente da instalação na qual será implementada e das estruturas das bases de dados a serem pesquisadas. Com tais características, será avaliada como um padrão a ser adotado em uma rede de acesso público a bases de dados, projeto que vem sendo desenvolvido pelo IBICT em conjunto com outras instituições. Este projeto está no âmbito do Subprograma de Informação em Ciência e Tecnologia do Programa de Apoio ao Desenvolvimento Científico e Tecnológico (PADCT).

Baseada em normas internacionais [5,6] esta linguagem procura ser a mais abrangente possível, englobando todas as funções próprias das atividades de recuperação de informações. Cada centro hospedeiro de bases de dados, que vier a fazer parte do sistema de acesso público a bases de dados, poderá implementar parte desta linguagem de acordo com sua conveniência, desde que incorpore um subconjunto mínimo de funções em fase de definição.

2 Características Gerais da Linguagem

A linguagem comum de recuperação de informações em linha (LINCE) serve de interface entre usuários e bases de dados. Possibilita a definição de uma série de parâmetros que facilitam os procedimentos inerentes à busca. Ao se conectar ao sistema, é aberta uma sessão durante a qual o usuário, através de comandos, poderá executar diferentes tarefas tais como: realizar buscas, exibir ou gravar referências recuperadas, reinicializar ou encerrar sessões, etc.

Como resposta a uma consulta, será informado o número de documentos recuperados e, opcionalmente, a relação dos tempos que satisfaçam a busca baseada em termos truncados.

Logo após a consulta, pode-se exibir os documentos recuperados e, à medida que são mostrados, é possível selecioná-los para posterior utilização.

Dentro de uma sessão, o usuário pode abrir uma subsessão para, por exemplo, atender a um outro usuário. A abertura de uma nova subsessão

** Especialistas do CIN/CNEN envolvidos no estabelecimento da LINCE: Elizabeth Cunha Lima, Carlos Cesar Fernandes, Carlos Henrique Marcondes de Almeida, José Augusto Alves Bernacchi, Luís Fernando Sayão e Norma Tricárico Orosco

provoca o encerramento da anterior. Terminada esta, retorna-se ao ambiente da sessão.

A linguagem permite, também, o uso de macros para facilitar a digitação de comandos. Com esta função é possível, por exemplo, nomear um comando ou parte dele, de forma a que toda vez que o usuário se referir a este nome, o sistema interpretará como uma menção à porção do comando que recebeu o nome mencionado. Suas definições serão mantidas no decorrer de uma sessão ou subsessão.

Alguns comandos oferecem a possibilidade de serem executados posteriormente. Esta modalidade deve ser utilizada quando a tarefa a ser realizada implicar um intenso uso de recursos do sistema, evitando uma queda no seu desempenho.

3 Expressão de Busca

Na formação das expressões de busca, são permitidos os seguintes operadores: lógicos, de relacionamento em tesauro, relacionais, de truncamento e mascaramento e operadores de proximidade. Pode-se utilizar também a identificação dos campos nos quais a expressão deve ser avaliada, assim como parênteses para mudança de precedência dos operadores.

Além desses, pode-se utilizar as palavras reservadas MAX e MIN quando desejar recuperar, respectivamente, as referências que contenham o maior ou menor valor de conteúdo no campo especificado. Desta forma SALARIO=MAX recuperará todas as referências que possuam o maior valor para o campo de salário.

3.1 Operadores de Mascaramento e Truncamento

Visam a facilitar a especificação, nas expressões de busca, de termos que possuam um conjunto de caracteres em comum. Estes operadores podem aparecer no início, meio ou final do termo. São eles:

o ponto indica a posição onde o número ilimitado de caracteres poderá ser incluído no termo.

indica a posição e o número exato de caracteres que deverão ser incluídos no termo (cada “#” substitui exatamente um caractere).

\$ indica a posição e o número máximo de caracteres que deverão ser incluídos no termo (cada “\$” substitui no máximo um caractere).

3.2 Operadores de Tesouro

Permitem formular buscas apoiadas em vocabulários controlados com estrutura hierárquica. São os seguintes os operadores de tesouro:

TG n – refere-se aos termos genéricos em relação a um determinado termo;

TE n – refere-se aos termos específicos em relação a um determinado termo;

TR – refere-se aos termos que não pertencem à mesma hierarquia de um determinado termo, porém tem alguma outra relação com este.

O parâmetro n é um número que, quando especificado, indica a profundidade com que será feita a pesquisa no vocabulário.

O uso destes operadores tem o mesmo efeito de uma expressão de busca que seja composta por todos os termos do tesouro que atendam às relações especificadas, conectados por operadores de união.

3.3 Operadores de Proximidade

Permitem a realização de buscas com a especificação da posição relativa entre termos (operandos). São os seguintes os operadores de proximidade:

ADJ – estabelece que os termos deverão estar em posição contígua e na ordem indicada.

SEG n – estabelece que os termos deverão estar separados por até n palavras, na ordem indicada.

SEG=n – similar ao operador anterior, sendo que os termos deverão estar separados por exatamente n palavras.

DIS n – indica a distância máxima entre os termos, independentemente da ordem em que apareçam nos documentos.

DIS=n – indica a distância exata entre os termos, independentemente da ordem em que apareçam nos documentos.

Os operadores de proximidade também admitem como operandos listas de termos.

3.4 Operadores Lógicos

Permitem formular expressões de busca combinando logicamente dois ou mais termos.

São eles, em ordem decrescente de precedência: subtração (-) (NÃO); interseção (&), (E), ou exclusivo @ (OUE) e união (+) (OU).

3.5 Operadores Relacionais

Permitem estabelecer limites ou intervalos que qualificam o conteúdo do campo e podem referir-se a campos numéricos, alfabéticos ou alfanuméricos. Estão disponíveis os seguintes operadores relacionais: igual a (=) (IG); superior a (>) (SU); inferior a (<) (IN); inferior ou igual a (\geq) (II); superior ou igual a (\leq) (SI); diferente de (<>) (DI).

4 Comandos

São identificados por uma palavra seguida ou não de opções e determinam uma ação a ser executada pelo sistema. Os nomes dos comandos e as demais palavras de uso restrito à linguagem podem ser abreviados usando-se o número de letras necessárias à sua identificação.

4.1 Comando de Busca

Identificado pela palavra PROCURE, localiza na base de dados todas as referências que satisfaçam à expressão de busca e ao seu refinamento, quando utilizado.

Usando-se o caractere “?” após a expressão de busca o sistema mostra o número de ocorrência de cada operando. São exibidos também os números máximo e mínimo de documentos que poderão ser recuperados. Se o usuário desejar que a busca seja executada, utilizará o comando EXECUTE.

O operador TEXTO, seguido de uma nova expressão de busca, permite refinar buscas anteriores através de pesquisa em textos livres.

Ex 1: PROCURE tit = (brasil + França) & tecno.)

recupera os documentos que contenham no campo de título os termos “brasil” ou “França” associados aos desdobramentos dos termos que tenham a raiz “tecno”;

Ex 2: PROCURE C5 & ano 1985

recupera, dentre os documentos da consulta 5, aqueles cujo conteúdo do campo ano seja superior a “1985”;

Ex 3: TEXTO res = ciência

recupera, dentre os documentos da consulta anterior, aqueles que contenham o termo “ciência” no campo de resumo.

4.2 Comando de Exibição de Resultado de Busca

Identificado pela palavra MOSTRE, exhibe na tela o resultado de uma busca. Permite a inclusão de indicações que definem opções como, por exemplo,

um subconjunto dos documentos recuperados numa consulta. O formato em que deverão ser apresentadas as informações na tela também pode ser indicado pelo usuário.

Após a exibição de cada tela, o sistema estará apto a receber os comandos de controle de exibição, explicados nos itens seguintes.

Ex 1: MOSTRE C3 1,3,7:10 aut, ent
 exhibe os autores e nomes de entidades constantes nos documentos 1,3 e de 7 a 10 recuperados na consulta de número 3.

4.3 Comando de Desdobramento e Pesquisa em Vocabulários

Identificado pela palavra **RELACIONE**, pode ser usado para exibir desdobramentos de termos truncados e para permitir busca de termos em vocabulários controlados.

O comando de exibição de desdobramento de termos permite a apresentação de uma lista ordenada alfabeticamente dos termos que satisfazam ao truncamento especificado. Também será mostrada a frequência de ocorrência desses termos na base de dados. Após a execução deste comando pode-se acionar os comandos **SELECIONE** e **EXECUTE** para realizar buscas com os termos selecionados.

Se o vocabulário pesquisado possuir estrutura de tesouro, pode-se especificar, para um termo, a lista de seus relacionamentos que deseja conhecer. Podem constar da lista um ou mais dos seguintes relacionamentos: termo genérico, termo específico, termo relacionado, uso de termos sinônimos, usado para e nota explicativa.

Este comando permite também a exibição dos termos selecionados para conferência.

Ex 1: RELACIONE ano = 1985:1988
 relacione os termos existentes no campo ano de 1985 até 1988, incluindo esses limites;

Ex 2: RELACIONE tit = agro.
 exhibe os desdobramentos de "agro.", no campo título com suas respectivas frequências;

Ex 3: RELACIONE VOCABULÁRIO TE2, RT alimentação
 exhibe os termos específicos até o segundo nível e os termos relacionados ao termo "alimentação".

4.4 Comandos de Controle de Exibição

Identificados pelas palavras **CONTINUE**, **AVANCE** e **SELECIONE**, pos-

sibilitam a continuação de exibição de informações que ultrapassem uma tela, avanço ou retrocesso de telas, ou a escolha de itens de informações que devam ser selecionados para uso posterior.

4.5 Comando de Execução de Busca

Identificado pela palavra EXECUTE, ativa um comando de busca pendente, de acordo com as ações efetuadas por comandos anteriores. Assim, por exemplo, se durante a execução de um comando RELACIONE tiverem sido selecionados termos de tesouros, este comando efetuará a busca na base de dados dos documentos que contenham pelo menos um desses termos.

4.6 Comando de Ordenamento de Busca

Identificado pela palavra ORDENE, classifica o resultado de uma busca de acordo com os campos e critérios especificados. Possui como opção uma lista de campos de ordenação e de rotinas de classificação que atendam a casos específicos.

Ex 1: ORDENE C3 tit, dat decrescente

ordena a terceira consulta, segundo os campos de título (chave primária) e data de publicação (chave secundária), em ordem decrescente.

4.7 Comando de Impressão

Identificado pela palavra IMPRIMA, produz saídas impressas dos documentos recuperados, de acordo com as especificações indicadas. É também possível definir um título para cada página impressa, o número de cópias, a impressão de uma consulta previamente armazenada e, opcionalmente, a estratégia de busca.

Ex 1: IMPRIMA 1,3:10 TÍTULO= "Periódicos sobre fibra ótica" CÓPIAS=2

imprime duas cópias das referências 1 e de 3 a 10 da última consulta, com o título especificado.

Ex 2: IMPRIMA C3 AUTOR, TÍTULO, DESTINATÁRIO= "Biblioteca Central"

imprime os campos de autor e título de todas as referências verificadas na consulta 3, indicando o destinatário da saída produzido.

4.8 Comandos de Abertura e finalização de Sessão ou Subsessão

Identificados respectivamente pelas palavras INICIE e ENCERRE. O primeiro abre uma nova subsessão e/ou reestabelece os valores dos atributos da sessão, previamente estabelecidos pelo usuário. O segundo desativa a sessão ou subsessão corrente.

4.9 Comando de Gravação

Identificado pela palavra **GUARDE**, permite o armazenamento temporário de todas as consultas, ou parte delas, realizadas no decorrer da sessão.

Ex 1: **GUARDE C1, C3:C5 sessão 2**

guarda em um arquivo de nome sessão 2 as consultas 1,3,4 e 5.

4.10 Comando de Reativação

Identificado pela palavra **RETOME**, recupera um arquivo armazenado anteriormente através do comando **GUARDE**.

4.11 Comando de Apagar

Identificado pela palavra **APAGUE**, elimina consultas realizadas na sessão ou subsessão corrente, definição de macros, seleção de documentos ou arquivos armazenados anteriormente.

Ex 1: **APAGUE C3,C5:C9**

apaga as consultas de números 3 e de 5 a 9 inclusive, da sessão corrente.

4.12 Comando de Mudança de Base de Dados

Identificado pela palavra **USE**, especifica dinamicamente a base a ser consultada mais de uma base de dados poderá ser ativada, sendo as buscas realizadas no conjunto dessas bases.

Ex 1: **USE inspec PERIODO=1988**

ativa o subconjunto de base inspec relativo ao ano de 1988.

4.13 Comando de Definição de Atributos da Sessão

Identificado pela palavra **DEFINA**, este comando é utilizado para alterar ou estabelecer parâmetros durante a sessão.

Dentre os parâmetros possíveis de serem estabelecidos ou alterados, pode-se destacar: o vocabulário de termos controlados a ser utilizado, o campo a ser pesquisado quando sua especificação for omitida, a composição de um campo formado por um conjunto de campos, as dimensões da tela e da página, a formação de macros e o idioma de diálogo.

4.14 Comando de Ajuda

Identificado pelo caractere “?” ou a palavra **EXPLIQUE**, visa fornecer, de forma simplificada, informações sobre a sintaxe da linguagem, e os parâmetros da sessão corrente e do sistema.

Ex 1: ?

explica como usar o comando de ajuda

Ex 2: ? COMANDOS

relaciona e explica resumidamente os comandos disponíveis

Ex 3: EXPLIQUE MOSTRE

fornece uma descrição detalhada de como utilizar o comando MOSTRE

4.15 Comando de Edição

Identificado pela palavra EDITE, permite a alteração de textos dos comandos.

5 Conclusão

A linguagem, como apresentada, não aborda a sintaxe própria de cada comando, mencionado apenas suas principais características. Apesar de não estar totalmente consolidada, uma versão preliminar encontra-se detalhada nos relatórios do projeto, incluindo sua descrição através da forma Backus-Naur (BNF). Um subconjunto desta linguagem já se encontra implantada no IBICT e atualmente é utilizada para testes e avaliações. Da mesma forma, o CIN/CNEN vem trabalhando em sua implementação no computador deste Centro. Ainda nesta fase do projeto, será realizada a interligação através da Rede Nacional de Comunicação de Dados por Comutação de Pacotes - RENPAC dos computadores das duas instituições, permitindo o acesso às bases neles hospedadas de uma forma única, sem que o usuário necessite saber, *a priori*, em qual computador reside a base a ser pesquisada.

Este projeto piloto tem término previsto para o início de 1989 e com a sua continuidade, prevista no PADCT fase II, espera-se formar uma rede de acesso público a base de dados com pelo menos seis hospedeiros de bases de dados interligados e operando de forma padronizada.

6 Referências Bibliográficas

1. BRASIL. Comissão Nacional de Energia Nuclear. Centro de Informações Nucleares. *SUPRIR; manual de instruções para usuários da Rede Nacional de Telex = Rev.* Rio de Janeiro, 1987, 19 p.
2. DATABASE descriptions. Columbus, STN International, 1988 1v.
3. GUIDE to Dialog searching. Palo Alto, Lockheed Dialog Information Retrieval Service, 1985 - 1v.
4. INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. *Módulo de recuperação de informações do Sistema Integrado de Publicações Seriadas SIPS.* Brasília, 1984. 33p.