MAC–IME–USP

Carlos Eduardo Ferreira
Sala 297A Tel.: 818 6140
e-mail cef@ime.usp.br

MAC 323 – Estruturas de Dados

Primeiro semestre de 1998

Exercício-Programa 4 – Entrega: **9 de junho**

Uma máquina de busca

Falar hoje da importância da internet é desnecessário. Todos estão ligados a essa rede mundial e a utilizam para diversos usos: educação, cultura, entretenimento, compras, etc. O grande problema é como encontrar as informações que você deseja na quantidade quase infinita disponível.

Para esse fim existem as máquinas de busca, como, por exemplo, a altavista (os nomes dados a este tipo de software são "robôs na internet", *Softbots, Wanderers* ou *Spiders*). Neste exercício-programa implementaremos uma máquina de busca semelhante, a fim de resolver o problema. A idéia é receber como entrada uma URL (como, por exemplo, o site www.linux.ime.usp.br/~cef) e o programa monta em uma primeira fase uma estrutura de busca rápida (B-árvore) com todas as palavras encontradas em documentos da URL dado. Então, o programa permite que diversas consultas possam ser feitas a essa estrutura, dizendo, para cada palavra buscada, quais os arquivos onde essa palavra ocorre. Você pode considerar que uma palavra tem pelo menos 3 letras.

Com isso, o exercício terá duas fases. Inicialmente todos os arquivos da URL são varridos e é montada uma tabela de índice remissivo com, para cada palavra que ocorre nos arquivos, a lista dos arquivos onde essa palavra é encontrada. Essa tabela deverá ser organizada como uma B-árvore. Depois, seu programa deverá permitir ao usuário buscas nesta URL, apresentando, para cada palavra dada, a lista dos arquivos que a contém (que é simplesmente percorrer a tabela de índices remissivos).