

Introdução

Muitos sistemas de busca (e.g., Google, Yahoo) usam técnicas de Recuperação de Informação para a indexação e busca rápida em uma base grande de dados (e.g., a Internet). Uma das técnicas mais usadas é a Lista Invertida (ou Índice Invertido).

Em uma lista invertida, as palavras dos documentos (e.g., casa, carro, etc) são indexadas em uma tabela hash e cada item (cada palavra) possui uma lista indicando em quais documentos ela ocorre. Por exemplo, considere os seguintes documentos e os seus conteúdos:

Documento	Conteúdo
document1.txt	Remember, the Force will be with you, always.
document2.txt	The Force is strong with you.
document3.txt	The Force is strong with this one.

Tal base de dados irá gerar a seguinte lista invertida:

Palavra	Documentos
"force"	document1.txt, document2.txt, document3.txt
"always"	document1.txt
"one"	document3.txt
"is"	document2.txt, document3.txt
"be"	document1.txt
"will"	document1.txt
"you"	document1.txt, document2.txt
"the"	document1.txt, document2.txt, document3.txt
"remember"	document1.txt
"this"	document3.txt
"strong"	document2.txt, document3.txt
"with"	document1.txt, document2.txt, document3.txt

Desta forma, para pesquisar todos os documentos que contém a palavra "**strong**", não precisaríamos pesquisar documento por documento, bastaria pesquisar na tabela hash a palavra (*chave da tabela*) e ver na lista (*valor da tabela*) quais os documentos que a contém (document2.txt e document3.txt).

Objetivo deste Trabalho

Implementar em Java uma lista invertida, que nada mais é do que uma tabela hash (classe [Hashtable](#) do Java) em que a *chave* da tabela será uma string (palavra) e em que o *valor* da tabela será uma lista encadeada de strings (lista de documentos). Para a lista encadeada, será usada a classe [LinkedList](#) do Java.

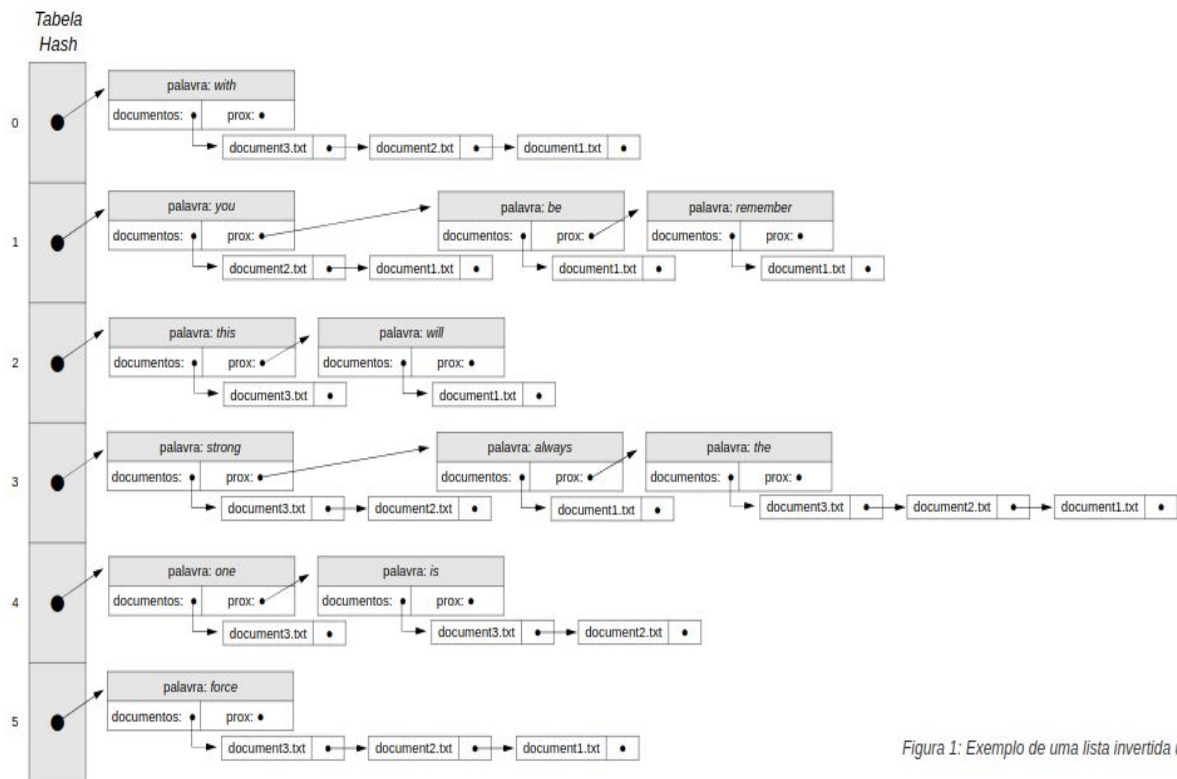


Figura 1: Exemplo de uma lista invertida usando tabelas hash.