

Recuperação de Informação

Parte II

João Lucas, Iago Raphael

Parte I - Indexação

- Arquivo invertido
- Lista índices por valores de atributo
- Formato:
 - `Atributo.value:[id,freq,[pos1,pos2...,posn]]`
- Variável “preço” discretizada

Parte I - Indexação (compressed)

- Arquivo invertido comprimido
- Não mostra frequência nem posições das palavras no documento
- Índices são “*shiftados*”, ocupando menos espaços
 - Produto.alaska: `[[165,1,[9]], [613,1,[9]]]`
 - Produto.alaska: `[166,448]`
- Tamanhos:
 - Arquivo Invertido: 163 KB
 - Arquivo Invertido comprimido: 68.6 KB

Parte III - Consulta

- Interface web com três campos de busca
 - Produto
 - Farmácia
 - Preço
- Exibe mutual information usando edit distance
 - Quanto menor a distância, maior a mutual information