

Recuperação de Informação

Projeto 2:

Indexador (*Larícia Mota - Immc2*)

Processamento de consulta (*Gabriel Vinicius - gvmgs*)

Composição da resposta (*Vinicius Emanuel - vems*)





Sites no domínio: Séries

- [Rotten Tomatoes](#)
- [IMDB](#)
- [The Movie DB](#)
- [Trakt](#)
- [TV Guide](#)
- [Metacritic](#)
- [The TV DB](#)
- [TV.com](#)

Indexador





Atributos

- Título da série
- Sinopse
- Data de lançamento
- Network (Exemplos: Netflix, HBO...)

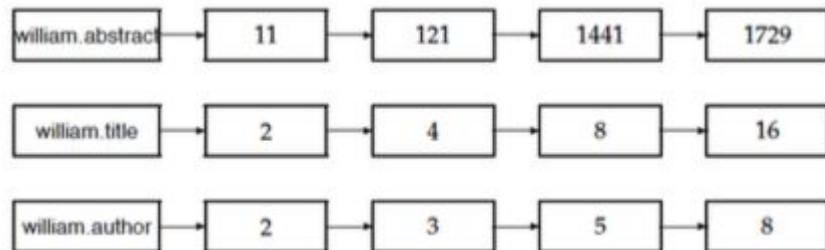


Dois índices

- Índice Termo Documento

Vocabulary	n_i	Occurrences as inverted lists
to	2	[1,4],[2,2]
do	3	[1,2],[3,3],[4,3]
is	1	[1,2]
be	4	[1,2],[2,2],[3,2],[4,2]
or	1	[2,1]
not	1	[2,1]
I	2	[2,2],[3,2]
am	2	[2,2],[3,1]
what	1	[2,1]
think	1	[3,1]
therefore	1	[3,1]
da	1	[4,3]
let	1	[4,2]
it	1	[4,2]

- Índice de Campos





Compressão

- Intervalo
 - Guardando intervalos ao invés de IDs

COMPUTER: 283154, 283159, 283202, ➡ COMPUTER: 283154, 5, 43



Versões de cada índice:

Em JSON (com e sem compressão) e em binário (com e sem compressão)

- Índice Termo Documento

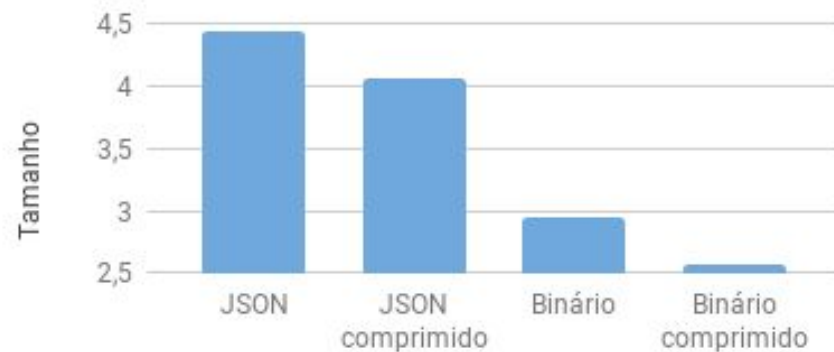
JSON sem compressão: 4,45 MB

JSON com compressão: 4,06 MB

BSON sem compressão: 2,95 MB

BSON com compressão: 2,56 MB

Índice Termo Documento





Versões de cada índice:

Em JSON (com e sem compressão) e em binário (com e sem compressão)

- Índice de Campos

JSON sem compressão: 154 KB

JSON com compressão: 130 KB

BSON sem compressão: 122 KB

BSON com compressão: 122 KB

Índice de Campos

