

# 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...)

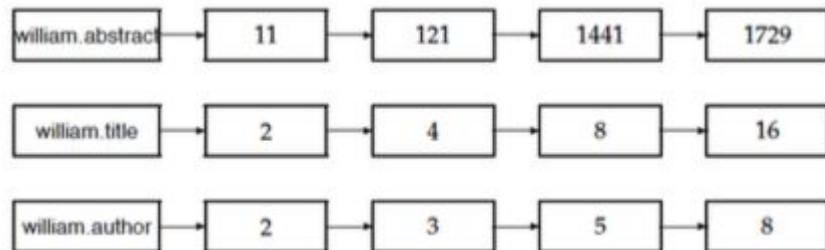


# Dois índices

- 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

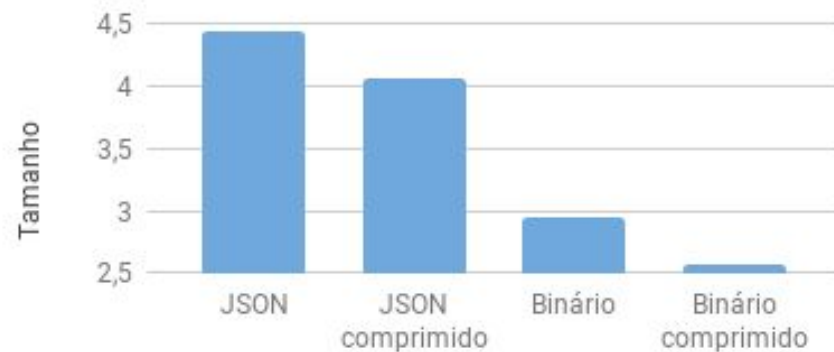
JSON sem compressão: 4,45 MB

JSON com compressão: 4,06 MB

BSON sem compressão: 2,95 MB

BSON com compressão: 2,56 MB

Índice Termo Documento





## Com codificação

- Índice Termo Documento

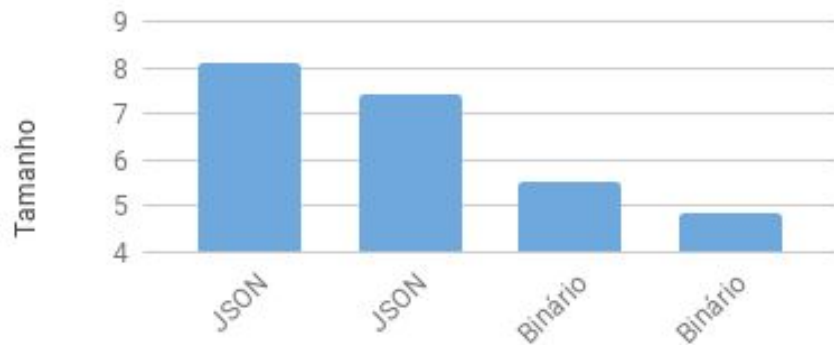
JSON sem codificação: 8,1 MB

JSON com codificação: 7,41 MB

BSON sem codificação: 5,53 MB

BSON com codificação: 4,84 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





## Com codificação

- Índice de Campos

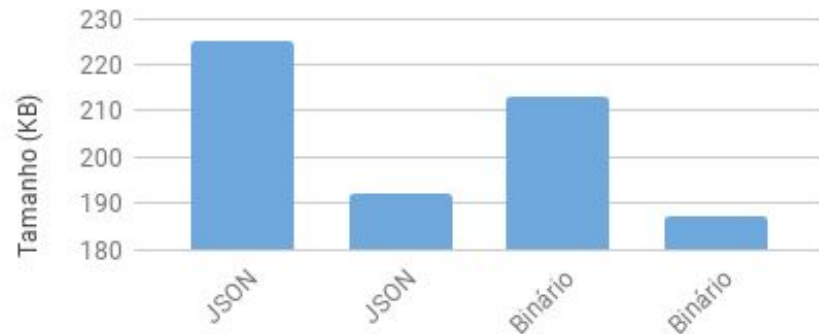
JSON sem codificação: 225 KB

JSON com codificação: 192 KB

BSON sem codificação: 213 KB

BSON com codificação: 187 KB

Índice de Campos



# Ranking





# Rankeamento

- Modelo de Espaço Vetorial
  - Sem peso por frequência
  - TFxIDF
- Composição da query
  - Considerando campos
- Leitura
  - Document-at-a-Time



# Consultas - Consulta 1

- Title
  - friends
- Storyline
  - new
  - york
  - friends
- Kendal tau
  - 0.76

0	7449604	0.959465
1	7449548	0.832212
2	7449624	0.706445
3	7449605	0.706445
4	7449629	0.703783
5	7449533	0.703783
6	7449574	0.682407
7	7449505	0.678343
8	7449632	0.500000
9	7449619	0.500000

Com TFxIDF

0	7449604	1.000000
1	7449548	0.866025
2	7449624	0.707107
3	7449629	0.707107
4	7449605	0.707107
5	7449574	0.707107
6	7449533	0.707107
7	7449505	0.707107
8	7449632	0.500000
9	7449619	0.500000

Sem Pesos



## Consultas - Consulta 2

- Title
  - how I meet your mother
- Storyline
  - new
  - york
  - friends
- Kendal tau
  - 0.68

0 7449573 0.693095

1 7449547 0.693095

2 7449526 0.611955

3 7449604 0.592399

4 7449624 0.499532

5 7449605 0.499532

6 7449629 0.497650

7 7449533 0.497650

8 7449548 0.479661

9 7449505 0.479661

Com TFxIDF

0 7449573 0.707107

1 7449547 0.707107

2 7449604 0.612372

3 7449526 0.612372

4 7449624 0.500000

5 7449629 0.500000

6 7449605 0.500000

7 7449548 0.500000

8 7449533 0.500000

9 7449505 0.500000

Sem Pesos



## Consultas - Consulta 3

- Network
  - netflix
- Title
  - things
- Storyline
  - upside
  - down
  - suspense
  - kids
  - strange
- Kendal tau
  - -1

0	7449519	0.541138
1	7449618	0.408248
2	7449621	0.408248
3	7449625	0.408248
4	7449630	0.408248
5	7449607	0.408248
6	7449610	0.408248
7	7449613	0.408248
8	7449584	0.408248
9	7449578	0.408248

Com TFxIDF

0	7449519	0.577350
1	7449618	0.408248
2	7449621	0.408248
3	7449625	0.408248
4	7449630	0.408248
5	7449607	0.408248
6	7449610	0.408248
7	7449613	0.408248
8	7449584	0.408248
9	7449578	0.408248

Sem Pesos



## Consultas - Consulta 4

- Title
  - doctor
- Storyline
  - every
  - show
  - about
  - doctor
  - and
  - hospital
- Kendal tau
  - 0.61

0	7449626	0.671060
1	7449622	0.576256
2	7449511	0.571210
3	7449514	0.558911
4	7449536	0.543294
5	7449632	0.540238
6	7449602	0.504261
7	7449572	0.504261
8	7449609	0.469014
9	7449538	0.460786

Com TFxIDF

0	7449626	0.816497
1	7449632	0.577350
2	7449620	0.577350
3	7449622	0.577350
4	7449629	0.577350
5	7449602	0.577350
6	7449609	0.577350
7	7449572	0.577350
8	7449536	0.577350
9	7449511	0.577350

Sem Pesos





# Consultas - Consulta 5

- Title
  - game of thrones
  - westworld
- Storyline
  - game of thrones west kindom
  - and cowboys
- Kendal tau
  - -1

0	7449617	0.625422
1	7449601	0.625422
2	7449571	0.625422
3	7449528	0.617325
4	7449504	0.547273
5	7449545	0.541643
6	7449529	0.523872
7	7449631	0.495981
8	7449602	0.481287
9	7449572	0.481287

Com TFxIDF

0	7449617	0.707107
1	7449601	0.707107
2	7449571	0.707107
3	7449528	0.707107
4	7449504	0.645497
5	7449602	0.577350
6	7449572	0.577350
7	7449545	0.577350
8	7449546	0.577350
9	7449529	0.577350

Sem Pesos