

Tentang IDF

$$\begin{aligned}\text{score}(Q, D) &= \sum_{t \in Q \cap D} TF \cdot IDF(t, D) \\ &= \sum_{t \in Q \cap D} (1 + \log_{10} TF(t, D)) \times \log_{10} \left(\frac{N}{DF(t)} \right)\end{aligned}\quad (2)$$

Pada persamaan fungsi *score* di atas, perhatikan bahwa $|Q| = 1$, dan $|Q \cap D| = 1$ jika dan hanya jika dokumen D mengandung terms t , satu satunya term yang diquery. Selain itu, $|Q \cap D| = 0$ dan dokumen D tersebut bernilai 0. Perhatikan bahwa nilai dari $\log_{10}(\frac{N}{DF(t)})$ akan menjadi sebuah nilai konstan yang sama untuk setiap dokumennya. Dengan kata lain, IDF tidak mempengaruhi score.

Elias-Gamma

Coding γ untuk suatu bilangan bulat positif k terdiri dari dua komponen:

- Sektor;
- Body;

Untuk mengencode:

- Cari nilai terbesar dari N , sehingga $2^N \leq X$, dengan kata lain, kita mencari Most significant bit dari X , kemudian kita akan melakukan encoding bilangan N menggunakan unary coding, itu untuk bagian sektornya. Unary coding untuk suatu bilangan N merupakan $N - 1$ buah 0 diikuti dengan satu buah bit 1.
- Bagian sisanya, yaitu $X - 2^N$, direpresentasikan dengan binary, yaitu bagian bodynya.
- Bagian sector pada intinya sama seperti penunjuk size dari bilangannya, sehingga nanti cara bacanya tinggal
- kira-kira:

```
while(*bit != 1) len++
bilangan++
while(len--> 1) bilangan <<= 1, bilangan += bit
```

Misal bilangan $7 = 2^2 + 3 = 00111$

Integer	Binary	Gamma Code	Delta Code	Fibonacci Code
1	1	1	1	11
2	10	010	0100	011
3	11	011	0101	0011
4	100	00100	01100	1011
5	101	00101	01101	00011
6	110	00110	01110	10011
7	111	00111	01111	01011
8	1000	0001000	00100000	000011
9	1001	0001001	00100001	100011
10	1010	0001010	00100010	010011

3. Average Encoding (AE)

Memori untuk mendecode suatu bilangan x menjadi $2\lfloor \log_2(x) \rfloor + 1$. Bila kita bandingkan tanpa kompresi, menyimpan suatu bilangan x membutuhkan memori sebanyak $\log_2(MAX) + 1$ untuk setiap bilangannya. Jadi, bila bilangan besarnya sedikit (atau untuk yang gapnya banyak pada postings list, kompresi ini cocok digunakan)

Elias-Delta

- Tuliskan $\lfloor \log_2(x) \rfloor + 1$ atau panjang dari bit nya dalam elias gamma, terus diikuti dengan bagian bodynya sama seperti elias- γ , perhatikan bahwa dibutuhkan:
 - $\lfloor \log_2(x) \rfloor$ untuk body
 - $2\lfloor \log_2(\log_2(x)) \rfloor + O(1)$ untuk sektor

nantinya, kalau x membengkak, ukurannya harusnya akan lebih kecil dari pada elias- γ , karena pada dasarnya payload dari bodynya itu dienkrp dengan elias- γ dengan kompleksitas memori yang konstantanya lebih besar dibandingkan dengan elias- δ

Referensi

https://www.youtube.com/watch?v=A_F94FV21Ek