

# Cara Menjalankan

Buat folder tmp, dan virtual environment, dan install requirements untuk mac dan gnu/linux:

```
mkdir tmp
python3 -m venv venv
source venv/bin/activate
pip3 install -r requirements.txt
```

Untuk windows:

```
mkdir tmp
python -m venv venv
venv\Scripts\activate
pip install -r requirements.txt
```

atau:

```
mkdir tmp
virtualenv venv
venv\Scripts\activate
pip install -r requirements.txt
```

Kalau yang windows venvnya salah commandnya, bisa di browsing kak hehe saya ga pake windows soalnya maaf kak 🙏

Selanjutnya jalankan bsbi.py

## Bonus Compression Elias-Gamma

Coding  $\gamma$  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 += 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)

Untuk benchmarking didapatkan hasil:

[illegible]

Compression Method	Waktu Rata-Rata Indexing	Waktu Rata-Rata Searching	Ukuran Dictionary	Ukuran Index
Standard (Tidak ada kompresi)	0.39821 detik	0.40845	75 KB	340.5 KB
Variable Byte Encoding	0.47410 detik	0.48441 detik	66.01 KB	48.56 KB
Elias Gamma	0.57993 detik	0.59071 detik	66 KB	41.704 KB

Dari sini kita mengetahui bahwa terdapat trade off waktu dan memori dari beberapa teknik kompresi yang dilakukan, dengan pencarian dilakukan 10 kali untuk query:

```
queries = ["olahraga", "tumor", "hidup sehat", "jantung", "penyakit",  
           "kuat", "badan sehat", "olahraga teratur", "tidur cukup"]
```

Serta dilakukan indexing 10 kali pula untuk masing-masing teknik kompresi.