

# UTS BIG DATA - SUSULAN

Program Studi Informatika, FMIPA Universitas Syiah Kuala  
2 Desember 2019

Dr. Taufik Fuadi Abidin, S.Si., M.Tech  
Alim Misbullah, S.Si., M.S

---

Pada Ujian Tengah Semester ini, anda diminta untuk mengolah data dalam sebuah file teks untuk dirangkum dan divisualisasikan dalam bentuk chart. File data dapat diunduh di <http://103.41.169.15/bd/data.txt>

File data tersebut memiliki 61.427 baris dengan gambaran data setiap baris terdiri dari 3 atribut yang dipisah dengan tanda koma. Format data kurang lebih:

[URL, pasangan akronim=>kepanjangannya, skor]

URL adalah alamat artikel dimana pasangan akronim dan kepanjangannya ditemukan, pasangan akronim dan kepanjangannya digabung dengan simbol =>, dan skor adalah bilangan real.

## Contoh:

[<http://news.okezone.com/read/2011/03/03/339/430981/mahfud-tolak-tanggapi-ancaman-recall-lily-wahid>, Unas=>upacara Wisuda Universitas Nasional, 0.899042262007784]

[<http://news.okezone.com/read/2011/03/03/339/430981/mahfud-tolak-tanggapi-ancaman-recall-lily-wahid>, Jakarta=>mafia pajak Terserah bagaimana partai, 0.725270686100131]

[<http://news.okezone.com/read/2011/03/03/339/430981/mahfud-tolak-tanggapi-ancaman-recall-lily-wahid>, JCC=>Jakarta Convention, 0.733945323017334]

[<http://news.okezone.com/read/2011/03/03/339/430981/mahfud-tolak-tanggapi-ancaman-recall-lily-wahid>, JCC=>Jakarta Convention Center, 0.952380952380952]

Anda diminta untuk mengolah data tersebut menggunakan **Python** agar dapat mendapatkan rangkuman dari data sebagai berikut:

1. Mengambil pasangan akronim dan kepanjangannya yang memiliki skor di atas atau sama dengan 0.92 dan menghimpun semua URL pendukung yang unik (tidak boleh duplikat). Pasangan akronim dan kepanjangannya yang memiliki skor di bawah 0.92 diabaikan saja.
2. Simpan data-data hasil olahan tersebut dalam sebuah file teks dengan format yang dipisah dengan koma:  
  
akronim=>kepanjangannya, skor, jumlah URL unik, daftar URL yang disusun dalam format URL1::URL2::URL3... dstnya.
3. Buat **pie chart** yang menampilkan 15 akronim dan kepanjangannya yang memiliki jumlah URL unik terbesar. Gunakan **matplotlib** yang telah dipelajari sebelumnya untuk menggambarkan chart dengan python. Amati hasil pie chart tersebut, apakah menurut pendapat anda hasil tersebut benar dan akurat. Jika sudah benar beri alasan, begitu juga jika ada yang tidak sesuai dari hasil tersebut. Tulis pendapat anda tersebut pada selembarnya terpisah.
4. Kumpulkan script python yang dibuat untuk menyelesaikan UTS ini melalui email [misbullah@unsyah.ac.id](mailto:misbullah@unsyah.ac.id). Jika ada lebih dari 1 file, mohon zip terlebih dahulu sebelum diemail.

--- Good Luck ---