
Minggu ke-13



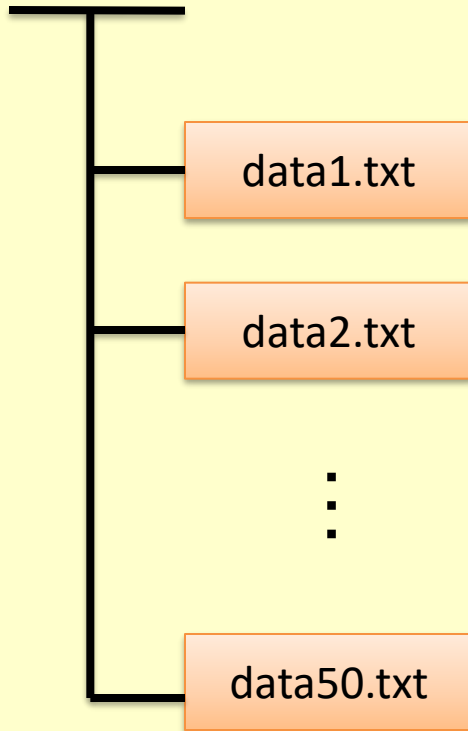
Praktikum Text Mining

Ali Ridho Barakbah, Entin Martiana, Tri Hadiah

Knowledge Engineering Research Group
Department of Information and Computer Engineering
Politeknik Elektronika Negeri Surabaya



News Dataset



label.csv

data1.txt	ekonomi
...	...
data10.txt	ekonomi
data11.txt	sepakbola
...	...
data20.txt	sepakbola
data21.txt	automotif
...	...
data30.txt	automotif
data31.txt	ponsel
...	...
data40.txt	ponsel
data41.txt	pariwisata
...	...
data50.txt	pariwisata

Assignment

Lakukan dari data1.txt-sampai data50.txt (i=1-50)

1. data \leftarrow bacalah file txt untuk data ke-i
 2. keywords \leftarrow lakukan Tokenizing, Filtering dan Stemming/Tagging pada data
 3. scores \leftarrow lakukan penghitungan TF dan hilangkan keyword yang mempunyai score TF dibawah 50% dari score tertinggi TF, dan tampilkan
-
4. querylist \leftarrow masukkan query “pertumbuhan ekonomi, perkembangan pasar dan pergerakan harga saham” (berkategori “ekonomi”), dan lakukan Tokenizing, Filtering dan Stemming/Tagging pada query
 5. rankdocs \leftarrow lakukan pencarian dari querylist terhadap scores masing-masing data, dan ambil 10 data yang mempunyai total scores tertinggi
 6. label \leftarrow baca file label.csv
 7. recall, precision \leftarrow hitunglah recall dan precision dari rankdocs dimulai dari ranking 1 sampai ranking 10, dengan membandingkan kategori rankdocs pada label terhadap kategori query
 8. Visualisasikan grafik recall dan precision

Pengumpulan Tugas

- Buatlah coding dengan Bahasa pemrograman/tools apapun untuk semua assignment
- Buatlah laporan dalam slide ppt. Laporan terdiri dari screenshot coding dan hasil running untuk setiap assignment.
- Simpan laporan dalam file pdf dengan format penamaan:
DM_M13_NRP_namadepan.pdf
- Upload file tersebut ke ETHOL
- Deadline upload: Kamis, 14 November 2024 pk. 18.00

