

Daftar Istilah

- Cosine Similarity*** : Salah satu tahap perhitungan dalam sistem dalam membandingkan vektor.
- Database*** : Data acuan program
- Document Clustering*** : Sistem pengelompokan dokumen.
- ESA*** : *Explicit Semantic Analysis*, yaitu salah satu metode yang digunakan dalam perhitungan nilai keterkaitan antar teks.
- Filtering*** : Penyaringan unsur-unsur yang tidak diperlukan dalam teks.
- Frekuensi Biner*** : Salah satu perkembangan metode *ESA* dalam perhitungan *tf.idf*.
- Frekuensi NonBiner*** : Salah satu perkembangan metode *ESA* dalam perhitungan *tf.idf*.
- Gold Standard*** : Skor yang dihasilkan dari sekumpulan pendapat manusia.
- Korelasi*** : Salah satu standar keakuratan suatu sistem.
- Lowercase*** : Huruf kecil dari suatu teks.
- MEN*** : Salah satu *dataset gold standard* berbahasa inggris.
- Pearson Correlation*** : Salah satu perhitungan dalam menentukan nilai korelasi .
- Preprocessing*** : Tahapan awal dalam mengolah data input.
- Question Answering*** : Sistem penjawab pertanyaan.
- Scoring*** : Proses pembuatan nilai .
- Semantic Similarity*** : Kesamaan semantik antar teks.
- Semantic Relatedness*** : Keterkaitan semantik antar teks.
- Semantik*** : Ilmu yang berkaitan dengan makna antar teks.
- Stemming*** : Perubahan struktur kata menjadi kata dasarnya.
- Stopword*** : Kata-kata umum yang tidak mengandung arti spesifik.
- Term*** : Sebuah kata.
- Tf.idf*** : Salah satu tahapan dalam metode *ESA*.

- Tokenization*** : Tahapan awal *preprocess* yang ditujukan memecah suatu teks menjadi kumpulan kata.
- Uppercase*** : Huruf besar dari suatu teks.
- Wikipedia*** : Dataset yang digunakan sebagai *database* sistem
- Word*** : Kata
- Wordsim353*** : Salah satu *dataset gold standard* berbahasa inggris.