

ABSTRAKSI

Vector Space Model dan *Binary Independence Model* adalah 2 metode yang dipakai untuk memodelkan hasil Pemerolehan Informasi. Pada metode *Vector Space Model*, setiap potongan kata (token) dalam dokumen dilambangkan sebagai vektor dan dihitung jarak kedekatannya dengan kata pencarian menggunakan *Cosine Similarity*. Metode *Binary Independence Model* menghitung peluang munculnya kata pencarian pada dokumen dengan menggunakan prinsip peluang Naïve Bayes.

Pada tugas akhir ini dibuat mesin pencari untuk membandingkan kedua metode tersebut. Perbandingan dilakukan dalam hal unjuk kerja dan lama waktu pencarian. Unjuk kerja diukur dari nilai precision dan recall untuk masing-masing metode. Lama waktu pencarian diukur dengan menghitung waktu setiap metode mulai dilakukan sampai metode selesai dilakukan. Koleksi yang digunakan berupa 72 buah ebook dan 34 buah Tugas Akhir dengan 2 bahasa, yaitu bahasa Inggris dan bahasa Indonesia.

Hasil percobaan menunjukkan bahwa rata-rata lama waktu pencarian untuk metode *Vector Space Model* lebih cepat dibandingkan dengan lama waktu pencarian metode *Binary Independence Model*. Akan tetapi hasil perhitungan *average precision* pada 11 titik untuk metode *Binary Independence Model* lebih baik dibandingkan dengan metode *Vector Space Model* dalam menangani pencarian pada koleksi Perpustakaan Universitas Sanata Dharma.

ABSTRACT

Vector Space Model and Binary Independence Model are two methods that been used to modelize retrieval result in Information Retrieval. In Vector Space Model, every word in a document represented as vector, and the similarities compare to search keyword's vector measured using Cosine Similarity. Binary Independence Model count the search keyword probabilistic shown up in a document using Naïve Bayes's principal.

In this thesis, a search engine were built to compare thus two methods. Comparison will be made in terms of performance and searching time. The performance will be measured by the value of precision and recall for each method. The searching time will be measured as the method starts to search until it has done the searching process. The collection that used in this thesis were 72 ebooks and 34 thesises within 2 languages, Bahasa Indonesia and English.

The result says that the average searching time for Vector Space Model is faster than Binary Independence Model's average searching time. However, result for measuring average of the precision within 11-point for Binary Independence Model is better than Vector Space Model's handling in the Sanata Dharma University Library collection search.