

ABSTRAKSI

Setiap tahun makalah ilmiah berbahasa Indonesia semakin bertambah banyak sehingga diperlukan sistem pemerolehan informasi untuk mencari dokumen yang relevan. Sebagian besar sistem pemerolehan informasi dan web pencarian menerapkan *inverted index* yang terbukti efisien dalam menjawab *query*. Pada penelitian sebelumnya, penggunaan *inverted index* pada *database* menawarkan beberapa keuntungan dan implementasi *inverted index* dengan Bahasa Indonesia menggunakan ORDBMS memiliki waktu akses yang lebih baik dibandingkan menggunakan RDBMS.

Dari latar belakang tersebut, diimplementasikan sebuah pencarian makalah ilmiah berbahasa Indonesia berdasarkan seluruh isi teks dalam dokumen menggunakan struktur data *inverted index* pada ORDBMS dengan metode pembobotan TF-IDF (Savoy) dan *stemming* menggunakan algoritma Nazief & Adriani. Sistem ini dikembangkan menggunakan bahasa pemrograman Java, ORDBMS pada Oracle 11g R2 Enterprise Edition dan Netbeans.

Percobaan ini menggunakan *corpus* 281 dokumen makalah ilmiah berbahasa Indonesia yang berisikan 25737 term. Waktu yang dibutuhkan untuk proses pencarian pada operasi AND dengan 1 sampai 2 kata kunci kurang dari 0,1 detik, 3 kata kunci kurang dari 0,4 detik, dan 4 kata kunci kurang dari 1 detik. Sedangkan, pada hasil perhitungan *recall precision* pada 11 titik sistem dapat menghasilkan nilai yang baik meskipun terdapat beberapa kelemahan yang ditemukan.

Kata kunci: Pemerolehan informasi, *inverted index*, ORDBMS, TF-IDF

ABSTRACT

Every year, Indonesian scientific papers are multiplying so it required information retrieval system to find relevant documents. Most of the information retrieval system and web search implement inverted index which proven efficient in answering queries. In previous research, the use of inverted index on database offers several advantages and implementation of inverted index by using Indonesian languages on ORDBMS had better access time than using RDBMS.

From these background, searching scientific paper with Indonesian language was implemented scientific paper searching, the entire contents of the text in a document using the inverted index data structure on ORDBMS with TF-IDF weighting method (Savoy) and stemming algorithm Nazief & Adriani. The system was developed using the Java programming language, ORDBMS Oracle 11g R2 Enterprise Edition and Netbeans.

This experiment uses corpus 281 Indonesian scientific papers documents that contains 25,737 terms. The time needed to process a search on the AND operation with 1 to 2 keywords less than 0.1 seconds, 3 keywords less than 0.4 seconds, and 4 keywords less than 1 second. Meanwhile, result on the 11-point recall precision computation system can produce a good value despite some weaknesses were found.

Keyword: Information retrieval, inverted index, ORDBMS, TF-IDF