

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Pada bagian ini akan ditarik beberapa kesimpulan setelah melakukan implementasi dan pengujian pada Bab 5. Berikut merupakan beberapa kesimpulan yang didapatkan setelah melakukan implementasi dan pengujian :

1. Ekspansi *query* yang dilakukan dengan menggunakan Matriks Asosiasi terbukti dapat meningkatkan kemampuan dari suatu aplikasi *Information Retrieval*. Hal ini dibuktikan pada bab 3 dan 5.2 bahwa nilai *recall* sesudah ekspansi pada setiap bagian pengujian menjadi bertambah.
2. Ekspansi *query* dengan Matriks Asosiasi bisa ditingkatkan efektivitasnya dengan penambahan *term* dari Thesaurus. Hal ini dibuktikan dalam eksperimen kedua pada bagian 5.3.2 bahwa salah satu hasil ekspansi adalah "asparagus luscious". Term "luscious" dari berasal dari kata "luscious" yang dihasilkan oleh Thesaurus.
3. Kombinasi *Boolean Model* dan *VSM* dapat membuat kelemahan dari kedua model tersebut bisa diatasi. Hasilnya aplikasi dapat menjawab *query* dengan *boolean* operator serta pengurutan peringkat menjadi bisa dilakukan. Hal ini dibuktikan pada bagian 3.11, 5.2 dan 5.3.
4. Peran NLP dalam *Information Retrieval* sangat penting terutama untuk melakukan normalisasi *term*. Hal ini dibuktikan dengan melihat perbedaan nilai *cosine similarity* pada bagian 3.10 dan 5.2.1.5. Jika *term* yang diindeks tidak dilakukan normalisasi maka akan mempengaruhi nilai *euclidean length vector*.
5. Pengurutan menggunakan *boolean* dan *cosine similarity* pada perangkat lunak ini memanfaatkan kelas dari Lucene. Kelas dari Lucene tersebut memakai list *stop word* dalam keadaan standar. Pengurutan dengan *cosine similarity* yang telah dibuat dalam perangkat lunak ini menggunakan *stop word* yang telah dilakukan penambahan *term*, sehingga daftar *stop word* yang digunakan dalam pengurutan menggunakan kombinasi *boolean* dan *cosine similarity* berbeda dengan pengurutan menggunakan *similarity* saja. Oleh karena itu nilai *cosine similarity* dari kedua tipe pengurutan peringkat yang telah dibuat menjadi berbeda. Hal ini dapat dilihat pada eksperimen bagian 5.3.1.1.
6. Saat melakukan perhitungan ekspansi *query* dengan Matriks Asosiasi, jika pasangan $term_i$ dan $term_j$ merupakan kandidat ekspansi, sementara $term_i$ sering muncul dalam setiap dokumen

dengan jumlah yang banyak tanpa kehadiran $term_j$ maka nilai *normalized association score* akan semakin kecil dan menyebabkan kesempatan untuk terpilih sebagai hasil ekspansi *query* akan semakin kecil. Hal ini berlaku untuk sebaliknya.

6.2 Saran

Pada bagian ini memuat saran terhadap perangkat lunak yang dibuat pada skripsi ini. Saran diharapkan dapat membuat perangkat lunak ini lebih baik lagi saat diimplementasikan. Berikut merupakan beberapa saran yang didapatkan setelah melakukan implementasi dan pengujian :

1. Diperlukan dokumen yang sangat banyak agar *corpus* lebih terisi dengan topik yang sangat beragam. Jika itu bisa dilakukan, Hasil ekspansi *query* akan lebih baik.
2. Diperlukan penggunaan *Database* agar proses berjalan lebih cepat. Semua vektor bobot dokumen disimpan dalam *Database* sehingga yang dihitung hanya vektor *query* saja. Selain dapat mempercepat proses, hal ini juga akan dapat mengurangi memori yang dipakai.
3. Untuk melakukan eksperimen yang lebih baik diperlukan dokumen yang banyak dan berisi topik yang beragam. Oleh karena itu disarankan menggunakan dokumen yang khusus yang telah disediakan untuk *benchmark*.

DAFTAR REFERENSI

- [1] Manning, Prabhakar, dan hinrich, S. (2009) *Introduction to Information Retrieval*. Cambridge University, England.
- [2] Chartree, J., Cankaya, E. C., dan Phithakkitnukoon, S. (2010) Query expansion using association matrix for improved information retrieval performance. Skripsi. University of North Texas Denton, Texas.
- [3] Essam, Z. U., Department of Computer Information System (2013) Similar thesaurus based on arabic document. *Computer Science*, **3**.
- [4] Cambria, B., Erik; White (2014) Jumping nlp curves: A review of natural language processing research. *Curve*, **9**.
- [5] Vaibhav Kant Singh, V. K. S. (2015) Vector space model: An information retrieval system. *Advanced Engineering Research and Studies*, **4**.
- [6] Hazra Imran, A. S. (2009) Thesaurus and query expansion. *Computer Science*, **1**.
- [7] Aliwy, A. H. (2012) Tokenization as preprocessing for arabic tagging system. *International Journal of Information and Education Technology*, **2**.
- [8] Giridhar N S, N. V. S. R., Prema K.V (2011) A prospective study of stemming algorithms for web text mining. *GANPAT UNIVERSITY JOURNAL OF ENGINEERING AND TECHNOLOGY*, **1**.
- [9] Karaa, W. B. A. (2013) A new stemmer to improve information retrieval. *International Journal of Network Security*, **5**.
- [10] Michael McCandless, O. G., Erik Hatcher (2004) *Lucene In Action*, 2 edition. Manning.