

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Dari penelitian yang telah dilakukan, didapatkan beberapa buah kesimpulan yaitu:

1. Secara performa FFT bekerja lebih baik ketimbang VSM ketika mengolah *query* yang mengandung *stopword* yang jarang muncul. Hal ini dapat dilihat dari rata-rata waktu pencarian serta performa AVP dan DCG. Ketika *query* mengandung *stopword* yang jarang muncul, jelas perhitungan tidak begitu bagus karena bobot dari *stopword* yang jarang muncul akan tinggi dan dianggap seperti kata biasa.
2. Secara kecepatan VSM bekerja lebih cepat ketimbang FFT ketika *query* yang diolah pendek atau jumlah dokumen yang relevan pada koleksi sedikit. Pada VSM perhitungan dilakukan secara langsung sedangkan pada FFT perlu beberapa langkah matematis sehingga untuk *query* pendek akan sedikit lebih lama dari yang seharusnya. Namun untuk *query* yang agak panjang FFT bekerja sedikit lebih baik dari VSM.
3. Hasil yang dikeluarkan FFT dan VSM selalu ideal ketika nilai *stopword* sudah tertekan oleh pembobotan dari TF-IDF. Secara umum kemunculan *stopword* cukup tinggi. Oleh karena itu nilai dari *stopword* dapat ditekan oleh TF-IDF sehingga tidak begitu mempengaruhi perhitungan. Hal ini dapat terlihat pada percobaan *query* ketiga dan keempat. Kedua *query* mengandung *stopword* namun performa pada *query* ketiga jauh lebih buruk ketimbang *query* keempat. Jika dilihat pada korpus, *stopword* "only" terkandung pada beberapa dokumen saja sehingga bobot TF-IDF cukup besar. Oleh karena itu perhitungan akan kacau karena bobot *stopword* "only" yang seharusnya tertekan menjadi tidak tertekan dan mengacaukan perhitungan.
4. Untuk *query* yang tidak mengandung *stopword*, hasil pencarian dari FFT dan VSM sudah sama-sama ideal. Salah satu faktor yang menyulitkan perhitungan dari penelitian ini adalah *stopword*. Penelitian ini tidak menggunakan *stemmer* sehingga *stopword* ikut terproses dalam pencarian. Ketika kemunculan *stopword* sedikit, nilai TF-IDF dari *stopword* itu akan cukup besar seperti kata yang bukan *stopword*. Lalu ketika diujicobakan ke dalam *query*, *stopword* itu mengacaukan perhitungan peringkat karena *stopword* memiliki bobot yang tinggi.

6.2 Saran

- Pada penelitian ini, program berhasil dibuat dan bekerja dengan cukup baik. Namun dalam pengujian performa mengalami kendala pada *stopword* dan kata yang berimbuhan. Secara normal, *stemmer* digunakan untuk menghilangkan *stopword* serta membuat kata yang berimbuhan menjadi kata dasar. Asumsi pada penelitian ini adalah secara umum *stopword* memiliki kemunculan yang tinggi sehingga nilai TF-IDF dari *stopword* tertekan dan tidak mempengaruhi pencarian. Akan tetapi pada penelitian ini pada koleksi dokumen memiliki beberapa *stopword* dengan tingkat kemunculan yang rendah.
- Saran untuk penelitian ini adalah lebih menspesifikasikan korpus atau tipe koleksi yang akan diujikan. Lalu dari jenis korpus yang dipilih komposisi dokumen juga harus diperhitungkan

dan dianalisis. Dari hal itu baru ditentukan apakah menggunakan *stemmer* atau tidak. Selain itu performa dari suatu mesin pencarian agak terlalu luas. Perlu juga spesifikasi performa yang akan diujikan apa. Dalam menguji suatu mesin pencari akan sulit menentukan mana yang lebih baik ketika algoritma yang diujikan sama-sama baik. Pada buku ChenXiang *Text Data Management and Analysis* [2], disebutkan performa mesin pencari satu dengan yang lain berbeda. Ada *query* yang bekerja baik pada suatu mesin pencari namun belum tentu baik pada mesin pencari lain. Banyak sekali faktor yang mempengaruhi performa mesin pencari seperti korpus atau koleksi dokumen, *query* yang digunakan, algoritma yang dipilih, bahasa yang digunakan, *indexing*, tujuan dari pencarian. Tujuan pencarian juga mempengaruhi dalam pembuatan *golden answer* yang mau tidak mau harus manusia yang membuat *golden answer*. Karena topik yang begitu luas, perlunya spesifikasi dari faktor-faktor itu dan tidak mungkin semua dibahas dalam satu penelitian.

DAFTAR REFERENSI

- [1] Yates, R. B. dan Neto, B. R. (2011) *Modern Information Retrieval*, 2nd edition. Pearson, United Kingdom.
- [2] Zhai, C. dan Samsung, S. (2016) *Text Data Management and Analysis*, 1st edition. ACM Books, New York.
- [3] Costa, A. dan Melucci, M. (2010) An information retrieval model based on discrete fourier transform. *Proceedings of the 1st Information Retrieval Facility Conference, IRFC 2010*, Vienna, Australia, May, pp. 84–99. ACM, New York.
- [4] Kreyszig, E., Kreyszig, H., dan Norminton, E. J. (2011) *Advance Engineering Mathematics*, 10th edition. Laurie Rosatone, USA.
- [5] Briandais, R. D. L. (1959) File searching using variable length keys. *Proceedings of Western J. Computer Conference*, New York, USA, March, pp. 295–298. ACM, New York.