

BAB 6

KESIMPULAN DAN SARAN

Pada bab ini akan dibahas mengenai kesimpulan akhir dan saran yang dapat diambil dari penelitian yang dilakukan.

6.1 Kesimpulan

Dari rangkaian penelitian yang telah dilaksanakan, dapat disimpulkan beberapa hal penting, antara lain:

1. Untuk menentukan kriteria pencarian sehingga proses *crawling* menjadi lebih fokus yaitu dengan menghitung nilai relevansi menggunakan *Vector Space Model*. Nilai relevansi pada *Vector Space Model* dapat ditingkatkan dengan menggunakan algoritma *Stemming*. Algoritma *Stemming* yang diterapkan belum sempurna, beberapa kata tidak dapat di-*stemming* dengan benar.
2. Berdasarkan hasil analisis yang dilakukan, penyimpanan informasi hasil *crawling* yang telah dikelompokkan berdasarkan kriteria disimpan pada basis data HBase yang berjalan diatas HDFS (*Hadoop Distributed File System*). Setiap *row key* dari *url* yang disimpan pada HBase memiliki nilai *row key* yang unik yaitu kombinasi antara topik pencarian, nilai relevansi, dan *url* yang bersangkutan.
3. Berdasarkan hasil pengujian eksperimen pada lingkungan HDFS (*Hadoop Distributed File System*), semakin banyak komputer *slave* dan agen *crawler* yang aktif, proses *crawling* semakin cepat sehingga menghasilkan *url* dan halaman web yang lebih banyak dalam suatu waktu tertentu.
4. Berdasarkan hasil pengujian eksperimen pada lingkungan HDFS (*Hadoop Distributed File System*), semakin banyak komputer *slave* yang aktif, waktu yang diperlukan untuk melakukan pencarian data semakin kecil.

6.2 Saran

Setelah melakukan proses analisis, perancangan, implementasi, dan pengujian pada penelitian ini maka penulis memberikan beberapa saran, yaitu:

1. Melakukan pencarian nilai relevansi pada saat proses pencarian, sehingga nilai *idf* pada *Vector Space Model* dapat dicari.
2. Melakukan perbaikan untuk aturan-aturan *stemming* dalam bahasa Indonesia, dikarenakan masih ada kondisi kata dalam bahasa Indonesia yang masih belum dapat di-*stemming* dengan aturan tersebut.
3. Ketika menjalankan perangkat lunak, dapat menggunakan koneksi internet yang tinggi agar *url* yang di-*crawl* lebih cepat.

4. Berdasarkan hasil penelitian dan pengujian yang telah dilakukan, ketika menjalankan perangkat lunak dengan basis data HBase yang berjalan diatas Hadoop, dapat menggunakan banyak komputer dan spesifikasi komputer yang tinggi agar berjalan dengan baik dan lancar.

DAFTAR REFERENSI

- [1] Udupure, T. V., Kale, R. D., dan Dharmik, R. C. (2014) Study of web crawler and its different types. *Computer Engineering*, **16**, 1–5.
- [2] Lublinsky, B., Smith, K. T., dan Yakubovich, A. (2013) *Professional Hadoop Solutions*. John Wiley and Sons Inc, Indianapolis, USA.
- [3] Olston, C. dan Najork, M. (2010) Web crawling. *Computer Engineering*, **4**, 175–246.
- [4] Manning, C. D., Raghavan, P., dan Schütze, H. (2009) *An Introduction to Information Retrieval*. Cambridge University Press, Cambridge, UK.
- [5] Elyasir, A. M. H. dan Anbananthen, K. S. M. (2012) Focused web crawler. *Information and Knowledge Management*, **45**, 149–153.
- [6] Batsakis, S., Petrakis, E., dan Milios, E. (2009) Improving the performance of focused web crawlers. *Data Knowledge Engineering*, **68**, 1001–1013.
- [7] Cormen, T. H., Leiserson, C. E., Rivest, R. L., dan Stein, C. (2009) *Introduction To Algorithms*, 3rd edition. Massachusetts Institute of Technology, Massachusetts, USA.
- [8] Indriyono, B. V., Utami, E., dan Sunyoto, A. (2015) Pemanfaatan algoritma porter stemmer untuk bahasa indonesia dalam proses klasifikasi jenis buku. *Jurnal Buana Informatika*, **4**, 301–310.
- [9] Sarkar, D. D., jaiswal, A., dan Saxena, A. (2015) Understanding architecture and framework of j2ee using web application. *Journal of Computer Science and Information Technologies*, **6**, 1253–1257.