

BAB 6

KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dan saran berdasarkan hasil analisis, perancangan, implementasi, pengujian, dan eksperimen yang telah dilakukan.

6.1 Kesimpulan

Berdasarkan hasil analisis, perancangan, implementasi, dan pengujian yang telah dilakukan pada penelitian ini, maka dapat diambil beberapa kesimpulan, antara lain:

1. *Web crawler* Apache Nutch dapat dipekerjakan menggunakan Nutch REST API yang dapat diakses melalui aplikasi Java yang berinteraksi dengan Nutch REST API.
2. Perangkat lunak berbasis Java REST API Client yang diberi nama agen *crawler* dibangun yang mempekerjakan *web crawler* Apache Nutch melalui Nutch REST API sesuai dengan siklus *crawling web crawler* Apache Nutch.
3. *Web crawler* Apache Nutch mengimplementasikan metode *crawling* yang terdiri dari 5 tahapan, yaitu *inject seed* URL ke dalam HBase, *generate*, *fetch*, *parse*, dan *updatedb*.
4. Berdasarkan pengujian fungsional, situs induk J2EE sudah berjalan dengan sesuai dan semestinya.
5. Berdasarkan hasil perbandingan eksperimen performa antar *web crawler*, pada saat 3 *region server*, *focused web crawler* lebih unggul dalam memproses banyaknya URL, sedangkan pada saat *region server* ditambah menjadi 5, banyaknya baris URL yang dapat diekstraksi oleh *web crawler* Apache Nutch akan mengungguli ketiga *web crawler* lainnya (dihitung menggunakan rumus *growth*).

6.2 Saran

Berdasarkan hasil analisis, perancangan, implementasi, dan pengujian yang telah dilakukan pada penelitian ini, maka penulis memberikan beberapa saran, antara lain:

1. Pada saat pengujian, penulis menghadapi permasalahan sumber daya komputer yang spesifikasinya tidak cukup baik sehingga sedikit menghambat jalannya program. Penulis menyarankan untuk menggunakan komputer dengan spesifikasi yang cukup tinggi ketika melakukan pengembangan pada lingkungan terdistribusi Hadoop, termasuk pengembangan menggunakan HBase.
2. Koneksi internet yang tidak stabil mempengaruhi kecepatan pengolahan baris jumlah URL yang diproses pada saat melakukan *crawling* sehingga Perlu diperhatikan adanya koneksi internet yang cepat dan stabil.

DAFTAR REFERENSI

- [1] Udupure, T. V., Kale, R. D., dan Dharmik, R. C. (2014) Study of web crawler and its different types. *IOSR Journal of Computer Engineering (IOSR-JCE)*, **16**, 1–5.
- [2] Abianti, M. N. (2018) Perangkat lunak incremental web crawler pada lingkungan hadoop. Skripsi. Universitas Katolik Parahyangan, Indonesia.
- [3] Maradenia, J. H. (2018) Focused web crawling pada lingkungan hadoop. Skripsi. Universitas Katolik Parahyangan, Indonesia.
- [4] Gabriella (2018) Web crawling terdistribusi pada lingkungan hadoop. Skripsi. Universitas Katolik Parahyangan, Indonesia.
- [5] Laliwala, D. Z. dan Shaikh, A. (2013) *Web Crawling and Data Mining with Apache Nutch*. Packt Publishing Ltd, Livery Place, 35 Livery Street, Birmingham B3 2PB, UK.
- [6] Metaxas, P. T. (2012) Why is the shape of the web a bowtie? *World Wide Web (WWW) Conference*, France, April. WebScience Track, Lyon.
- [7] Donato, D., Laura, L., Leonardi, S., dan Millozzi, S. (2004) Large scale properties of the webgraph. *The European Physical Journal*, **38**, 239–243.
- [8] Manning, C. D., Raghavan, P., dan SchÄijtzte, H. (2009) *An Introduction to Information Retrieval*. Cambridge University Press, Cambridge, UK.
- [9] Meusel, R., Vigna, S., dan Lehmborg, O. (2014) Graph structure in the web-revisited, . Seoul, Korea, 7-11 April. ACM, New York.
- [10] Zheng, S., Dmitriev, P., dan Giles, C. L. (2009) Graph based crawler seed selection. *Conference: Proceedings of the 18th International Conference on World Wide Web*, Madrid, Spain, 20-24 April, pp. 1089–1090. ACM, New York.
- [11] Yang, C. dan Liao, H. J. (2010) Using the robots.txt and robots meta tags to implement online copyright and a related amendment. *Library Hi Tech*, **28**, 94–106.
- [12] Lublinsky, B., Smith, K. T., dan Yakubovich, A. (2013) *Professional Hadoop® Solutions*. John Wiley and Sons, Inc, 10475 Crosspoint Boulevard Indianapolis, IN 46256.
- [13] Rani, C. S. dan Rama, D. B. (2017) Mapreduce with hadoop for simplified analysis of big data. *International Journal of Advanced Research in Computer Science*, **8**, 853–856.
- [14] Sharma, V. dan Dave, M. (2012) Sql and nosql databases. *International Journal of Advanced Research in Computer Science and Software Engineering*, **2**, 20–27.
- [15] Moniruzzaman, A. B. M. dan Hossain, S. A. (2013) Nosql database: New era of databases fot big data analytics-classification, characteristics, and comparison. *International Journal of Database Technology and Application*, **6**, 1–13.

- [16] George, L. (2011) *HBase: The Definitive Guide*, 1 edition. O'Reilly Media, Inc, 1005 Gravenstein Highway North, Sebastopol, CA 95472.
- [17] Cesare Pautasso, O. Z. dan Leymann, F. (2008) Restful web services vs. "big" web services: Making the right architectural decision. *Web Engineering - Web Service Deployment* , ?, 805–814.
- [18] Kahate, A. dan Godbole, A. (2013) *Web Technologies: TCP/IP, Web/Java Programming and Cloud Computing*, 3 edition. Tata McGraw-Hill Education Pvt. Ltd, New York, NY Columbus.