Abstract

Research in the field of Natural Language Processing (NLP) is currently increasing especially with the arrival of a new term that is “big data”. The needs of the programming library that ready-to-use becomes very important to speed up the phases of research. Some libraries that have already been mature is available but generally for English language and its dependently. So, it can’t be used for other languages. Stemming is one of the basic processes that exist in NLP. Indonesian stemming algorithm that often used is ECS (Enhanced Confix-Stripping). One of the libraries that already implemented the algorithm is Sastrawi. Results from the experiment show that the time of stemming processing by Sastrawi is still slow. Therefore, this research will optimize the speed of stemming processing using multiprocessing (MP). The data test are used in this research has manually taken from Wikipedia. The experiment results show that the MP technique can decrease the average time of stemming processing about 98.45%.

References
1. Sugiyama H., Meguro T., and Higashinaka R., 2017, Evaluation of Question-answering System about Conversational Agents Personality, Dialogues with Social Robots, Springer, Singapore, 183
2. Yousefi-Azar M. and Hamey L., 2017, Text Summarization Using Unsupervised Deep Learning, Expert Systems with Applications, 93
3. Negri M., Ataman D., Sabet M. J., Turchi M., and Federico M., 2017, Automatic Translation Memory Cleaning, Machine Translation, 1
4. Abdiansah A. and Winarko E., 2015, Question Classification Menggunakan Support Vector Machines dan Stemming, Seminar Nasional Aplikasi Teknologi Informasi (SNATI), 34
5. Tala F. Z., 2003, A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia, thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands
6. Asian J., Williams H. E., and Tahaghoghi S. M., 2005, Stemming Indonesian, the Twenty-eighth Australasian conference on Computer Science, 307
7. Adriani M., Asian J., Nazief B., Tahaghoghi S. M., and Williams H. E., 2007, Stemming Indonesian: A Confi xStripping Approach, ACM Transactions on Asian Language Information Processing (TALIP), 1
8. Arifin A. Z., Mahendra I. P., and Ciptaningtyas H. T., 2009, Enhanced Confi x Stripping Stemmer and Ants Algorithm for Classifying News Document in Indonesian Language, International Conference on Information and Communication Technology and Systems (ICTS), 60
9. Tahitoe A. D. and Puriwasasari D., 2010, Implementasi Modifikasi Enhanced Confi x Stripping Stemmer Untuk Bahasa Indonesia Dengan Metode Corpus Based Stemming, thesis, Fakultas Teknologi Informasi, Institut Teknologi Surabaya, Surabaya
10. Haris M., Maqsood N., Haq U., Zaman T., and Zubair M., 2016, UNI Processor and Multi-processor Performance Comparison, International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), 72
11. Kertsz G., Kiss D., Lovrics A., Sznsi S., and Vmossy Z., 2016, Multiprocessing of an Individual-Cell Based Model for Parameter Testing, Applied Computational Intelligence and Informatics (SACI), IEEE, 491
12. Siddiqui I. F., Abbas A., Ariffin A. R., and Lee S. U., 2016, A Comparative Study of Multithreading APIs for Software of ICT Equipment, Indian Journal of Science and Technology, 9

Index Terms
Computer Science Information Sciences

Keywords
Multiprocessing, Stemming, ECS, Sastrawi
Multiprocessing Stemming: A Case Study of Indonesian Stemming