

{tag}

{/tag}International Journal of Computer Applications

© 2012 by IJCA Journal

Volume 46 - Number 8

Year of Publication: 2012

Authors:

Tahar Dilekh

Ali Behloul

10.5120/6927-9344

{bibtex}pxc3879344.bib{/bibtex}

Abstract

In this paper, we propose a hybrid method that combines the application of three previously used techniques. These techniques deal with three key issues related to Arabic stemming including affix removal proposed by Kadri [1], dictionaries [2] and morphological analysis [3] [4] [5]. Thus, when solving these problems these techniques are applied individually and independently to solve associated stemming problems, which requires some adjustments to be implemented on each one of them. Therefore, the main contribution of this experiment is to demonstrate the effectiveness of the hybrid method compared to other methods, and the choice of removing the suffix before prefix during the operation of Arabic stemming process.

References

- Kadri, Y. & Nie, J. (2006). "Effective Stemming for Arabic Information Retrieval" in proceedings of the Challenge of Arabic for NLP/ MT Conference, Londres, Royaume-Uni.
- Al-Kharashi, I. and Evens, M. W. Comparing words, stems, and roots as index terms in an Arabic information retrieval system. JASIS, 45 (8), pp. 548-560, 1994.
- Kenneth R. Beesley. 1998. Arabic Morphological Analysis on the Internet. To appear in the Proceedings of the International Conference and Exhibition on Multi-lingual Computing

(Arabic and English), ICEMCO-98.

- Attia, Mohamed, A. : 2000 A large-scale computational processor of the Arabic morphology, A Master's Thesis, Cairo University, (Egypt) (2000).
- Mohamadi, T. S. Mokhnache: 2002, Design and development of Arabic speech synthesis, WSEAS 2002, Greece, Sept. 25-28, (2002).
- <http://www.internetworldstats.com/stats.htm>.
- Khoja S. and Garside S. (1999). 'Stemming Arabic Text'. Computing Department, Lancaster University, Lancaster, U. K.
- Larkey L. S. and Connell M. E. (2001). 'Arabic information retrieval at UMass in TREC-10'. TREC-10 conference, Gaithersburg, Maryland 2001.
- Darwish, K. and Oard, D. W. CLIR Experiments at Maryland for TREC-2002: Evidence combination for Arabic-English retrieval. In TREC 2002. Gaithersburg: NIST, pp 703-710, 2002.
- Chen, A. , and Gey, F. Building an Arabic stemmer for information retrieval. In TREC 2002. Gaithersburg: NIST, pp 631-639, 2002.
- Wightwick, J. and Gaafar, M. Arabic verbs and essentials of grammar. Chicago: Passport Books, 1998.
- Larkey L. S, L. Ballesteros, and M. E. Connell, 'Improving stemming for Arabic information retrieval: light stemming and co-occurrence analysis,' Tampere, Finland: ACM, 2002, pp. 275-282.
- P. Schauble, Multimedia Information Retrieval: content-based Information Retrieval from Large Text and Audio Databases, Kluwer Academic Publishers, 1997
- Pirkola, A. Morphological typology of languages for IR. Journal of Documentation, 57 (3), pp. 330-348, 2001.
- Popovic, M. and Willett, P. The effectiveness of stemming For natural-language access to Slovene textual data. JASIS, 43 (5), pp. 384-390, 1992.
- Ntais, G. Development of a stemmer for the greek language. Master's thesis, Stockholm University, 2006.
- Sankupellay, M. 'Malay-Language Stemmer,' Sunway Academic Journal, vol. 3, pp. 147–153, 2006.
- Al-Sughaiyer, I. A. and Al-Kharashi, I. A. (2004) 'Arabic morphological analysis techniques: A comprehensive survey', Journal of the American Society for Information Science and Technology, 55(3):189–213.

Index Terms

Computer Science

Artificial Intelligence

Keywords

Information Retrieval Indexation Tokenization stemming Arabic Language