Finding semantic similarity in Vietnamese

Nguyen D.T., Pham S.B.
Faculty of Information Technology, Information Technology Institute, Vietnam National University, Hanoi, Viet Nam

Abstract: Finding semantic similarity is an important task in many natural language processing applications. Despite numerous works for popular languages, there is still limited research done for Vietnamese. In this paper, we tackle the problem of finding semantic similarity for Vietnamese using Random Indexing and Hyperspace Analogue to Language to represent the semantics of words and documents. We build a system to find synonyms in Vietnamese. Experimental results show that our system achieves accuracies of 75% for finding synonyms for verbs and 65% for synonyms for nouns. © 2010 IEEE.

Author Keywords: Hyperspace Analogue to Language; Random projection; Semantic vector; Word space model

Index Keywords: Hyperspace analogue to languages; NAtrual language processing; Random indexing; Random projection; Semantic similarity; Word space model; Computational linguistics; Semantics; Vector spaces; Natural language processing systems

Year: 2010
Source title: Proceedings - 2010 International Conference on Asian Language Processing, IALP 2010
Art. No.: 5681551
Page : 91-94

Link: Scopus Link
Correspondence Address: Nguyen, D. T.; Faculty of Information Technology, Information Technology Institute, Vietnam National University, Hanoi, Viet Nam; email: datnt88@gmail.com

Conference name: 2010 International Conference on Asian Language Processing, IALP 2010
Conference date: 28 December 2010 through 30 December 2010
Conference location: Harbin
Conference code: 83700
DOI: 10.1109/IALP.2010.78

Language of Original Document: English

Abbreviated Source Title: Proceedings - 2010 International Conference on Asian Language Processing, IALP 2010
Document Type: Conference Paper
Source: Scopus

Authors with affiliations:
• Nguyen, D.T., Faculty of Information Technology, Information Technology Institute, Vietnam National University, Hanoi, Viet
References:

- Banerjee, S., Pedersen, T., Extended gloss overlaps as a measure of semantic relatedness. in (2003) IJCAI, pp. 805-810
- Phillips, W., Riloff, E., Exploiting strong syntactic heuristics and co-training to learn semantic lexicons (2002) Proceedings of EMNLP'02
- Van Der Plas, L., Tiedemann, J., Finding synonyms using automatic word alignment and measures of distributional similarity (2006) Association for Computational Linguistics Morristown, , NJ, USA
- Van Der Plas, L., Bouma, G., Syntactic contexts for finding semantically similar words (2005) Proceedings of the Meeting of Computational Linguistics in the Netherlands (CLIN)
- Pham, D.D., Tran, G.B., Pham, S.B., A hybrid approach to Vietnamese word segmentation using part of speech tags (2009) International Conference on Knowledge and Systems Engineering