

Vietnamese noun phrase chunking based on conditional random fields

Thao N.T.H., Thai N.P., Minh N.L., Thuy H.Q.

College of Technology, Vietnam National University, Hanoi, Viet Nam; Japan Advanced Institute of Science and Technology, Japan

Abstract: Noun phrase chunking is an important and useful task in many natural language processing applications. It is studied well for English, however with Vietnamese it is still an open problem. This paper presents a Vietnamese Noun Phrase chunking approach based on Conditional random fields (CRFs) models. We also describe a method to build Vietnamese corpus from a set of hand annotated sentences. For evaluation, we perform several experiments using different feature settings. Outcome results on our corpus show a high performance with the average of recall and precision 82.72% and 82.62% respectively. © 2009 IEEE.

Index Keywords: Conditional random field; NAatural language processing; Noun phrase; Open problems; Recall and precision; Computational linguistics; Content based retrieval; Natural language processing systems; Systems engineering; Knowledge engineering

Year: 2009

Source title: KSE 2009 - The 1st International Conference on Knowledge and Systems Engineering

Art. No.: 5361712

Page : 172-178

Link: Scopus Link

Correspondence Address: Thao, N. T. H.; College of Technology, Vietnam National University, Hanoi, Viet Nam; email: thaonth@vnu.edu.vn

Sponsors: College of Technology; Vietnam National University

Conference name: 1st International Conference on Knowledge and Systems Engineering, KSE 2009

Conference date: 13 October 2009 through 17 October 2009

Conference location: Hanoi

Conference code: 79895

ISBN: 9.78E+12

DOI: 10.1109/KSE.2009.43

Language of Original Document: English

Abbreviated Source Title: KSE 2009 - The 1st International Conference on Knowledge and Systems Engineering

Document Type: Conference Paper

Source: Scopus

Authors with affiliations:

- Thao, N.T.H., College of Technology, Vietnam National University, Hanoi, Viet Nam
- Thai, N.P., College of Technology, Vietnam National University, Hanoi, Viet Nam
- Minh, N.L., Japan Advanced Institute of Science and Technology, Japan

- Thuy, H.Q., College of Technology, Vietnam National University, Hanoi, Viet Nam

References:

- Quang Ban, D., Bân, H., (2004) Vietnamese Grammar, , Education Publishing House, HaNoi
- Chí Hoà, N., (2004) Practical Vietnamese Grammar, , Hanoi National University Publishing House
- Steven, A., (1991) Parsing by Chunks, , In Berwick, Abney, and Tenny, editors, Principle-Based Parsing. Kluwer Academic Publishers
- McCallum, A., Freitag, Pereira, Maximum entropy markov models for information extraction and segmentation (2000) Proc. International Conference on Machine Learning
- McCallum, A., Efficiently inducing features of conditional random fields (2003) Conference on Uncertainty in Artificial Intelligence (UAI)
- Bharathi, A., Mannem, P.R., Introduction to the shallow parsing contest for South Asia languages (2007) Proceedings of the IJCAI-2007 Workshop on Shallow Parsing for South Asian Languages
- Brill, E., (1993) A Corpus-Based Approach to Language Learning, , PhD thesis, University of Pennsylvania
- Tjong, E.F., Sang, K., Buchholz, S., Introduction to the CoNLL-2000 shared task: Chunking (2000) Proceedings of CoNLL-2000 and LLL-2000, pp. 127-132. , Lisbon, Portugal
- Xu, F., Zong, C., A hybrid approach to Chinese base noun phrase chunking (2006) Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing, pp. 87-93. , Sydney, July
- Ramshaw, L.A., Marcus, M.P., Text chunking using transformation-based learning (1995) Proceedings of the ACL 3rd Workshop on Very Large Corpora, pp. 82-94
- Lafferty John, D., McCallum, A., Pereira, Fernando, C.N., Conditional random fields: Probabilistic models for segmenting and labeling sequence data (2001) Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), p. 282. , Morgan Kaufmann Publishers
- Thai, N.P., Luong, V.X., Huyen, N.T.M., Van Hiep, N., Phuong, L.H., Building a large syntactically-annotated corpus of vietnamese (2009) Proceedings of the 3rd Linguistic Annotation Workshop (LAW) at ACL-IJCNLP, , (to appear)
- Phan, X.-H., Nguyen, L.-M., Inoguchi, Y., Horiguchi, S., High-performance training of conditional random fields for large-scale applications of labeling sequence data (2007) IEICE Transactions on Information and Systems, E90-D (1), pp. 13-21. , DOI 10.1093/ietisy/e90-1.1.13
- Ando, R.K., Zhang, T., A high-performance semi-supervised learning method for text chunking (2005) Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 1-9. , June 25-30, Ann Arbor, Michigan
- Sha, F., Pereira, F., Shallow parsing with conditional random fields (2003) Technical Report MS-CIS-02-35, University of Pennsylvania
- Peng, F., McCallum, A., Information extraction from research papers using conditional random fields (2006) Information Processing and Management, 42 (4), pp. 963-979. , DOI 10.1016/j.ipm.2005.09.002, PII S0306457305001172
- Hunag, S., Xu, W., Chinese chunking based on maximum entropy markov models (2006) International Journal of Computational Linguistics and Chinese Language Processing, pp. 115-136
- Kudo, T., Matsumoto, Y., Use of support vector learning for chunk identification (2000) Proceedings of CoNLL-2000 and LLL-2000, pp. 142-144. , Lisbon, Portugal
- Kudo, T., Matsumoto, Y., Chunking with support vector machines Proceedings of the NAACL 2001, pp. 192-19
- Wang, T., Li, J., Diao, Q., Hu, W., Zhang, Y., Dulong, C., Semantic event detection using conditional random fields (2006) 2006 Conference on Computer Vision and Pattern Recognition Workshops, 2006, p. 1640552. , DOI 10.1109/CVPRW.2006.190, 2006 Conference on Computer Vision and Pattern Recognition Workshop

- Goldberg, Y., Adler, M., Elhadad, M., Noun phrase chunking in hebrew influence of lexical and morphological features (2006) Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL, pp. 689-696. , July 17-18, Sydney, Australia
- Lee, Y.-H., Kim, M.-Y., Lee, J.-H., Chunking using conditional random fields in korean texts Lecture Notes in Artificial Intelligence IJCNLP 2005
- Tan, Y., Yao, T., Chen, Q., Zhu, J., Applying conditional random fields to Chinese shallow parsing The 6th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2005)
- Chen, W., Zhang, Y., Isahara, H., An empirical study of Chinese chunking Proceedings of the 44th Annual Meeting of ACL, pp. 97-10