

# Named entity recognition for Vietnamese

Nguyen D.B., Hoang S.H., Pham S.B., Nguyen T.P.

Faculty of Information Technology, University of Engineering and Technology, Vietnam National University, Hanoi, Viet Nam

**Abstract:** Named Entity Recognition is an important task but is still relatively new for Vietnamese. It is partly due to the lack of a large annotated corpus. In this paper, we present a systematic approach in building a named entity annotated corpus while at the same time building rules to recognize Vietnamese named entities. The resulting open source system achieves an F-measure of 83%, which is better compared to existing Vietnamese NER systems. ?? 2010 Springer-Verlag Berlin Heidelberg.

**Index Keywords:** F-measure; In-buildings; Named entities; Named entity recognition; NER system; Open source system; Database systems; Open systems; Natural language processing systems

Year: 2010

Source title: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)

Volume: 5991 LNAI

Issue: PART 2

Page : 205-214

Link: [Scopus Link](#)

Correspondence Address: Nguyen, D. B.; Faculty of Information Technology, University of Engineering and Technology, Vietnam National University, Hanoi, Viet Nam; email: [datnb@vnu.edu.vn](mailto:datnb@vnu.edu.vn)

Sponsors: Hue University; Wroclaw University of Technology

Conference name: 2010 Asian Conference on Intelligent Information and Database Systems, ACIIDS 2010

Conference date: 24 March 2010 through 26 March 2010

Conference location: Hue City

Conference code: 80259

ISSN: 3029743

ISBN: 9.78E+12

DOI: 10.1007/978-3-642-12101-2\_22

Language of Original Document: English

Abbreviated Source Title: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)

Document Type: Conference Paper

Source: Scopus

Authors with affiliations:

1. Nguyen, D.B., Faculty of Information Technology, University of Engineering and Technology, Vietnam National University, Hanoi, Viet Nam
2. Hoang, S.H., Faculty of Information Technology, University of Engineering and Technology, Vietnam National University,

Hanoi, Viet Nam

3. Pham, S.B., Faculty of Information Technology, University of Engineering and Technology, Vietnam National University, Hanoi, Viet Nam
4. Nguyen, T.P., Faculty of Information Technology, University of Engineering and Technology, Vietnam National University, Hanoi, Viet Nam

#### References:

1. Bikel, D., Miller, S., Schwartz, R., Weischedel, R., A High- Performance learning namefinder (1998) Proceedings of the Fifth Conference on Applied Natural Language Processing, pp. 194-201
2. Borthwick, A., Sterling, J., Agichtein, E., Grishman, R., Exploiting diverse knowledge sources via maximum entropy in named entity recognition (1998) Proceedings of the Sixth Workshop on Very Large Corpora Montreal Canada
3. Budi, I., Bressan, S., Association rules mining for name entity recognition (2003) Proceedings of the Fourth International Conference on Web Information Systems Engineering
4. Nguyen, V.T.T., Cao, T.H., Automatic extraction of vietnamese named-entities on the web (2007) Proceedings of the Journal of New Generation Computing, , Ohmsha, Ltd. Springer
5. Fang, X., Sheng, H., A hybrid approach for chinese named entity recognition (2002) Proceedings of the Fifth International Conference on Discovery Science
6. Kim, J., Kang, I., Choi, K., Unsupervised named entity classification models and their ensembles (2002) Proceedings of the Nineteenth International Conference on Computational Linguistics
7. Mai, T.D., (2008) Named Entity Guideline for Vietnamese, , Bachelor thesis, College of Technology, Vietnam National University, Hanoi
8. Mansouri, A., Affendey, L., Mamat, A., Named entity recognition using a new fuzzy support vector machine (2008) Proceedings of the International Journal of Computer Science and Network Security, IJCSNS, 8 (2), pp. 320-325
9. Maynard, D., Tablan, V., Ursu, C., Cunningham, H., Wilks, Y., Named entity recognition from diverse text types (2001) Proceedings Recent Advances in Natural Language Processing 2001 Conference, , Tzigor Chark, Bulgaria
10. Maynard, D., Bontcheva, K., Cunningham, H., Towards a semantic extraction of named entities (2003) Proceedings Recent Advances in Natural Language Processing Borovets Bulgaria
11. Nguyen, T.C., Tran, O.T., Phan, H.X., Ha, T.Q., Named entity recognition in vietnamese free-text and web documents using conditional random fields (2005) Proceedings of the Eighth Conference on Some Selection Problems of Information Technology and Telecommunication, Hai Phong Viet Nam
12. Nguyen, T.P., Vu, L.X., Nguyen, H.M.T., Nguyen, H.V., Le, P.H., Building a large syntactically-annotated corpus of vietnamese (2009) Proceedings of the Third Linguistic Annotation Workshop (LAW) at ACL-IJCNLP 2009
13. Pastra, K., Maynard, D., Hamza, O., Cunningham, H., Wilks, Y., How feasible is the reuse of grammars for named entity recognition? (2002) Proceedings of the Conference on Language Resources and Evaluation, LREC 2002
14. Pham, T., Kawazoe, A., Dinh, D., Collier, N., Construction of vietnamese corpora for named entity recognition (2007) Conference RIAO 2007, , Pittsburgh PA, U.S.A, May 30- June 1
15. Pham, D.D., Tran, G.B., Pham, S.B., Vietnamese word segmentation using part of speech tags (2009) Proceedings of the First International Conference on Knowledge and Systems Engineering Hanoi Vietnam
16. Srihari, R., Niu, C., Li, W., A hybrid approach for named entity and sub-type tagging (2000) Proceedings of the Sixth Conference on Applied Natural Language Processing
17. Wu, Y.-C., Fan, T.-K., Lee, Y.-S., Yen, S.-J., Extracting named entities using support vector machines (2006) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 3886

LNBI, pp. 91-103. , Knowledge Discovery in Life Science Literature - PAKDD 2006 International Workshop, KDLL 2006, Proceedings