

{tag}

{/tag}

International Journal of Computer Applications
© 2011 by IJCA Journal

Volume 33 - Number 3

Year of Publication: 2011

Authors:

Vishal Gupta

Gurpreet Singh Lehal

10.5120/4001-5668

{bibtex}pxc3875668.bib{/bibtex}

Abstract

Named Entity Recognition (NER) is used to locate and classify atomic elements in text into predetermined classes such as the names of persons, organizations, locations, concepts etc. NER is used in many applications like text summarization, text classification, question answering and machine translation systems etc. For English a lot of work has already done in

field of NER, where capitalization is a major clue for rules, whereas Indian Languages do not have such feature. This makes the task difficult for Indian languages. This paper explains the Named Entity Recognition System for Punjabi language text summarization. A Condition based approach has been used for developing NER system for Punjabi language. Various rules have been developed like prefix rule, suffix rule, propername rule, middlename rule and lastname rule. For implementing NER, various resources in Punjabi, have been developed like a list of prefix names, a list of suffix names, a list of proper names, middle names and last names. The Precision, Recall and F-Score for condition based NER approach are 89.32%, 83.4% and 86.25% respectively.

Reference

- Hideki Isozaki. 2001 Japanese named entity recognition based on a simple rule generator and decision tree learning” in the proceedings of the Association for Computational Linguistics, (pp 306- 313). India.
- Takeuchi K. and Collier N. 2002 Use of Support Vector Machines in extended named entity Recognition. In the proceedings of the sixth Conference on Natural Language Learning (CoNLL-02), Taipei, Taiwan, China.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001 Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In the proceedings of International Conference on Machine Learning (pp 282-289). Williams College,
- R. Grishman 1995 The NYU system for MUC-6 or Where’s the Syntax. In the proceedings of Sixth Message Understanding Conference (MUC-6) (pp167-195). Fairfax, Virginia.
- Wakao T., Gaizauskas R. and Wilks Y. (1996). Evaluation of an algorithm for the Recognition and Classification of Proper Names. In the proceedings of COLING-96.
- Andrew Borthwick. 1999 Maximum Entropy Approach to Named Entity Recognition, doctoral dissertation, New York University.
- Collins, Michael and Y. Singer 1999 Unsupervised models for Named Entity Classification. In the proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora.
- 8 J. Kim, I. Kang, K. Choi. 2002 Unsupervised Named Entity Classification Models and their Ensembles. In the proceedings of the 19th International Conference on Computational Linguistics.
- Darvinder Kaur and Vishal Gupta. 2010 A survey of Named Entity Recognition in English and other Indian Languages. In Proceedings of (IJCSI) International Journal of Computer Science Issues Vol. 7 Issue 6 (pp 239-245).
- Alireza Mansouri, Lilly Suriani Affendey, Ali Mamat 2008 Named Entity Recognition Approaches. In Proceedings of IJCSNS International Journal of Computer Science and Network Security VOL.8 No.2 pp 339-344
- Srihari R., Niu C. and Li W. 2000 A Hybrid Approach for Named Entity and Sub-Type Tagging. In the proceedings of the sixth Conference on Applied Natural Language Processing.
- Amandeep Kaur, Gurpreet Singh Josan and Jagroop Kaur. 2009 Named

Entity Recognition for Punjabi: A Conditional Random Field Approach. In Proceedings of 7th international conference on Natural Language Processing/ICON-09. Macmillan Publishers, India.

- Praneeth M Shishtla, Karthik, Prasad Pingali and Vasudeva Verma
2008 Experiments in Telugu NER: A Conditional Random Field Approach. In Proceedings of the IJCNLP-08 workshop on NER for South and South East Asian Languages (pp105-110). Hyderabad, India.

- Asif Ekbal, Sivaji Bandyopadhyay. 2008 Bengali Named Entity Recognition using Support Vector Machine. In the Proceedings of the IJCNLP-08 workshop on NER for South and South East Asian Languages (pp 51-58). Hyderabad, India.

- Karthik Gali, Harshit Surana, Ashwini Vaidya, Praneeth Shishtla and Dipti Misra Sharma. 2008 Aggregating Machine Learning and Rule Based Heuristics for NER. In the Proceedings of the IJCNLP-08 workshop on NER for South and South East Asian Languages (pp 25-32). Hyderabad, India.

- Awaghad Ashish Krishnarao 2009 A Comparison of Performance of Sequential Learning Algorithm on task of NER for Indian Languages. In the Proceedings of the 9th International Conference on Computer Science (pp 123-132). Baton Rouge, LA, USA.

Index Terms

Computer Science
Natural Language
Processing

Key words

Punjabi language Named Entity Recognition (NER) Condition based approach

Punjabi proper nouns identification

Named Entities (NE) etc.

