

{tag}

{/tag}

Recent Advances in Information Technology

IJCA Proceedings on National Conference on

© 2014 by IJCA Journal

NCRAIT - Number 4

Year of Publication: 2014

Authors:

H. B. Patil

A. S. Patil

B. V. Pawar

{bibtex}ncrait1414.bib{/bibtex}

Abstract

Part-of-speech tagging in Marathi language is a very complex task as Marathi is highly inflectional in nature & free word order language. In this paper we have demonstrated a rule-based Part-of-Speech tagger for Marathi Language. The hand-constructed rules that are learned from corpus and some manual addition after studying the grammar of Marathi language are added and that are used for developing the tagger. Disambiguation is done by analyzing

the linguistic feature of the word, its preceding word, its following word, etc. After testing the system with three data sets we got encouraging results. The accuracy of our system is of an average 78.82% after testing it on three different data sets.

References

ences

- "A Part of Speech Tagger for Indian Languages", http://shiva.iiit.ac.in/SPSAL2007/iiit_tagset_guidelines.pdf
- A. Ratnaparkhi "A maximum entropy model for Part-of-Speech tagging", 1st Conference on Empirical Methods in Natural Language Processing (EMNLP-1996). PP133-142
- A. Bharati, V. Chaitanya and R. Sangal, "Computational Linguistics in India: An Overview", Proceedings of the 38th Annual Meeting on Association for Computational Linguistics 2000, VOL 38; PART 1, PP 595-596
- A. Azimzadeh, M. M. Arab, S. R. Quchani "Persian part of speech tagger based on Hidden Markov Model", JADT 2008 : 9es Journées internationales d'Analyse statistique des Données Textuelles , 2008, PP 121-128.
- A. Ramanathan, D. D. Rao, "A Lightweight Stemmer for Hindi", In Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2003. ,Workshop on Computational Linguistics for South Asian Languages (Budapest, April 2003).
- A. Dalal, K. Nagaraj, U. Sawant, S. Shelke and P. Bhattacharyya
- : "Building a Future Rich POS Tagger for Morphologically Rich Languages: Experiences in Hindi", ICON 2007, Hyderabad, India.
- Arulmozhi. P, Sobha. L. Kumara Shanmugam. B,
- : "Parts of Speech Tagger for Tamil", Symposium on Indian Morphology, Phonology & language Engineering IIT Kadagpur India March 19-21 2004 PP 55-57.
- A. Voutilainen, "A Syntax-based part-of-speech analyser", Conference of the European Chapter of the Association for Computational Linguistics, 1995, EACL – 95 PP 157-164.
- B. N. Patnaik, "Computational linguistics for Indian Languages", Symposium on Indian Morphology Phonology and Language Engineering 2004, PP 3-4.
- C. Samuelsson and A. Voutilainen "Comparing a Linguistic and a Stochastic Tagger", Proceedings of the 35th Annual meeting of the ACL and 8th Conference of the European chapter of the ACL 1997 PP 246-253.
- Dinesh Kumar & Gurpreet Singh Josan "Part of Speech Tagger for Morphologically rich Indian languages : A Survey", International Journal of Computer Applications Vol. 6, No. 5, September 2010.
- E. Brill, "Unsupervised Learning of Disambiguation Rules for Part of Speech Tagging", Proceedings of the Third Workshop on Very Large Corpora, 1995.
- E. Brill, "A Simple Rule Based Part of Speech Tagger", In Proceeding of the Third Conference on Applied Natural Language Processing 1992 Toronto, Italy, PP 152-155.
- F. M. Hasan, N. UzZaman and M. Khan "Comparison of Different POS Tagging Techniques (n-gram, HMM and Brill's Tagger) for Bangla", Proceedings of the

International Conference on Systems, Computing Sciences and Software Engineering (SCS2 06) of International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CIS2E 06), December 4 - 14, 2006.

- F. M. Hasan, N. UzZaman, and M. Khan, "Comparison of Unigram, Bigram, HMM and Brill's POS Tagging Approaches for some South Asian Languages.", Proceedings of the Conference on Language and Technology (CLT07), Pakistan, August 7 - 11, 2007
- F. Karlsson "Constraint grammar as a framework for parsing running text", In COLING-1990, PP 163-173.
- H. Schmid, "Part-of-Speech Tagging with Neural Networks", In Proceeding of the International Conference on Computational Linguistics 1994, Kyoto, Japan, PP 172-176.
- H. Schmid, "Probabilistic Part-of-Speech Tagging Using Decision Trees", In International Conference on New Methods in Language Processing 1994.
- J. Chanod and P. Tapanainen, "Tagging French- comparing a statistical and a constraint-based method", In EACL- 1995 PP 149-157.
- Jyoti Singh, Nisheeth Joshi Iti Mathur, "Part of Speech Tagging of Marathi text using trigram method", International Journal of Advanced Information Technology, Vol. 3, No. 2, April 2013.
- K. Bali, S. Baskaran, T. Bhattacharya, P. Bhattacharyya, M. Choudhury, G. Nath Jha, and et. al. , "A Common Part-of-Speech Tagset Framework for Indian Languages", Lexical Resources Engineering Conference (LREC08), Marrakech, Morocco, May 26-June 1, 2008.
- Kh. Raju Singha, Bipul Syam Purkayastha & kh. Dhiren Singha "Part of Speech Tagging in Manipuri with Hidden Markov Model", International Journal of Computer Science Issues, Vol. 9, No. 2, November 2012 PP: 146-149.
- Kh. Raju Singha, Bipul Syam Purkayastha & kh. Dhiren Singha "Part of Speech Tagging in Manipuri : A rule-based approach", International Journal of Computer Applications, Vol. 15, No. 14, August 2012.
- K. W. Church, "Current practice in Part of Speech Tagging and Suggestion for the Future", In Simmons (ed.) 1992 Sbornik Praci : In honor of Henry Kucera Michigan Slavic studies.
- K. Gupta, M. Shrivastava, S. Singh and P. Bhattacharyya, "Morphological Richness Offsets Resource Poverty- an Experience in Building a POS Tagger for Hindi", In Proceedings of the COLING/ACL on Main conference poster sessions , Sydney, Australia 2006,. PP: 779 – 786.
- K. T. Lua, "Part of Speech Tagging of Chinese Sentences Using Genetic Algorithm", Proceedings of ICC96, National University of Singapore, 1996.
- L. V. Gilder, "Automated Part Of Speech Tagging A Brief Overview", Handout for LING 361 Georgetown University Fall 1995.
- N. Agrawal, M. Shrivastava, S. Singh, B. Mohapatra, P. Bhattacharya, "Morphology Based Natural Language Processing tools for Indian Languages.", Workshop on Morphology 2005 . PP 71-75. Online link: http://www.cse.iitk.ac.in/users/iriss05/m_shrivastava.pdf
- R. M. Carrasco and A. Gelbukh, "Evaluation of TnT Tagger for Spanish.", Computer Science, 2003. ENC 2003. Proceedings of the Fourth Mexican International Conference on, ISBN:0-7695-1915-6 . PP: 18- 25 .

- S. Abney "Part-of-Speech Tagging and Partial Parsing", Corpus-Based Methods in Language and Speech Processing 1996,.
- S. Singh, K. Gupta, M. Shrivastava, P. Bhattacharyya "Morphological Richness Offsets Resource Demand- Experiences in Construction a Pos Tagger for Hindi. ", Proceedings of the COLING/ACL- 2006, on Main conference poster sessions. PP: 779 – 786. Sydney, Australia
- U. Sawant, S. Shelke, K. Nagaraj, and A. Dalal, "Hindi Part-of-Speech Tagging and Chunking: A Maximum Entropy Approach. ", Proceeding of the NLP AI Machine Learning, 2006.
- Y. Tlili-Guiassa, L. M. Tayeb "Tagging by Combining Rules-Based and Memory-based Learning", Information Technology Journal 5 (4), PP 679-684. 2006. ISSN:1812-5638.

Computer Science

Index Terms

Natural Language Processing

Keywords

Pos Tagger Morphological Analysis Rule-based.