May 2025: Top 10 Downloaded Articles in Natural Language Computing

International Journal on Natural Language Computing (IJNLC)

https://airccse.org/journal/ijnlc/index.html

ISSN: 2278 - 1307 [Online]; 2319 - 4111 [Print]

NAMED ENTITY RECOGNITION USING HIDDEN MARKOV MODEL (HMM)

Sudha Morwal [1] , Nusrat Jahan [2] and Deepti Chopra [3]

[1]Associate Professor, Banasthali University, Jaipur, Rajasthan-302001
[2] M.Tech (CS), Banasthali University, Jaipur, Rajasthan-302001
[3] M. Tech (CS), Banasthali University, Jaipur, Rajasthan-302001

ABSTRACT:

Named Entity Recognition (NER) is the subtask of Natural Language Processing (NLP) which is the branch of artificial intelligence. It has many applications mainly in machine translation, text to speech synthesis, natural language understanding, Information Extraction, Information retrieval, question answering etc. The aim of NER is to classify words into some predefined categories like location name, person name, organization name, date, time etc. In this paper we describe the Hidden Markov Model (HMM) based approach of machine learning in detail to identify the named entities. The main idea behind the use of HMM model for building NER system is that it is language independent and we can apply this system for any language domain. In our NER system the states are not fixed means it is of dynamic in nature one can use it according to their interest. The corpus used by our NER system is also not domain specific.

KEYWORDS

Named Entity Recognition (NER), Natural Language processing (NLP), Hidden Markov Model (HMM).

Full Text: **http://airccse.org/journal/ijnlc/papers/1412ijnlc02.pdf**

Volume Url: http://airccse.org/journal/ijnlc/vol1.html

REFERENCES

[1] Pramod Kumar Gupta, Sunita Arora "An Approach for Named Entity Recognition System for Hindi: An Experimental Study" in Proceedings of ASCNT – 2009, CDAC, Noida, India, pp. 103 – 108.

[2] Shilpi Srivastava, Mukund Sanglikar & D.C Kothari. "Named Entity Recognition System for Hindi Language: A Hybrid Approach" International Journal of Computational Linguistics (IJCL), Volume(2):Issue(1):2011.Availableat: http://cscjournals.org/csc/manuscript/Journals/IJCL/volume2/Issue1/IJCL-19.pdf

[3] "Padmaja Sharma, Utpal Sharma, Jugal Kalita"Named Entity Recognition: A Survey for the Indian Languages"(Language in India www.languageinindia.com 11:5 May 2011 Special Volume: Problems of Parsing in Indian Languages.) Available at: http://www.languageinindia.com/may2011/padmajautpaljugal.pdf.

[4] Lawrence R. Rabiner, " A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", In Proceedings of the IEEE, VOL.77,NO.2, February 1989.Available at: http://www.cs.ubc.ca/~murphyk/Bayes/rabiner.pdf.

[5] Sujan Kumar Saha, Sudeshna Sarkar, Pabitra Mitra "Gazetteer Preparation for Named Entity Recognition in Indian Languages" in the Proceeding of the 6th Workshop on Asian Language Resources, 2008 . Available at: http://www.aclweb.org/anthology-new/I/I08/I08-7002.pdf

[6] B. Sasidhar#1, P. M. Yohan*2, Dr. A. Vinaya Babu3, Dr. A. Govardhan4" A Survey on Named Entity Recognition in Indian Languages with particular reference to Telugu" in IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011 available at : http://www.ijcsi.org/papers/IJCSI-8-2-438-443.pdf.

[7] GuoDong Zhou Jian Su," Named Entity Recognition using an HMM-based Chunk Tagger" in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 473-480.

[8] http://en.wikipedia.org/wiki/Forward–backward_algorithm

[9] http://en.wikipedia.org/wiki/Baum-Welch_algorithm.

[10] Dan Shen, jie Zhang, Guodong Zhou,Jian Su, Chew-Lim Tan" Effective Adaptation of a Hidden Markov Model-based Named Entity Recognizer for Biomedical Domain" available at: http://acl.ldc.upenn.edu/W/W03/W03-1307.pdf.

RESUME INFORMATION EXTRACTION WITH A NOVEL TEXT BLOCK SEGMENTATION ALGORITHM

Shicheng Zu and Xiulai Wang

Post-doctoral Scientific Research Station in East War District General Hospital, Nanjing, Jiangsu 210000, China

ABSTRACT

In recent years, we have witnessed the rapid development of deep neural networks and distributed representations in natural language processing. However, the applications of neural networks in resume parsing lack systematic investigation. In this study, we proposed an end-to-end pipeline for resume parsing based on neural networks-based classifiers and distributed embeddings. This pipeline leverages the position-wise line information and integrated meanings of each text block. The coordinated line classification by both line type classifier and line label classifier effectively segment a resume into predefined text blocks. Our proposed pipeline joints the text block segmentation with the identification of resume facts in which various sequence labelling classifiers perform named entity recognition within labelled text blocks. Comparative evaluation of four sequence labelling classifiers confirmed BLSTMCNNs-CRF's superiority in named entity recognition task. Further comparison among three publicized resume parsers also determined the effectiveness of our text block classification method.

KEYWORDS

Resume Parsing, Word Embeddings, Named Entity Recognition, Text Classifier, Neural Networks

Full Text: https://aircconline.com/ijnlc/V8N5/8519ijnlc03.pdf

Volume URL: https://airccse.org/journal/ijnlc/vol8.html

REFERENCES

[1] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu (2016) "Attention-based Bidirectional Long Short-term Memory Networks for Relation Classification", In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL'16), Berlin, Germany, August 7-12, 2016, pp 207-212.

[2] Xuezhe Ma, & Eduard Hovy (2016) "End-to-End Sequence Labelling via Bi-directional LSTMCNNs-CRF", In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL'16), Berlin, Germany, August 7-12, 2016, pp 1064-1074.

[3] Kun Yu, Gang Guan, and Ming Zhou (2005) "Resume Information Extraction with Cascaded Hybrid Model" In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), Stroudsburg, PA, USA, June 2005, pp 499-506.

[4] Jie Chen, Chunxia Zhang, and Zhendong Niu (2018) "A Two-Step Resume Information Extraction Algorithm" Mathematical Problems in Engineering pp1-8. [5] Jie Chen, Zhendong Niu, and Hongping Fu (2015) "A Novel Knowledge Extraction Framework for Resumes Based on Text Classifier" In: Dong X., Yu X., Li J., Sun Y. (eds) Web-Age Information Management (WAIM 2015) Lecture Notes in Computer Science, Vol. 9098, Springer, Cham.

[6] Hui Han, C. Lee Giles, Eren Manavoglu, HongYuan Zha (2003) "Automatic Document Metadata Extraction using Support Vector Machine" In Proceedings of the 2003 Joint Conference on Digital Libraries, Houston, TX, USA, pp 37-48.

[7] David Pinto, Andrew McCallum, Xing Wei, and W. Bruce Croft (2003) "Table Extraction Using Conditional Random Field" In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, Toronto, Canada, pp 235-242.

[8] Amit Singh, Catherine Rose, Karthik Visweswariah, Enara Vijil, and Nandakishore Kambhatla (2010) "PROSPECT: A system for screening candidates for recruitment" In Proceedings of the 19th ACM international conference on Information and knowledge management, (CIKM'10), Toronto, ON, Canada, October 2010, pp 659-668.

[9] Anjo Anjewierden (2001) "AIDAS: Incremental Logical Structure Discovery in PDF Documents" In Proceedings of 6th International Conference on Document Analysis and Recognition (ICDAR'01) pp 374-378.

[10] Sumit Maheshwari, Abhishek Sainani, and P. Krishna Reddy (2010) "An Approach to Extract Special Skills to Improve the Performance of Resume Selection" Databases in Networked Information Systems, Vol. 5999 of Lecture Notes in Computer Science, Springer, Berlin, Germany, 2010, pp 256-273.

[11] Xiangwen Ji, Jianping Zeng, Shiyong Zhang, Chenrong Wu (2010) "Tag tree template for Web information and schema extraction" Expert Systems with Applications Vol. 37, No.12, pp 8492- 8498.

[12] V. Senthil Kumaran and A. Sankar (2013) "Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping (EXPERT)" International Journal of Metadata, Semantics and Ontologies, Vol. 8, No. 1, pp 56-64.

[13] Fabio Ciravegna (2001) "(LP)2, an Adaptive Algorithm for Information Extraction from Webrelated Texts" In Proceedings of the IJCAI-2001 Workshop on Adaptive Text Extraction and Mining. Seattle, WA.

[14] Fabio Ciravegna, and Alberto Lavelli (2004) "LearningPinocchio: adaptive information extraction for real world applications" Journal of Natural Language Engineering Vol. 10, No.

2, pp145- 165.
[15] Yan Wentan, and Qiao Yupeng (2017) “Chinese resume information extraction based on semistructure text” In 36th Chinese Control Conference (CCC), Dalian, China.
[16] Zhang Chuang, Wu Ming, Li Chun Guang, Xiao Bo, and Lin Zhi-qing (2009) “Resume Parser: Semi-structured Chinese document analysis” In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, USA, Vol. 5 pp 12-16.
[17] Zhixiang Jiang, Chuang Zhang, Bo Xiao, and Zhiqing Lin (2009) “Research and Implementation of Intelligent Chinese resume Parsing” In 2009 WRI International Conference on Communications and Mobile Computing, Yunan, China, Vol. 3 pp 588-593.
[18] Duygu Çelik, Askýn Karakas, Gülsen Bal , Cem Gültunca , Atilla Elçi , Basak Buluz, and Murat Can Alevli (2013) “Towards an Information Extraction System based on Ontology to Match Resumes and Jobs” In Proceedings of the 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops, Japan, pp 333-338.
[19] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean (2013) “Efficient Estimation of Word Representations in Vector Space” Computer Science, arXiv preprint arxiv:1301.3781.
[20] Jeffrey Pennington, Richard Socher, and Christopher D. Manning (2014) “GloVe: Global Vectors for Word Representation” In Empirical Methods in Natural Language Processing (EMNLP) pp 1532-1543.

SURVEY OF MACHINE TRANSLATION SYSTEMS IN INDIA

G V Garje[1] and G K Kharate[2]

[1]Department of Computer Engineering and Information Technology PVG's College of Engineering and Technology, Pune, India
[2] Principal, Matoshri College of Engineering and Research Centre, Nashik, India

ABSTRACT

The work in the area of machine translation has been going on for last few decades but the promising translation work began in the early 1990s due to advanced research in Artificial Intelligence and Computational Linguistics. India is a multilingual and multicultural country with over 1.25 billion population and 22 constitutionally recognized languages which are written in 12 different scripts. This necessitates the automated machine translation system for English to Indian languages and among Indian languages so as to exchange the information amongst people in their local language. Many usable machine translation systems have been developed and are under development in India and around the world. The paper focuses on different approaches used in the development of Machine Translation Systems and also briefly described some of the Machine Translation Systems along with their features, domains and limitations.

KEYWORDS

Machine Translation, Example-based MT, Transfer-based MT, Interlingua-based MT

Full Text: http://airccse.org/journal/ijnlc/papers/2513ijnlc04.pdf

Volume URL: http://airccse.org/journal/ijnlc/vol2.html

REFERENCES

[1] Sitender & Seema Bawa, (2012) "Survey of Indian Machine Translation Systems", International Journal Computer Science and Technolgy, Vol. 3, Issue 1, pp. 286-290, ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

[2] Sanjay Kumar Dwivedi & Pramod Premdas Sukhadeve, (2010) "Machine Translation System in Indian Perspectives", Journal of Computer Science 6 (10): 1082-1087, ISSN 1549-3636, © 2010 Science

[3] John Hutchins, (2005) "Current commercial machine translation systems and computer-based translation tools: system types and their uses", International Journal of Translation vol.17, no.1-2, pp.5-38.

[4] Vishal Goyal & Gurpreet Singh Lehal, (2009) "Advances in Machine Translation Systems", National Open Access Journal, Volume 9, ISSN 1930-2940 http://www.languageinindia.

[5] Latha R. Nair & David Peter S., (2012) "Machine Translation Systems for Indian Languages", International Journal of Computer Applications (0975 – 8887) Volume 39– No.1

[6] Vishal Goyal & Gurpreet Singh Lehal, (2010) "Web Based Hindi to Punjabi Machine Translation System", International Journal of Emerging Technologies in Web Intelligence, Vol. 2, no. 2, pp. 148-151, ACADEMY PUBLISHER

[7] Shachi Dave, Jignashu Parikh & Pushpak Bhattacharyya, (2002) "Interlingua-based English-Hindi Machine Translation and Language Divergence", Journal of Machine Translation, pp. 251-304.

[8] Sudip Naskar & Shivaji Bandyopadhyay, (2005) "Use of Machine Translation in India: Current status" AAMT Journal, pp. 25-31.

[9] Sneha Tripathi & Juran Krishna Sarkhel, (2010) "Approaches to Machine Translation", International journal of Annals of Library and Information Studies, Vol. 57, pp. 388-393

[10] Gurpreet Singh Josan & Jagroop Kaur, (2011) "Punjabi To Hindi Statistical Machine Transliteration", International Journal of Information Technology and Knowledge Management , Volume 4, No. 2, pp. 459-463.

[11] S. Bandyopadhyay, (2004) "ANUBAAD - The Translator from English to Indian Languages", in proceedings of the VIIth State Science and Technology Congress. Calcutta. India. pp. 43-51

[12] R.M.K. Sinha & A. Jain, (2002) "AnglaHindi: An English to Hindi Machine-Aided Translation System", International Conference AMTA(Association of Machine Translation in the Americas)

[13] Murthy. K, (2002) "MAT: A Machine Assisted Translation System", In Proceedings of Symposium on Translation Support System(STRANS-2002), IIT Kanpur. pp. 134-139.

[14] Lata Gore & Nishigandha Patil, (2002) "English to Hindi - Translation System", In proceedings of Symposium on Translation Support Systems. IIT Kanpur. pp. 178-184.

[15] Kommaluri Vijayanand, Sirajul Islam Choudhury & Pranab Ratna "VAASAANUBAADA - Automatic Machine Translation of Bilingual Bengali-Assamese News Texts", in proceedings of Language Engineering Conference-2002, Hyderabad, India © IEEE Computer Society.

[16] Bharati, R. Moona, P. Reddy, B. Sankar, D.M. Sharma & R. Sangal, (2003) "Machine Translation: The Shakti Approach", Pre-Conference Tutorial, ICON-2003.

[17] S. Mohanty & R. C. Balabantaray, (2004) "English to Oriya Translation System (OMTrans)" cs.pitt.edu/chang/cpol/c087.pdf
[18] Ananthakrishnan R, Kavitha M, Jayprasad J Hegde, Chandra Shekhar, Ritesh Shah, Sawani Bade & Sasikumar M., (2006) "MaTra: A Practical Approach to Fully- Automatic Indicative EnglishHindi Machine Translation", In the proceedings of MSPIL-06.
[19] G. S. Josan & G. S. Lehal, (2008) "A Punjabi to Hindi Machine Translation System", in proceedings of COLING-2008: Companion volume: Posters and Demonstrations, Manchester, UK, pp. 157-160.
[20] Sanjay Chatterji, Devshri Roy, Sudeshna Sarkar & Anupam Basu, (2009) "A Hybrid Approach for Bengali to Hindi Machine Translation", In proceedings of ICON-2009, 7th International Conference on Natural Language Processing, pp. 83-91.
[21] Vishal Goyal & Gurpreet Singh Lehal, (2011) "Hindi to Punjabi Machine Translation System", in proceedings of the ACL-HLT 2011 System Demonstrations, pages 1–6, Portland, Oregon, USA, 21 June 2011.
[22] Ankit Kumar Srivastava, Rejwanul Haque, Sudip Kumar Naskar & Andy Way, (2008) "The MATREX (Machine Translation using Example): The DCU Machine Translation System for ICON 2008", in Proceedings of ICON-2008: 6th International Conference on Natural Language Processing, Macmillan Publishers, India, http://ltrc.iiit.ac.in/proceedings/ICON-2008.
[23] hutchinsweb.me.uk/Nutshell-2005.pdf
[24] John Hutchins "Historical survey of machine translation in Eastern and Central Europe", Based on an unpublished presentation at the conference on Crosslingual Language Technology in service of an integrated multilingual Europe, 4-5 May 2012, Hamburg, Germany. (www.hutchinsweb.me.uk/Hamburg-2012.pdf)
[25] Sampark: Machine Translation System among Indian languages (2009) http://tdildc.in/index.php?option=com_vertical&parentid=74, http://sampark.iiit.ac.in/
[26] Akshar Bharti, Chaitanya Vineet, Amba P. Kulkarni & Rajiv Sangal, (1997) "ANUSAARAKA: Machine Translation in stages', Vivek, a quarterly in Artificial Intelligence, Vol. 10, No. 3, NCST Mumbai, pp. 22-25
[27] Akshar Bharti, Chaitanya Vineet, Amba P. Kulkarni & Rajiv Sangal, (2001) "ANUSAARAKA: overcoming the language barrier in India", published in Anuvad: approaches to Translation
[28] Hemant Darabari, (1999) "Computer Assisted Translation System- An Indian Perspective", in proceedings of MT Summit VII, Thialand [29] R. Mahesh K. Sinha & Anil Thakur, (2005) "Machine Translation of Bi-lingual Hindi-English (Hinglish) Text", in proceedings of 10th Machine Translation Summit organized by Asia-Pacific Association for Machine Translation (AAMT), Phuket, Thailand
[30] Parameswari K, Sreenivasulu N.V., Uma Maheshwar Rao G & Christopher M, (2012) "Development of Telugu-Tamil Bidirectional Machine Translation System: A special focus on case divergence", in proceedings of 11th International Tamil Internet conference, pp 180- 191
[31] Salil Badodekar, (2004) "Translation Resources, Services and Tools for Indian Languages", a report of Centre for Indian Language Technology, IITB, http://www.cfilt.iitb.ac.in/Translationsurvey/survey.pdf
[32] Ananthakrishnan R, Kavitha M, Jayprasad J Hegde, Chandra Shekhar, Ritesh Shah, Sawani Bade & Sasikumar M, (2006) "MaTra: A Practical Approach to Fully-Automatic

Indicative EnglishHindi Machine Translation", in proceedings of the first national symposium on Modelling and shallow parsing of Indian languages (MSPIL-06) organized by IIT Bambay, 202.141.152.9/clir/papers/matra_mspil06.pdf
[33] CDAC Mumbai, (2008) "MaTra: an English to Hindi Machine Translation System", a report by CDAC Mumbai formerly NCST.
[34] Sanjay Chatterji, Praveen Sonare, Sudeshna Sarkar & Anupam Basu, (2011) "Lattice Based Lexical Transfer in Bengali Hindi Machine Translation Framework", in Proceedings of ICON2011: 9th International Conference on Natural Language Processing, Macmillan Publishers, India. Also accessible from ltrc.iiit.ac.in/proceedings/ICON-2011.
[35] R. Ananthakrishnan, Jayprasad Hegde, Pushpak Bhattacharyya, Ritesh Shah & M. Sasikumar, (2008) "Simple Syntactic and Morphological Processing Can Help English-Hindi Statistical Machine Translation", in proceedings of International Joint Conference on NLP (IJCNLP08), Hyderabad, India.
[36] Yanjun Ma, John Tinsley, Hany Hassan, Jinhua Du & Andy Way, (2008) "Exploiting Alignment Techniques in MATREX: the DCU Machine Translation System for IWSLT 2008', in proceedings of IWSLT 2008, Hawaii, USA
[37] projects.uptuwatch.com/cs-it/anubharti-an-hybrid-example-based-approach-for-machine-aidedtrapnslation/
[38] Sugata Sanyal & Rajdeep Borgohain, (2013) "Machine Translation Systems in India", Cornel University Library, arxiv.org/ftp/arxiv/papers/1304/1304.7728.pdf [39] Antony P. J., (2013) "Machine Translation Approaches and Survey for Indian Languages", International journal of Computational Linguistics and Chinese Language Processing Vol. 18, No. 1, pp. 47-78.
[40] Manoj Jain & Om P. Damani, (2009) "English to UNL (Interlingua) Enconversion", in proceedings of 4th Language and Translation Conference (LTC-09).
[41] Smriti Singh, Mrugank Dalal, Vishal Vachhani, Pushpak Bhattacharyya & Om P. Damani, (2007) "Hindi Generation from Interlingua (UNL)", in proceedings of MT Summit, 2007
[42] language.worldofcomputing.net
[43] sampark.iiit.ac.in [44] www.cdacmumbai.in/xlit [
45] www.cdacmumbai.in/rupantar
[46] translationjournal.net/journal/29computers.htm
[47] www.cfilt.iitb.ac.in/resources/surveys/MT-Literature%20Survey-2012-Somya.pdf
[48] www.cdacmumbai.in/e-ilmt
[49] www.iiit.net/ltrc/Anusaaraka/anu_home.html
[50] cdac.in/html/aai/mantra.asp
[51] translate.google.com/about/intl/en_ALL/

GRAMMAR CHECKERS FOR NATURAL LANGUAGES: A REVIEW

Nivedita S. Bhirud[1] R.P. Bhavsar[2] B.V. Pawar[3]

[1]Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, India
[2,3]School of Computer Sciences, North Maharashtra University, Jalgaon, India

ABSTRACT

Natural Language processing is an interdisciplinary branch of linguistic and computer science studied under the Artificial Intelligence (AI) that gave birth to an allied area called 'Computational Linguistics' which focuses on processing of natural languages on computational devices. A natural language consists of many sentences which are meaningful linguistic units involving one or more words linked together in accordance with a set of predefined rules called 'grammar'. Grammar checking is fundamental task in the formal world that validates sentences syntactically as well as semantically. Grammar Checker tool is a prominent tool within language engineering. Our review draws on the till date development of various Natural Language grammar checkers to look at past, present and the future in the present context. Our review covers common grammatical errors , overview of grammar checking process, grammar checkers of various languages with the aim of seeking their approaches, methodologies and performance evaluation, which would be great help for developing new tool and system as a whole. The survey concludes with the discussion of different features included in existing grammar checkers of foreign languages as well as a few Indian Languages.

KEYWORDS

Natural Language Processing, Computational Linguistics, Writing errors,Grammatical mistakes, Grammar Checker

Full Text: https://aircconline.com/ijnlc/V6N4/6417ijnlc01.pdf

Volume URL: https://airccse.org/journal/ijnlc/vol6.html

REFERENCES

[1] Misha Mittal, Dinesh Kumar, Sanjeev Kumar Sharma, "Grammar Checker for Asian Languages: A Survey", International Journal of Computer Applications & Information Technology Vol. 9, Issue I, 2016

[2] DebelaTesfaye, "A rule-based Afan Oromo Grammar Checker", International Journal of Advanced Computer Science and Applications, Vol. 2, No. 8, 2011

[3] Aynadis Temesgen Gebru, 'Design and development of Amharic Grammar Checker', 2013

[4] Arppe, Antti. "Developing a Grammar Checker for Swedish". The 12th Nordic conference computational linguistic. 2000. PP. 13 – 27.

[5] "A prototype of a grammar checker for Icelandic", available at www.ru.is/~hrafn/students/BScThesis_Prototype_Icelandic_GrammarChecker.pdf

[6] Bal Krishna Bal, Prajol Shrestha, "Architectural and System Design of the Nepali Grammar Checker", www.panl10n.net/english/.../Nepal/Microsoft%20Word%20-%208_OK_N_400.pdf

[7] Kinoshita, Jorge; Nascimento, Lanѕ do; Dantas ,Carlos Eduardo. "CoGrOO: a Brazilian-Portuguese Grammar Checker based on the CETENFOLHA Corpus". Universidade da Sro Paulo (USP), Escola Politйcnica. 2003.

[8] Domeij, Rickard; Knutsson, Ola; Carlberger, Johan; Kann, Viggo. "Granska: An efficient hybrid system for Swedish grammar checking". Proceedings of the 12th Nordic conference in computational linguistic, Nodalida- 99. 2000.

[9] Jahangir Md; Uzzaman, Naushad; Khan, Mumit. "N-Gram Based Statistical Grammar Checker For Bangla And English". Center for Research On Bangla Language Processing. Bangladesh, 2006.

[10] Singh, Mandeep; Singh, Gurpreet; Sharma, Shiv. "A Punjabi Grammar Checker". Punjabi University. 2nd international conference of computational linguistics: Demonstration paper. 2008. pp. 149 – 132.

[11] Steve Richardson. "Microsoft Natuaral language Understanding System and Grammar checker". Microsoft.USA, 1997.

[12] Daniel Naber. "A Rule-Based Style And Grammar Checker". Diplomarbeit. Technische Fakultät Bielefeld, 2003

[13] "Brief History of Grammar Check Software", available at: http://www.grammarcheck.net/briefhistory-of-grammar-check software/, Accessed On October 28, 2011.

[14] Gelbukh, Alexander. "Special issue: Natural Language Processing and its Applications". InstitutoPolitécnico Nacional. Centro de InvestigaciónenComputación. México 2010.

[15] H. Kabir, S. Nayyer, J. Zaman, and S. Hussain, "Two Pass Parsing Implementation for an Urdu Grammar Checker."

[16] Jaspreet Kaur, Kamaldeep Garg, " Hybrid Approach for Spell Checker and Grammar Checker for Punjabi," vol. 4, no. 6, pp. 62–67, 2014.

[17] LataBopche, GauriDhopavkar, and ManaliKshirsagar, "Grammar Checking System Using Rule Based Morphological Process for an Indian Language", Global Trends in Information Systems and Software Applications, 4th International Conference, ObCom 2011 Vellore, TN, India, December 9-11, 2011.

[18] MadhaviVaralwar, Nixon Patel. “Characteristics of Indian Languages” available at “http://http://www.w3.org/2006/10/SSML/papers/CHARACTERISTICS_OF_INDIAN_LANGUAG ES.pdf” on 30/12/2013
[19] Kenneth W. Church andLisa F. Rau, “CommercialApplications of Natural Language Processing”, COMMUNICATIONS OF THE ACM ,Vol. 38, No. 11, November 1995
[20] ChandhanaSurabhi.M, “Natural Language Processing Future”, Proceedings of International Conference on Optical Imaging Sensor and Security, Coimbatore, Tamil Nadu, India, July 2-3, 2013
[21] ER-QING XU , “ NATURAL LANGUAGE GENERATION OF NEGATIVE SENTENCES IN THE MINIMALIST PARADIGM”, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005
[22] Jurafsky Daniel, H., James. Speech and Language Processing: “An introduction to natural language processing, computational linguistics and speech recognition” June 25, 2007.
[23] Aronoff ,Mark; Fudeman, Kirsten. “What is Morphology?”. Blackwell publishing. Vol 8. 2001. [24] S., Philip; M.W., David. “A Statistical Grammar Checker”. Department of Computer Science. Flinders University of South Australia.South Australia, 1996
[25] Mochamad Vicky Ghani Aziz, Ary Setijadi Prihatmanto, Diotra Henriyan, Rifki W�jaya, “Design and Implementation of Natural Language Processing with Syntax and Semantic Analysis for Extract Traffic Conditions from Social Media Data” , IEEE 5th International Conference on System Engineering and Technology,Aug.10-11,UiTM,ShahAlam,Malaysia,2015
[26] Chandhana Surabhi.M,”Natural Language Processing Future”, Proceedings of International Conference on Optical Imaging Sensor and Security, Coimbatore, Tamil Nadu, India, July 2-3, 2013
[27] Blossom Manchanda, Vijay Anant Athvale, Sanjeev Kumar Sharma, “ Various Techniques used for Grammar Checking”, Internation Journal of Computer Application & information Technology, Vol. 9, Issue 1, 2016
[28] Simon Ager,1998-2017, Language Index[online].Available: http://www.omniglot.com/writing/languages.htm
[29] Nitin Indurkhya & Fred J. Damerau ,”A Handbook of Natural Language Processing” Cambridge UK [30] Available:http://authority.pub/common-grammar-mistakes/
[31] Mo.Ra. Walambe, “Sugam Marathi Vyakaran va Lekhan”, Nitin Prakashan, 1988.

MACHINE TRANSLATION DEVELOPMENT FOR INDIAN LANGUAGES AND ITS APPROACHES

Amruta Godase[1] and Sharvari Govilkar[2]

[1]Department of Information Technology (AI & Robotics), PIIT, Mumbai University, India
[2]Department of Computer Engineering, PIIT, Mumbai University, India

ABSTRACT

This paper presents a survey of Machine translation system for Indian Regional languages. Machine translation is one of the central areas of Natural language processing (NLP). Machine translation (henceforth referred as MT) is important for breaking the language barrier and facilitating inter-lingual communication. For a multilingual country like INDIA which is largest democratic country in whole world, there is a big requirement of automatic machine translation system. With the advent of Information Technology many documents and web pages are coming up in a local language so there is a large need of good MT systems to address all these issues in order to establish a proper communication between states and union governments to exchange information amongst the people of different states. This paper focuses on different Machine translation projects done in India along with their features and domain.

KEYWORDS

Machine translation, computational linguistics, Indian Languages, Rule-based, Statistical, Empirical MT, Principle-based, Knowledge-based, Hybrid

Full Text: https://airccse.org/journal/ijnlc/papers/4215ijnlc05.pdf

Volume URL: https://airccse.org/journal/ijnlc/vol4.html

SURVEY ON MACHINE TRANSLITERATION AND MACHINE LEARNING MODELS

M L Dhore[1] ,R M Dhore[2] and P H Rathod[3]

[1,3]Vishwakarma Institute of Technology, Savitribai Phule Pune University, India

[2] Pune Vidhyarthi Girha's College of Engineering and Technology, SPPU, India

ABSTRACT

Globalization and growth of Internet users truly demands for almost all internet based applications to support local languages. Support of local languages can be given in all internet based applications by means of Machine Transliteration and Machine Translation. This paper provides the thorough survey on machine transliteration models and machine learning approaches used for machine transliteration over the period of more than two decades for internationally used languages as well as Indian languages. Survey shows that linguistic approach provides better results for the closely related languages and probability based statistical approaches are good when one of the languages is phonetic and other is nonphonetic.Better accuracy can be achieved only by using Hybrid and Combined models.

KEYWORDS

CRF, Grapheme, HMM, Machine Transliteration, Machine Learning, NCM, Phoneme, SVM

Full Text: https://airccse.org/journal/ijnlc/papers/4215ijnlc02.pdf

Volume URL: https://airccse.org/journal/ijnlc/vol4.html

REFERENCES

[1] Karimi S, Scholer F, & Turpin, (2011) "Machine Transliteration Survey", ACM Computing Surveys, Vol. 43, No. 3, Article 17, pp.1-46.

[2] Antony P J &Soman K P, (2011) "Machine Transliteration for Indian Languages: A Literature Survey", International Journal of Scientific and Engineering Research, Vol 2, Issue 12, pp. 1-8.

[3] Jong-Hoon Oh, Key-Sun Choi & Hitoshi Isahara, (2006) "A Comparison of Different Machine Transliteration Models", Journal of Artificial Intelligence Research, pp. 119-151.

[4] Brown P F, Pietra V J D, Pietra S A D, & Mercer R L, (1993) "The Mathematics of Statistical Machine Translation: Parameter estimation", Computational Linguistic, 19, 2 pp. 263–311.

[5] Knight Kevin &Graehl Jonathan, (1998) "Machine Transliteration", In Proceedings of the 35th Annual Meetings of The Association for Computational Linguistics, pp. 128-135..

[6] Li Haizhou et al., (2004) "A Joint Source-Channel Model for Machine Transliteration", ACL.

[7] L Rabiner, (1989) "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of IEEE, Vol. 77, No. 2, pp. 257-296.

[8] Phil Blunsom, (2004) "Hidden Markov Models".

[9] J Lafferty et al., (2001) "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data", In International Conference on Machine Learning.

[10] Hanna M. Wallach, (2004) "Conditional Random Fields: An Introduction", University of Pennsylvania CIS Technical Report MS-CIS-04-21.

[11] Charles S et al., "An Introduction to CRF Relational Learning", University of Massachusetts.

[12] A L Berger, S D Pietra, & V J Della Pietra, (1996) "A Maximum Entropy Approach to Natural Language Processing", Computational Linguistics, vol. 22, no. 1, pp. 39–71.

[13] K.P.Soman et al, Machine Learning with SVM and Other Kernel Methods, Book, PHI.

[14] Y. Yuan et al. (1995) Fuzzy sets and Systems, pp 125-139.

[15] Lee J S & Choi K S, (1998) "English to Korean Statistical Transliteration For Information Retrieval", Computer Processing of Oriental Languages.

[16] Kang I H et al., (2000) "English-to-Korean Transliteration Using Multiple Unbounded Overlapping Phoneme Chunks", In Proceedings of the 18th Conference on Coling, pp. 418–424.

[17] Kang B J et al., (2000) "Automatic Transliteration & Back-Transliteration by Decision Tree Learning", 2nd International Conference on Language Resources and Evaluation.

[18] Kang B J (2001) "A Resolution of Word Mismatch Problem Caused by Foreign Word Transliterations and English Words in Korean Information Retrieval", Ph.D. Thesis, KAIST.

[19] Goto I et al, (2003) "Transliteration Considering Context Information Based on the Maximum Entropy Method", In Proceedings of MT-Summit IX, pp. 125-132.

[20] Jaleel et al, (2003) “Statistical Transliteration For English-Arabic Cross Language Information Retrieval”, 12th International Conference on Information and Knowledge Management.

[21] Lee, J et al., (2003), “Acquisition of English-Chinese Transliterated Word Pairs from Parallel-Aligned Texts using a Statistical Machine Transliteration Model”, HLT-NAACL 2003.

[22] Li H et al., (2004) “A Joint Source-Channel Model for Machine Transliteration”, ACL.

[23] Malik M G A, (2006) “Punjabi Machine Transliteration”, In Proceedings of the 21st International Conference on Computational Linguistics, ACL, pp.1137-1144.

[24] Ekbal A, Naskar S &Bandyopadhyay S, (2006) “A Modified Joint Source Channel Model for Transliteration”, In Proceedings of the COLING-ACL, Australia, pp.191-198.

[25] Kumaran A et al., (2007) “A Generic Framework for Machine Transliteration”, 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

[26] Hermjakob, U. Et al., (2008) “Name Translation in Statistical Machine Translation Learning When to Transliterate”, Proceedings of Association for Computational Linguistics, pp. 389–397.

[27]Ganesh S, Harsha S, Pingali P, &Verma V, (2008) “Statistical Transliteration for Cross Language Information Retrieval Using HMM Alignment and CRF”, In Proceedings of the Workshop on CLIA, Addressing the Needs of Multilingual Societies.

[28] Rama T. Et al., (2009) “Modeling Machine Transliteration as a Phrase Based Statistical Machine Translation Problem”, Proceedings of the 2009 Named Entities Workshop, pp. 124-127.

[29] Martin Jansche& Richard Sproat, (2009) “Named Entity Transcription with Pair n-Gram Models”, Google Inc., Proceedings of the 2009 Named Entities Workshop, Singapore pp. 32–35.

[30] Jong-Hoon Oh et al., (2009) “Ma-chine Transliteration Using Target-Language Grapheme and Phoneme: Multi-Engine Transliteration Approach”, Named Entities Workshop, pp. 36–39.

[31] SittichaiJiampojamarn et al, (2009) “DirecTL: a Language Independent Approach to Transliteration”, Proceedings of the 2009 Named Entities Workshop, Singapore, pp. 28–31.

[32] Paul, M. Et al., (2009) “Model Adaptation and Transliteration for Spanish-English SMT”, Proceedings of the 4th EACL Workshop on Statistical Machine Translation, pp. 105- 109.

[33] KommaluriVijayanand, (2009) “Testing and Performance Evaluation of Machine Transliteration System for Tamil Language”, Proceedings of the 2009 NEWS, pp. 48–51.

[34] Finch, A. &Sumita, E, (2009) “Transliteration by Bidirectional Statistical Machine Translation”, Proceedings of the 2009 Named Entities Workshop, pp. 52-56.

[35] Xue Jiang, Le Sun &Dakun Zhang, (2009) “A Syllable-Based Name Transliteration System”, Proceedings of the 2009 Named Entities Workshop, Singapore, pp. 96–99.

[36] Vijaya M.S. et al., (2009) “English to Tamil Transliteration using WEKA”, International

Journal of Recent Trends in Engineering, Vol. 1, No. 1, pp. 498-500.

[37] Das A., Ekbal A., Mandal T. &Bandyopadhyay S, (2009) "English to Hindi Machine Transliteration System at NEWS", Proceedings of the 2009 Named Entities Workshop pp.80-83. [38] Chai Wutiwiwatchai and AusdangThangthai, (2010) "Syllable-based Thai-English Machine Transliteration", Named Entities Workshop Sweden pp. 66-70.

[39] Josan, G. &Lehal, G, (2010) "A Punjabi to Hindi Machine Transliteration System", Computational Linguistics and Chinese Language Processing, Vol. 15, No. 2, pp. 77-102, 2010.

[40] Chinnakotla M K, Damani O P, and Satoskar A, (2010) "Transliteration for Resource-Scarce Languages", ACM Transactions on Asian Language Information Processing, 9, 4, pp. 1-30.

[41] Fehri H et al., (2011) "Recognition and Translation of Arabic Named Entities with NooJ Using a New Representation Model", 9th International Workshop on FSM and NLP, pp.134–142.

[42] Deep, K. &Goyal, V, (2011) "Development of a Punjabi to English Transliteration System", International Journal of Computer Science and Communication, Vol. 2, No. 2, pp. 521-526. [43] Kaur, J. &Josan, G, (2011) "Statistical Approach to Transliteration from English to Punjabi", International Journal on Computer Science and Engineering, Vol. 3, No. 4, pp. 1518-1527.

[44] Josan, G. &Kaur, J, (2011) "Punjabi To Hindi Statistical Machine Transliteration", International Journal of Information Technology and Knowledge Management, pp. 459-463.

[45] Dhore Manikrao L, Dixit Shantanu K and Sonwalkar Tushar D, (2012) "Hindi to English Machine Transliteration of Named Entities using Conditional Random Fields", International Journal of Computer Applications, Vol. 48– No.23, pp. 31-37.

[46] Sharma S. Et al., (2012) "English-Hindi Transliteration using Statistical Machine Translation in different Notation", International Conference on Computing and Control Engineering.

[47] Kumar, P. and Kumar, V, (2013) "Statistical Machine Translation Based Punjabi to English Transliteration System for Proper Nouns", International Journal of Application or Innovation in Engineering & Management, Vol. 2, Issue 8, pp. 318-321.

[48] Rathod P H, Dhore M L and Dhore R M, (2013) "Hindi And Marathi To English Machine Transliteration Using SVM", International Journal on Natural Language Computing (IJNLC) Vol. 2, No.4, pp. 55-71.

[49] Bhalla, D. and Joshi, N, (2013) "Rule Based Transliteration Scheme For English To Punjabi", International Journal on Natural Language Computing, Vol. 2, No. 2, pp. 67-73.

[50] Joshi, H., Bhatt, A. & Patel. H, (2013) "Transliterated Search using Syllabification Approach", Forum for Information Retrieval Evaluation.

[51] Arbabi M, Fischthal S M, Cheng V C & Bart E, (1994) "Algorithms for Arabic Name Transliteration", IBM Journal of Research and Development, pp. 183-194.

[52] Stephen Wan & Cornelia Maria Verspoor, (1998) "Automatic English-Chinese Name Transliteration for Development of Multilingual Resources", NSW 2109, pp. 1352-1356.

[53] Stalls, B. & Knight K, (1998) "Translating Names and Technical Terms in Arabic Text",

COLING ACL Workshop on Computational Approaches to Semitic Languages, pp. 34-41, 1998.

[54] Lee J S, (1999) “An English-Korean Transliteration and Re-transliteration Model for Cross-Lingual Information Retrieval”, Computer Science Dept., KAIST.

[55] Jeong K S et al., (1999) “Automatic Identification and Back-Transliteration of Foreign Words for Information Retrieval”, Information Processing and Management, 35, 4, pp. 523–540.

[56] Jung S Y et al., (2000) “An English to Korean Transliteration Model of Extended Markov Window”, In Proceedings of the 18th Conference on Computational linguistics, pp. 383–389.

[57] Meng H et al., (2001) “Generating Phonetic Cognates to Handle Named Entities in English-Chinese Cross-Language Spoken Document Retrieval”, ASRU '01, pp. 311-314.

[58] Oh J H, & Choi K S, (2002) “An English-Korean Transliteration Model using Pronunciation and Contextual Rules”, In Proceedings of COLING 2002, pp. 758-764.

[59] Lin W H & Chen H H, (2002) “Backward Machine Transliteration by Learning Phonetic Similarity”, In Proceedings of the 6th Conference on Natural Language Learning, pp. 1–7.

[60] Yan, Q et al., (2003) “Automatic Transliteration For Japanese-to-English Text Retrieval”, ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 353-360.

[61] Paola Virga et al.,(2003) “Transliteration of Proper Names in Cross-Lingual Information Retrieval”, Proceedings of the ACL Workshop on Multilingual and Mixed-language NER.

[62] Gao W, Wong K F, & Lam W, (2004) “Improving Transliteration with Precise Alignment of Phoneme Chunks and Using Contextual Features”, vol. 3411, Springer, Berlin, pp. 106–117.

[63] Gao W, Wong K F, & Lam W, (2004) “Phoneme-based Transliteration of Foreign Names for OOV Problem”, First IJCNLP, vol. 3248, Springer, pp. 110–119.

[64] DebasisMandal, D., Dandapat, S., Gupta, M., Banerjee, P. &Sarkar, S, (2007) “Bengali and Hindi to English CLIR Evaluation”, Cross-Language Evaluation Forum CLEF, pp. 95-102.

[65] HarshitSurana& Anil Kumar Singh, (2008) “A More Discerning and Adaptable Multilingual Transliteration Mechanism for Indian Languages”, Proceedings of the Third IJCNLP, pp. 64-71.

[66] Saha S et al., (2008) “NE Recognition in Hindi Using Maximum Entropy and Transliteration”.

[67] M L Dhore, S K Dixit and J B Karande, (2011) “Cross Language Representation for Commercial Web Applications in Context of Indian Languages using Phonetic model”, CiiT International Journal of Artificial Intelligent Systems and Machine Learning, Volume 3, No. 4, pp 174-179.

[68] M L Dhore and S K Dixit, (2011) “Development of Bilingual Application Using Machine Transliteration: A Practical Case Study”, CiiT International Journal of Artificial Intelligent Systems and Machine Learning, Volume 3, No. 13, pp 859-864.

[69] M L Dhore, S K Dixit and R M Dhore, (2012) “Hindi and Marathi to English NE Transliteration Tool using Phonology and Stress Analysis”, 24th International Conference on Computational Linguistics, Proceedings of COLING: Demonstration Papers, at III, Bombay, pp 111-118.

[70] Al-Onaizan& Knight K, (2002) “Machine Transliteration of Names in Arabic Text”, Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages.

[71] Bilac S, & Tanaka H, (2004) “Improving Back-Transliteration by Combining Information Sources”, In Proceedings of IJCNLP2004, pp. 542-547.

[72] Oh J H & Choi K S, (2005) “Machine Learning Based English-to-Korean Transliteration using Grapheme and Phoneme Information”, IEICE Transaction on Information and Systems.

[73] Oh J H & Choi K S, (2006) “An Ensemble of Transliteration Models for Information Retrieval”, Information Processing and Management, 42, 4, pp. 980–1002.

[74] Abbas Malik et al, (2009) “A Hybrid Model for Urdu Hindi Transliteration”, Proceedings of the 2009 Named Entities Workshop, ACL-IJCNLP 2009, pages 177–185.

[75] M L Dhore, S K Dixit and R M Dhore, (2012) “Optimizing Transliteration for Hindi/Marathi to English Using only Two Weights”, Proceedings of the First International Workshop on Optimization Techniques for Human Language Technology, COLING,IITB, pp 31–48,

[76] Oh J H and Ishara H, (2007) “Machine Transliteration using Multiple Transliteration Engines and Hypothesis Re-Ranking”, In Proceedings of the 11th Machine Translation Summit.

[77] SarvnazKarimi, (2008) “Machine Transliteration of Proper Names between English and Persian”, Thesis, RMIT University, Melbourne, Victoria, Australia.

SENTIMENT ANALYSIS ON PRODUCT FEATURES BASED ON LEXICON APPROACH USING NATURAL LANGUAGE PROCESSING

Ameya Yerpude, Akshay Phirke, Ayush Agrawal and Atharva Deshmukh

Department of Computer Science and Engineering, RCOEM, Nagpur, India

ABSTRACT

Sentiment analysis has played an important role in identifying what other people think and what their behavior is. Text can be used to analyze the sentiment and classified as positive, negative or neutral. Applying the sentiment analysis on the product reviews on e-market helps not only the customer but also the industry people for taking decision. The method which provides sentiment analysis about the individual product's features is discussed here. This paper presents the use of Natural Language Processing and SentiWordNet in this interesting application in Python: 1. Sentiment Analysis on Product review [Domain: Electronic]2. sentiment analysis regarding the product's feature present in the product review [Sub Domain: Mobile Phones]. It usesa lexicon based approach in which text is tokenized for calculating the sentiment analysis of the product reviews on a e-market. The first part of paper includessentiment analyzer whichclassifiesthe sentiment present in product reviews into positive, negative or neutral depending on the polarity. The second part of the paper is an extension to the first part in which the customer review's containing product's features will be segregated and then these separated reviews are classified into positive, negative and neutral using sentiment analysis. Here, mobile phones are used as the product with features as screen, processors, etc. This gives a business solution for users and industries for effective product decisions.

KEYWORDS

Sentiment Analysis, Natural Language Processing, SentiWordNet, lexicon based approach

Full Text: https://aircconline.com/ijnlc/V8N3/8319ijnlc01.pdf

Volume URL: https://airccse.org/journal/ijnlc/vol8.html

REFERENCES

[1] Allen, James, "Natural Language Understanding", Second edition (Redwood City: Benjamin/Cummings, 1995).
[2] Baxendale, P. (1958). Machine-made index for technical literature - an experiment. IBM Journal of Research Development, 2(4):354–361. [2, 3, 5]
[3] Bird Steven, Klein Ewan, Loper Edward June 2009, "Natural Language Processing with Python", Pages 16,27,79
[4] Cortez Eli, Altigran S da da Silva 2013, " Unsupervised Information Extraction by Text Segmentation", Ch 3
[5] Kumar Ela, "Artificial Intelligence", Pages 313-315
[6] Goddard Cliff Second edition 2011,"Semantic Analysis: A practical introduction ", Section 1.1- 1.5
[7] Lukaszewski Albert 2010, "MySQL for Python", Ch 1,2,3
[8] Manning Christopher D., SchützeHinrich Sixth Edition 2003,"Foundations of Statistical Natural Language Processing", Ch 4 Page no. 575 [9] Martelli Alex Second edition July 2006, "Python in a Nutshell", Pages 44,201. [10] Natural Language Toolkit, Retrieved from http://www.nltk.org
[11] Pattern 2.6, Retrieved from http://www.clips.ua.ac.be/pattern
[12] Prasad Reshma, Mary Priya Sebastian, International Journal on Natural Language Computing (IJNLC) Vol. 3, No.2, April 2014, " A survey on phrase structure learning methods for text classification"
[13] Python Language, Retrieved from https://www.python.org/ [14] Rodrigues Mário , Teixeira António , "Advanced Applications of Natural Language Processing for Performing ", Ch 1,2,4
[15] Sobin Nicholas 2011, "Syntactic Analysis: The Basics", Ch 1,2 [16] Swaroop C H, “A Byte of Python: Basics and Syntax of Python”, Ch 5,8,9,10
[17] TextBlob: Simplified Text Processing, Retrieved from http://textblob.readthedocs.org/en/dev
[18] ThanosCostantino ,"Research and Advanced Technology for Digital Libraries", Page 338-362 [19] Tosi Sandro November 2009, "Matplotlib for Python Developers", Ch 2,3
[20] Aashutosh Bhatt et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5107-5110
[21] Akshaya R. Garjeet al, International Journal of Advanced Research in Computer Science, 8(9), Nov– Dec, 2017,554-557 [22] Andrea Esuli and Fabrizio Sebastiani. 2006. SENTIWORDNET:A publicly available lexical resource for opinion mining. In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC’06), pages 417–422, Genova, IT.
[23] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1/2):1–135.

ALGORITHM FOR TEXT TO GRAPH CONVERSION AND SUMMARIZING USING NLP: A NEW APPROACH FOR BUSINESS SOLUTIONS

Prajakta Yerpude and Rashmi Jakhotiya and Manoj Chandak

Department of Computer Science and Engineering, RCOEM, Nagpur

ABSTRACT

Text can be analysed by splitting the text and extracting the keywords .These may be represented as summaries, tabular representation, graphical forms, and images. In order to provide a solution to large amount of information present in textual format led to a research of extracting the text and transforming the unstructured form to a structured format. The paper presents the importance of Natural Language Processing (NLP) and its two interesting applications in Python Language:
1. Automatic text summarization [Domain: Newspaper Articles] 2. Text to Graph Conversion [Domain: Stock news]. The main challenge in NLP is natural language understanding i.e. deriving meaning from human or natural language input which is done using regular expressions, artificial intelligence and database concepts. Automatic Summarization tool converts the newspaper articles into summary on the basis of frequency of words in the text. Text to Graph Converter takes in the input as stock article, tokenize them on various index (points and percent) and time and then tokens are mapped to graph. This paper proposes a business solution for users for effective time management.

KEYWORDS

NLP, Automatic Summarizer, Text to Graph Converter, Data Visualization, Regular Expression, Artificial Intelligence

Full Text: https://airccse.org/journal/ijnlc/papers/4415ijnlc03.pdf

Volume URL: https://airccse.org/journal/ijnlc/vol4.html

REFERENCES

[1] Allen, James, "Natural Language Understanding", Second edition (Redwood City: Benjamin/Cummings, 1995).
[2] Baxendale, P. (1958). Machine-made index for technical literature - an experiment. IBM Journal of Research Development, 2(4):354–361. [2, 3, 5]
[3] BeautifulSoup4 4.3.2, Retrieved from https://pypi.python.org/pypi/beautifulsoup4
[4] Bird Steven, Klein Ewan, Loper Edward June 2009, "Natural Language Processing with Python", Pages 16,27,79 [5] Cortez Eli, Altigran S da da Silva 2013, " Unsupervised Information Extraction by Text Segmentation", Ch 3
[6] Economic Times Archives Jan 2014-Dec 2014, Retrieved from http://economictimes.indiatimes.com/
[7] Edmundson, H. P. (1969). New methods in automatic extracting. Journal of the ACM, 16(2):264–285. [2, 3, 4]
[8] Friedl Jeffrey E.F. August 2006,"Mastering Regular Expressions", Ch 1
[9] Goddard Cliff Second edition 2011,"Semantic Analysis: A practical introduction ", Section 1.1- 1.5
[10] Kumar Ela, "Artificial Intelligence", Pages 313-315
[11] Luhn, H. P. (1958). The automatic creation of literature abstracts. IBM Journal of Research Development, 2(2):159–165. [2, 3, 6, 8]
[12] Lukaszewski Albert 2010, "MySQL for Python", Ch 1,2,3
[13] Manning Christopher D., Schütze Hinrich Sixth Edition 2003,"Foundations of Statistical Natural Language Processing", Ch 4 Page no. 575
[14] Martelli Alex Second edition July 2006, "Python in a Nutshell", Pages 44,201.
[15] Natural Language Toolkit, Retrieved from http://www.nltk.org [16] Pattern 2.6, Retrieved from http://www.clips.ua.ac.be/pattern
[17] Prasad Reshma, Mary Priya Sebastian, International Journal on Natural Language Computing (IJNLC) Vol. 3, No.2, April 2014, " A survey on phrase structure learning methods for text classification"
[18] Pressman Rodger S 6th edition," Software Engineering – A Practitioner's Approach "
[19] Python Language, Retrieved from https://www.python.org/
[20] Rodrigues Mário , Teixeira António , "Advanced Applications of Natural Language Processing for Performing ", Ch 1,2,4
[21] Stubblebine Tony, "Regular Expression Pocket Reference: Regular Expressions for Perl, Ruby, PHP, Python, C, Java and .NET " [22] Sobin Nicholas 2011, "Syntactic Analysis: The Basics", Ch 1,2
[23] Swaroop C H, "A Byte of Python: Basics and Syntax of Python", Ch 5,8,9,10
[24] TextBlob: Simplified Text Processing, Retrieved from http://textblob.readthedocs.org/en/dev
[25] Thanos Costantino ,"Research and Advanced Technology for Digital Libraries", Page 338-362
[26] Tosi Sandro November 2009, "Matplotlib for Python Developers", Ch 2,3.

AMBIGUITY RESOLUTION IN INFORMATION RETRIEVAL

RekhaJain[1], Rupal Bhargava[2] and G.N Purohit[3]

[1, 2, 3]Deparment of Computer Science, Banasthali Vidyapith, Jaipur

ABSTRACT

With the advancement of the web it is very difficult to keep up with the amplifying requirements of learning on web, to satisfy user's expectation. Users demand with the updated and accurate results. To solve the queries Search Engines use different techniques. Google the most famous search engine uses Page Ranking Algorithm. Ranking Algorithms arrange the results according to the user's needs. This paper deals with "Page Rank Algorithm". Our proposed algorithm is an extension of page rank algorithm which refines the results so that user gets what he/she expects. We have used a measure Average Precision to compare Page Rank algorithm and the proposed algorithm, and proved that our algorithm provides better results.

KEYWORDS

Information retrieval, page ranking algorithms, weighted page rank

Full Text: https://airccse.org/journal/ijnlc/papers/2113ijnlc01.pdf

Volume URL: https://airccse.org/journal/ijnlc/vol2.html

REFERENCES

[1] Ashutosh Kumar Singh, Ravi Kumar P, A Comparative study of Page Ranking Algorithms for Information Retrieval, International Journal of Electrical and Computer Engineering 4:7 2009
[2] David Hawkin,Web Search Engines, CSIRO ICT Center
[3] Eric J. Glover, Steve Lawrence, Michael D. Gordon, William P. Birmingham, C. Lee Giles, Web Search – Your Way
[4] "Information Retrival" available at http://en.wikipedia.org/wiki/Information_retrieval
[5] J. Kleinberg, "Authoritative Sources in a Hyper-Linked Environment", Journal of the ACM 46(5), pp. 604-632, 1999.
[6] J. Kleinberg, "Hubs, Authorities and Communities", ACM Computing Surveys, 31(4), 1999.
[7] L. Page, S. Brin, R. Motwani, and T. Winograd, "The Pagerank Citation Ranking: Bringing order to the Web". Technical Report, Stanford Digital Libraries SIDL-WP-1999-0120, 1999.
[8] "Precision and Recall" available at http://en.wikipedia.org/wiki/Precision_and_recall
[9] R. Cooley, B. Mobasher and J. Srivastava, "Web Minig: Information and Pattern Discovery on the World Wide Web". Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence, pp. (ICTAI'97), 1997.
[10] R. Kosala, H. Blockeel, "Web Mining Research: A Survey", SIGKDD Explorations, Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining Vol. 2, No. 1 pp 1-15, 2000.
[11] S. Brin, and L. Page, The Anatomy of a Large Scale Hypertextual Web Search Engine, Computer Network and ISDN Systems, Vol. 30, Issue 1-7, pp. 107-117,1998.
[12] W. Xing and Ali Ghorbani, "Weighted PageRank Algorithm", Proc. Of the Second Annual Conference on Communication Networks and Services Research (CNSR '0), IEEE, 2004.

HINDI AND MARATHI TO ENGLISH MACHINE TRANSLITERATION USING SVM

P H Rathod[1] , M L Dhore[2] , R M Dhore[3]

[1,2]Department of Computer Engineering, Vishwakarma Institute of Technology, Pune

[3] Pune Vidhyarthi Griha's College of Engineering and Technology, Pune

ABSTRACT

Language transliteration is one of the important areas in NLP. Transliteration is very useful for converting the named entities (NEs) written in one script to another script in NLP applications like Cross Lingual Information Retrieval (CLIR), Multilingual Voice Chat Applications and Real Time Machine Translation (MT). The most important requirement of Transliteration system is to preserve the phonetic properties of source language after the transliteration in target language. In this paper, we have proposed the named entity transliteration for Hindi to English and Marathi to English language pairs using Support Vector Machine (SVM). In the proposed approach, the source named entity is segmented into transliteration units; hence transliteration problem can be viewed as sequence labeling problem. The classification of phonetic units is done by using the polynomial kernel function of Support Vector Machine (SVM). Proposed approach uses phonetic of the source language and n-gram as two features for transliteration.

KEYWORDS

Machine Transliteration, n-gram, Support Vector Machine, Syllabification

Full Text: https://airccse.org/journal/ijnlc/papers/2413ijnlc04.pdf

Volume URL: https://airccse.org/journal/ijnlc/vol2.html

REFERENCES

[1] Padariya Nilesh, Chinnakotla Manoj, Nagesh Ajay, Damani Om P.(2008) "Evaluation of Hindi to English, Marathi to English and English to Hindi", IIT Mumbai CLIR at FIRE.

[2] Saha Sujan Kumar, Ghosh P. S, Sarkar Sudeshna and Mitra Pabitra (2008) "Named entity recognition in Hindi using maximum entropy and transliteration." [3] BIS (1991) "Indian standard code for information interchange (ISCII)", Bureau of Indian Standards, New Delhi.

[4] Joshi R K, Shroff Keyur and Mudur S P (2003) "A Phonemic code based scheme for effective processing of Indian languages", National Centre for Software Technology, Mumbai, 23rd Internationalization and Unicode Conference, Prague, Czech Republic, pp 1-17.

[5] Arbabi M, Fischthal S M, Cheng V C and Bart E (1994) "Algorithms for Arabic name transliteration", IBM Journal of Research and Development, pp 183-194.

[6] Knight Kevin and Graehl Jonathan (1997) "Machine transliteration", In proceedings of the 35th annual meetings of the Association for Computational Linguistics, pp 128-135.

[7] Stalls Bonnie Glover and Kevin Knight (1998) "Translating names and technical terms in Arabic text."

[8] Al-Onaizan Y, Knight K (2002) "Machine translation of names in Arabic text", Proceedings of the ACL conference workshop on computational approaches to Semitic languages.

[9] Jaleel Nasreen Abdul and Larkey Leah S. (2003) "Statistical transliteration for English- Arabic cross language information retrieval", In Proceedings of the 12th international conference on information and knowledge management, pp 139 – 146.

[10] Jung S. Y., Hong S., S., Paek E.(2003) "English to Korean transliteration model of extended Markov window", In Proceedings of the 18th Conference on Computational Linguistics, pp 383–389.

[11] Ganapathiraju M., Balakrishnan M., Balakrishnan N., Reddy R. (2005) "OM: One Tool for Many (Indian) Languages", ICUDL: International Conference on Universal Digital Library, Hangzhou.

[12] Malik M G A (2006) "Punjabi Machine Transliteration", Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL, pp 1137–1144.

[13] Sproat R.(2002) "Brahmi scripts, In Constraints on Spelling Changes", Fifth International Workshop on Writing Systems, Nijmegen, The Netherlands.

[14] Sproat R.(2003) "A formal computational analysis of Indic scripts", In International Symposium on Indic Scripts: Past and Future, Tokyo.

[15] Sproat R.(2004) "A computational theory of writing systems, In Constraints on Spelling Changes", Fifth International Workshop on Writing Systems, Nijmegen, The Netherlands.

[16] Kopytonenko M. , Lyytinen K. , and Krkkinen T.(2006) "Comparison of phonological representations for the grapheme-to-phoneme mapping, In Constraints on Spelling Changes", Fifth International Workshop on Writing Systems, Nijmegen, The Netherlands.

[17] Ganesh S, Harsha S, Pingali P, and Verma V (2008) "Statistical transliteration for cross language information retrieval using HMM alignment and CRF", In Proceedings of the Workshop on CLIA, Addressing the Needs of Multilingual Societies.

[18] Sumaja Sasidharan, Loganathan R, and Soman K P (2009) "English to Malayalam Transliteration Using Sequence Labeling Approach" International Journal of Recent Trends in Engineering, Vol. 1, No. 2, pp 170-172

[19] Oh Jong-Hoon, Kiyotaka Uchimoto, and Kentaro Torisawa (2009) “Machine transliteration using target-language grapheme and phoneme: Multi-engine transliteration approach”, Proceedings of the Named Entities Workshop ACL-IJCNLP Suntec, Singapore,AFNLP, pp 36–39
[20] Antony P.J, Soman K.P (2010) “Kernel Method for English to Kannada Transliteration”, Conference on Machine Learning and Cybernetics, pp 11-14
[21] Ekbal A. and Bandyopadhyay S. (2007) “A Hidden Markov Model based named entity recognition system: Bengali and Hindi as case studies”, Proceedings of 2nd International conference in Pattern Recognition and Machine Intelligence, Kolkata, India, pp 545–552.
[22] Ekbal A. and Bandyopadhyay S. (2008) “Bengali named entity recognition using support vector machine”, In Proceedings of the IJCNLP-08 Workshop on NER for South and South East Asian languages, Hyderabad, India, pp 51–58.
[23] Ekbal A. and Bandyopadhyay S. (2008), “Development of Bengali named entity tagged corpus and its use in NER system”, In Proceedings of the 6th Workshop on Asian Language Resources.
[24] Ekbal A. and Bandyopadhyay S. (2008) “A web-based Bengali news corpus for named entity recognition”, Language Resources & Evaluation, vol. 42, pp 173–182.
[25] Ekbal A. and Bandyopadhyay S.(2008) “Improving the performance of a NER system by postprocessing and voting”, In Proceedings of Joint IAPR International Workshop on Structural Syntactic and Statistical Pattern Recognition, Orlando, Florida, pp 831–841.
[26] Ekbal A. and Bandyopadhyay S.(2009) “Bengali Named Entity Recognition using Classifier Combination”, In Proceedings of Seventh International Conference on Advances in Pattern Recognition, pp 259–262.
[27] Ekbal A. and Bandyopadhyay S. (2009) “Voted NER system using appropriate unlabelled data”, In Proceedings of the Named Entities Workshop, ACL-IJCNLP.
[28] Ekbal A. and Bandyopadhyay S. (2010) “ Named entity recognition using appropriate unlabeled data, post-processing and voting”, In Informatica, Vol 34, No. 1, pp 55-76.
[29] Chinnakotla Manoj K., Damani Om P., and Satoskar Avijit (2010) “Transliteration for ResourceScarce Languages”, ACM Trans. Asian Lang. Inform,Article 14, pp 1-30.
[30] Kishorjit Nongmeikapam (2012) “Transliterated SVM Based Manipuri POS Tagging”, Advances in Computer Science and Engineering and Applications, pp 989-999
[31] K.P.Sonam, V. Ajay, R. Laganatha.(2009) “Machine Learning with SVM and Other Kernel Methods”, Machine Learning Book, PHI.
[32] Koul Omkar N. (2008) “Modern Hindi Grammar”, Dunwoody Press [33] Walambe M. R. (1990) “Marathi Shuddalekhan”, Nitin Prakashan, Pune
[34] Walambe M. R. (1990) “Marathi Vyakran”, Nitin Prakashan, Pune
[35] Dhore M L, Dixit S K and Dhore R M (2012) “Hindi and Marathi to English NE Transliteration Tool using Phonology and Stress Analysis”, 24th International Conference on Computational Linguistic,s Proceedings of COLING Demonstration Papers, at IIT Bombay, pp 111-118