

August 2025: Top 10 Read Article 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]

ASSAMESE-ENGLISH BILINGUAL MACHINE TRANSLATION

Kalyanee Kanchan Baruah¹ , Pranjal Das¹ , Abdul Hannan¹ and Shikhar Kr Sarma¹

¹Department of Information Technology, Gauhati University, Guwahati, Assam

ABSTRACT

Machine translation is the process of translating text from one language to another. In this paper, Statistical Machine Translation is done on Assamese and English language by taking their respective parallel corpus. A statistical phrase based translation toolkit Moses is used here. To develop the language model and to align the words we used two other tools IRSTLM, GIZA respectively. BLEU score is used to check our translation system performance, how good it is. A difference in BLEU scores is obtained while translating sentences from Assamese to English and vice-versa. Since Indian languages are morphologically very rich hence translation is relatively harder from English to Assamese resulting in a low BLEU score. A statistical transliteration system is also introduced with our translation system to deal basically with proper nouns, OOV (out of vocabulary) words which are not present in our corpus.

KEYWORDS

Assamese, Machine translation, Moses, Corpus, BLEU

Volume URL: <https://airccse.org/journal/ijnlc/vol3.html>

Full Text: <https://airccse.org/journal/ijnlc/papers/3314ijnlc07.pdf>

REFERENCES

- [1] Statistical Machine Translation System User Manual and Code Guide”, Available: <http://www.statmt.org//moses/manual/manual.pdf>.
- [2] F.J.Och., “GIZA++: Training of statistical translation models”, Available: <http://fjoch.com/GIZA++.html>.
- [3] “IRSTLM”, Available: <http://hlt.fbk.eu/en/irstlm>.
- [4] Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu, “Bleu: a method for Automatic Evaluation of Machine Translation”, “Proceedings of the 40th Annual Meeting of the Association for Computer Linguistics (ACL)” Philadelphia, July 2002 pp. 311-318.
- [5] N.Sharma, P.Bhatia, V.Singh, “English to Hindi Statistical Machine Translation”, June 2011.
- [6] P.F. Brown, S. De.Pietra, V. D. Pietra and R. Mercer, “The mathematics of statistical machine translation: parameter estimation”. “Journal Computational Linguistics”, vol. 10, no.2, June 1993.
- [7] “Machine Translation”, Available: <http://faculty.ksu.edu.sa/homiedan/Publications/Machine%20Translation.pdf>.
- [8] D. D. Rao, “Machine Translation A Gentle Introduction”, RESONANCE, July 1998.
- [9] “Assamese Language”, Available: http://en.wikipedia.org/wiki/Assamese_Language.
- [10] R.M.K Sinha, K Shivaraman, A Agarwal, R Jain, A Jain, “ANGLABHARATI: a multilingual machine aided translation project on translation from English to Indian languages.
- [11] Akshar Bharati, Vineet Chaitanya, Amba P. Kulkarni, Rajeev Sangal, G Umamaheshwara Rao “ Anusaaraka Machine Translation in stages.”.
- [12] Jayprasad J Hegde, Chandra Shekhar, Ritesh Shah, Sawani Bade, “MaTra A Practical Approach to Fully-Automatic Indicative English.
- [13] “Text corpus”, Available:http://en.wikipedia.org/Text_corpus.
- [14] Aneena George,“English To Malayalam Statistical Machine Translation System”.

NAMED ENTITY RECOGNITION USING HIDDEN MARKOV MODEL (HMM)

Sudha Morwal¹, Nusrat Jahan² and Deepti Chopra³

¹Associate 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).

Volume URL: <https://airccse.org/journal/ijnlc/vol1.html>

Full Text: <https://airccse.org/journal/ijnlc/papers/1412ijnlc02.pdf>

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://csejournals.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>.

GRAMMAR CHECKERS FOR NATURAL LANGUAGES: A REVIEW

Nivedita S. Bhirud¹ R.P. Bhavsar² B.V. Pawar³

¹Department 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

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

Full Text: <https://airconline.com/ijnlc/V6N4/6417ijnlc01.pdf>

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] Debel Tesfaye, “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, Prajod Shrestha, “Architectural and System Design of the Nepali Grammar Checker”, www.pan110n.net/english/.../Nepal/Microsoft%20Word%20-%20208_OK_N_400.pdf
- [7] Kinoshita, Jorge; Nascimento, Laís do; Dantas, Carlos Eduardo. “CoGrOO: a Brazilian-Portuguese Grammar Checker based on the CETENFOLHA Corpus”. Universidade da São Paulo (USP), Escola Politécnica. 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 linguistics, 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 Natural 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”. Instituto Politécnico Nacional. Centro de Investigación en Computación. 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] Lata Bopche, Gauri Dhopavkar, and Manali Kshirsagar, “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_LANGUAGES.pdf" on 30/12/2013

[19] Kenneth W. Church and Lisa F. Rau, "Commercial Applications of Natural Language Processing", COMMUNICATIONS OF THE ACM, Vol. 38, No. 11, November 1995

[20] 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

[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 Widyajaya, "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, Shah Alam, 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", International 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 Indurkha & 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.

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.

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

Full Text: <https://airconline.com/ijnlc/V8N5/8519ijnlc03.pdf>

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 LSTM CNNs-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.
- [21] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019) “BERT: Pre- training of Deep Bidirectional Transformers for Language Understanding” arxiv:1810.04805.
- [22] Yoon Kim (2014) “Convolutional Neural Networks for Sentence Classification” In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* pp 1746- 1751.
- [23] Siwei Lai, Liheng Xu, Kang Liu, Jun Zhao (2005) “Recurrent Convolutional Neural Networks for Text Classification” In *Proceedings of Conference of the Association for the Advancement of Artificial Intelligence* Vol. 333 pp 2267-2273.
- [24] Takeru Miyato, Andrew M. Dai, and Ian Goodfellow (2017) “Adversarial Training Methods for Semi-supervised Text Classification” In *ICLR 2017*.
- [25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin (2017) “Attention Is All You Need” In 31st Conference on Neural Information Processing Systems (NIPS’ 2017), Long Beach, CA, USA.
- [26] Zoubin Ghahramani, and Michael I. Jordan (1997) “Factorial Hidden Markov Model” *Machine Learning* Vol. 29 No. 2-3, pp 245-273.
- [27] Andrew McCallum, Dayne Freitag, and Fernando Pereira (2000) “Maximum Entropy Markov Models for Information Extraction and Segmentation” In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML’00)* pp 591-598.
- [28] John Lafferty, Andrew McCallum, and Fernando Pereira (2001) “Conditional Random Fields: Probabilistic Models for Segmenting and Labelling Sequence Data” In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML’01)* Vol. 3 No. 2,

pp 282-289.

- [29] Zhiheng Huang, Wei Xu, and Kai Yu (2015) “Bidirectional LSTM-CRF Models for Sequence Tagging” arXiv preprint arXiv:1508.01991, 2015.
- [30] Zhenyu Jiao, Shuqi Sun, and Ke Sun (2018) “Chinese Lexical Analysis with Deep Bi-GRU-CRF Network” arXiv preprint arXiv:1807.01882.
- [31] Emma Strubell, Patrick Verga, David Belanger, and Andrew McCallum (2017) “Fast and Accurate Entity Recognition with Iterated Dilated Convolutions” In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing arXiv preprint arXiv: 1702.02098.
- [32] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio (2014) “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling” arXiv preprint arXiv:1412.3555, 2014.
- [33] Chuanhai Dong, Jiajun Zhang, Chengqing Zong, Masanori Hattori, and Hui Di (2016) “CharacterBased LSTM-CRF with Radical-Level Features for Chinese Named Entity Recognition” International Conference on Computer Processing of Oriental Languages Springer International Publishing pp 239-250.
- [34] Sanyal, S., Hazra, S., Adhikary, S., & Ghosh, N. (2017) “Resume Parser with Natural Language Processing” International Journal of Engineering Science, 4484.

SURVEY OF MACHINE TRANSLATION SYSTEMS IN INDIA

G V Garje¹ and G K Kharate²

**¹Department 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

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

Full Text: <https://airccse.org/journal/ijnlc/papers/2513ijnlc04.pdf>

REFERENCES

- [1] Sitender & Seema Bawa, (2012) "Survey of Indian Machine Translation Systems", International Journal Computer Science and Technology, 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.

SURVEY ON MACHINE transliteration AND MACHINE LEARNING MODELS

M L Dhore¹, R M Dhore² and P H Rathod³

^{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

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

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

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] Debasis Mandal, 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] Harshit Surana & 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.

MACHINE TRANSLATION DEVELOPMENT FOR INDIAN LANGUAGES AND ITS APPROACHES

Amruta Godase¹ and Sharvari Govilkar²

¹Department of Information Technology (AI & Robotics), PIIT, Mumbai University, India

²Department 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

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

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

REFERENCES

- [1] 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.

- [2] Sudip Naskar & Shivaji Bandyopadhyay, (2005) "Use of Machine Translation in India: Current status" AAMT Journal, pp. 25-31
- [3] Lata Gore & Nishigandha Patil, (2002) "English to Hindi - Translation System", In proceedings of Symposium on Translation Support Systems. IIT Kanpur. pp. 178-184
- [4] 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
- [5] Choudhary,A. Singh, M. (2009) "GB theory based Hindi to english translation system", Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference PP.293 – 297
- [6] Ruchika A. Sinhal, Kapil O. Gupta (2014) "A Pure EBMT Approach for English to Hindi Sentence Translation System" I.J.Modern Education and Computer Science, 2014, 7, 1-8 Published Online July 2014 in MECS (<http://www.meecs-press.org/>)
- [7] Shachi Dave, Jignashu Parikh And Pushpak Bhattacharyya Department of Computer Science and Engineering, Indian Institute of Technology, Mumbai, India, "Interlingua based English-Hindi Machine Translation system and Language Divergence."
- [8] M.L.Dhore & S.X.Dixit (2011) "English to Devnagari Translation for UI Labels of Commercial web based Interactive Applications" International Journal of Computer Applications (0975-8887) Volume 35-No.10 , December 2011
- [9] Devika Pishartoy, Priya, Sayli Wandkar (2012) "Exteneding capabilities of English to Marathi machine Translator" , International journal of Computer Science Issues, Vol.9, Issues 3, No. 3, May 2012 ISSN (Online): 1694-0814
- [10] Abhay A, Anita G, Purnima T, Prajakta G (2013), "Rule based English to Marathi translation of Assertive sentence" , International Journal of Scientific & Engineering Research, Volume 4, Issues 5, May- 2013 ISSN 2259-5518 pp. 1754-1756
- [11] Krushnadeo B, Vinod W, S.V.Phulari, B.S.Kankate (2014), "A novel approach for Interlingual example-based translation of English to Marathi", International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 3, March 2014)
- [12] G.V.Gajre, G.Kharate, H. Kulkarni (2014), "Transmuter: An approach to Rule-based English to Marathi Machine Translation" , International Journal of Computer Applications (0975 – 8887) Volume 98 – No.21, July 2014
- [13] Vimal Mishra, R.B.Mishra Research Scholar, Department of Computer Engineering, Institute of Technology, Banaras, Hindu University, (IT-BHU), Varanasi-221005, U.P., India, "ANN and Rule Based Model for English to Sanskrit Machine Translation"
- [14] Ms.Vaishali.M.Barkade, Prof. Prakash R. Devale , Dr. Suhas H. Patil, "ENGLISH TO SANSKRIT MACHINE TRANSLATOR LEXICAL PARSER AND SEMANTIC MAPPER", National Conference On "Information and Communication Technology" NCICT-10
- [15] Promila Bahadur ,D.S.Chauhan, , A.K.Jain , Indian Institute of Technology Kanpur, India, "EtranSA Complete Framework for English To Sanskrit Machine Translation", IJACSA Special Issue on Selected Papers from International Conference & Workshop On Emerging Trends In Technology 2012 pp. 52-59
- [16] Sarita G. Rathod, Shanta Sondur, Information Technology Department, VESIT, Mumbai University, Maharashtra, India, "English to Sanskrit Translator and Synthesizer", International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN

2250-2459, ISO 9001:2008 Certified Journal, Volume 2, Issue 12, December 2012)

[17] Sandeep R. Warhade, Prakash R. Devale ,Suhass H. Patil , “ English-to-Sanskrit Statistical Machine Translation with Ubiquitous Application” , International Journal of Computer Applications (0975 – 8887) Volume 51– No.1, August 2012

[18] Pankaj Upadhyay, Umesh Chandra Jaiswal, Kumar Ashish, “ TranSish: Translator from Sanskrit to English-A Rule based Machine Translation” , International Journal of Current Engineering and Technology E-ISSN 2277 – 4106, P-ISSN 2347 - 5161 ©2014 INPRESSCO®, All Rights Reserved Available at <http://inpressco.com/category/ijcet>

[19] 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

[20] 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.

[21] 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.

[22] 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.

[23] Sanjay Chatterji, Praveen Sonare, Sudeshna Sarkar, and Anupam Basu, 2011, “Lattice Based Lexical Transfer in Bengali Hindi Machine Translation Framework”, In Proceedings of ICON-2011: 9th International Conference on Natural Language Processing, Macmillan Publishers, India.

[24] Shibli A, Humayun K, Musfique A, K.M.Noman, 2013, “English To Bengali Machine Translation Using Context Free Grammars”, International journal of Computer Science Issues, vol.10, Issues 3, No.2, May 2013 ISSN: 1694-0814 pp. 144-153

[25] Harjinder Kaur, Dr. Vijay Laxmi, 2013 “A Web Based English to Punjabi MT System for News Headlines” In International Journal of Advanced Research in Computer Science and Software Engineering 3(6), June - 2013, pp. 1092-1094

[26] Pankaj Kumar and Er.Vinod Kumar, 2013, “Statistical Machine Translation Based Punjabi to English Transliteration System for Proper Nouns”, In International Journal of Application or Innovation in Engineering & Management (IJAEM) Volume 2, Issue 8, August 2013 ISSN 2319 – 4847, pp . 318- 320

[27] Kamaljeet Kaur Batra and G S Lehal, 2010, “Rule Based Machine Translation of Noun Phrases from Punjabi to English”, In IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010 ISSN (Online): 1694-0814, pp. 409-413

[28] Vishal Goyal and Gurpreet Singh Lehal , 2010, “Web Based Hindi to Punjabi Machine Translation System”, JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, VOL. 2, NO. 2, MAY 2010, pp.148-151

[29] Vishal Goyal and Gurpreet Singh Lehal , 2011, “HINDI TO PUNJABI MACHINE TRANSLATION SYSTEM”, Proceedings of the ACL-HLT 2011 System Demonstrations, pages 1–6, Portland, Oregon, USA, 21 June 2011. Association for Computational Linguistics

[30] Naila Ata, Bushra Jawaid, Amir Kamran “Rule Based English to Urdu Machine Translation”

[31] Aasim Ali, Arshad , Hussain and Muhammad Kamran Malik, 2013, “Model for English-Urdu Statistical Machine Translation”, World Applied Sciences Journal 24 (10): 1362-1367, 2013 ISSN 1818-4952 © IDOSI Publications, 2013 DOI: 10.5829/idosi.wasj.2013.24.10.760

[32] Asad Habib, Asad Abdul Malik ,Kohat University of Science and Technology, Kohat,

Pakistan , 2013, “Urdu to English Machine Translation using Bilingual Evaluation Understudy” International Journal of Computer Applications (0975 – 8887) Volume 82 – No 7, November 2013, pp. 5-12

[33] R. Mahesh K. Sinha Department of Computer Science & Engineering, Indian Institute of Technology, Kanpur, India, “Developing English-Urdu Machine Translation Via Hindi”

[34] Nadeem khan, Waqas Anwar, Nadir Durrani, 2013, “English to Urdu Hierarchical Phrase based statistical Machine translation”The 4th Workshop on South and Southeast Asian NLP (WSSANLP), International Joint Conference on Natural Language Processing, pages 72–76

[35] Nayyara Karamat December, 2006, “VERB TRANSFER FOR ENGLISH TO URDU MACHINE TRANSLATION (Using Lexical Functional Grammar (LFG)) – MS Thesis” at the National University of Computer & Emerging Sciences

[36] SHAHNAWAZ, R. B. MISHRA, 2011, “Translation Rules and ANN based model for English to Urdu Machine Translation”, INFOCOMP, v. 10, no. 3, p. 36-47, September of 2011

[37] Nadir Durrani ,Hassan Sajjad ,Alexander Fraser & Helmut Schmid, Institute for Natural Language Processing University of Stuttgart, “Hindi to Urdu Machine Translation Through Transliteration”

[38] Aasim Ali, Shahid, Muhammad Malik, 2010, “ Development of Parallel Corpus and English to Urdu Statistical Machine Translation”, International Journal of Engineering & Technology IJET-IJENS Vol:10 No:05, pp.31-33

[39] Mary Priya Sebastian, Sheena Kurian K, G. Santhosh Kumar, 2009 “ English to Malayalam Translation: A Statistical Approach”

[40] Nithya B, Shibily Joseph, 2013 “ A Hybrid Approach to English to Malayalam Machine Translation”, International Journal of Computer Applications (0975 – 8887) Volume 81 – No.8, November 2013, pp. 11-15

[41] Latha R Nair, David Peter & Renjith P Ravindran, 2012, “Design and Development of a Malayalam to English Translator-A Transfer Based Approach”, International Journal of Computational Linguistics (IJCL), Volume (3) : Issue (1) : 2012

[42] Anju E S, Manoj Kumar K V, 2014, “ Malayalam To English Machine Translation: An EBMT System”, IOSR Journal of Engineering (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 2278-8719 Vol. 04, Issue 01 (January. 2014), ||V1|| PP 18-23 [43] Murthy. K, (2002) “MAT: A Machine Assisted Translation System”, In Proceedings of Symposium on Translation Support System(STRANS-2002), IIT Kanpur. pp. 134-139

[44] Mr. Chethan Chandra S Basavaraddi , Dr. H. L. Shashirekha, 2014, “A Typical Machine Translation System for English to Kannada” , International Journal of Scientific & Engineering Research, Volume 5, Issue 4, April-2014 ISSN 2229-5518

[45] 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)

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

[47] http://www.iiit.net/ltrc/Anusaaraka/anu_home.html

[48] <http://www.cdac.in/html/aai/mantra.asp>

[49] http://www.academia.edu/7986160/Machine_Translation_of_Bilingual_Hindi-English_Hinglish_Text

[50] <http://www.cfilt.iitb.ac.in/machine-translation/eng-hindi-mt>

[51] <http://www.ncst.ernet.in/matra/>

- [52] <http://ebmt.serc.iisc.ernet.in/mt/login.html>
- [53] <http://shakti.iiit.net>
- [54] <https://translate.google.co.in>
- [55] www.cflit.iitb.ac.in/indic-translator/
- [56] <http://sampark.iiit.ac.in/sampark/>.

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

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

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

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.

NERHMM: A TOOL FOR NAMED ENTITY RECOGNITION BASED ON HIDDEN MARKOV MODEL

Sudha Morwa¹ and Deepti Chopra

¹Department of Computer Engineering, Banasthali Vidyapith, Jaipur (Raj.), INDIA

ABSTRACT

Named Entity Recognition (NER) is considered as one of the key task in the field of Information Retrieval. NER is the method of recognizing Named Entities (NEs) in a corpus and then organizing these NEs into diverse classes of NEs e.g. Name of Location, Person, Organization, Quantity, Time, Percentage etc. Today, there is a great need to develop a tool for NER, since the existing tools are of limited scope. In this paper, we would discuss the functionality and features of our tool of NER with some experimental results.

KEYWORDS

HMM, NER, F-Measure, Accuracy, NEs

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

Full Text: <https://airccse.org/journal/ijnlc/papers/2213ijnlc04.pdf>

REFERENCES

- [1] Animesh Nayan,, B. Ravi Kiran Rao, Pawandeep Singh,Sudip Sanyal and Ratna Sanya “Named Entity Recognition for Indian Languages” .In Proceedings of the IJCNLP-08 Workshop on NER for South and South East Asian Languages ,Hyderabad (India) pp. 97–104, 2008.
- [2] Asif Ekbal and Sivaji Bandyopadhyay. “Named Entity Recognition using Support Vector Machine: A Language Independent Approach” International Journal of Electrical and Electronics Engineering 4:2 2010.
- [3] Asif Ekbal, Rejwanul Haque, Amitava Das, Venkateswarlu Poka and Sivaji Bandyopadhyay

“Language Independent Named Entity Recognition in Indian Languages” .In Proceedings of the IJCNLP-08 Workshop on NER for South and South East Asian Languages, pages 33–40,Hyderabad, India, January 2008.

[4] Asif Ekbal and Sivaji Bandyopadhyay 2008 “ Bengali Named Entity Recognition using Support Vector Machine” Proceedings of the IJCNLP-08 Workshop on NER for South and South East Asian Languages, pages 51–58, Hyderabad, India, January 2008.

[5] B. Sasidhar, P. M. Yohan, Dr. A. Vinaya Babu³, Dr. A. Govardhan. “A Survey on Named Entity Recognition in Indian Languages with particular reference to Telugu” IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011

[6] Darvinder kaur, Vishal Gupta. “A survey of Named Entity Recognition in English and other Indian Languages” . IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 6, November 2010.

[7] Georgios Paliouras, Vangelis Karkaletsis, Georgios Petasis and Constantine D. Spyropoulos.”Learning Decision Trees for Named-Entity Recognition and Classification”

[8] G.V.S.RAJU, B.SRINIVASU, Dr.S.VISWANADHA RAJU, 4K.S.M.V.KUMAR “Named Entity Recognition for Telugu Using Maximum Entropy Model”

[9] Hideki Isozaki “Japanese Named Entity Recognition based on a Simple Rule Generator and Decision Tree Learning” .Available at:<http://acl.ldc.upenn.edu/acl2001/MAIN/ISOZAKI.PDF>

[10] James Mayfield and Paul McNamee and Christine Piatko “Named Entity Recognition using Hundreds of Thousands of Features” .Available at: <http://acl.ldc.upenn.edu/W/W03/W03-0429.pdf>

[11] Kamaldeep Kaur, Vishal Gupta.” Name Entity Recognition for Punjabi Language” IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 .Vol. 2, No.3, June 2012

[12] Lawrence R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", In Proceedings of the IEEE, 77 (2), p. 257-286February 1989.Available at: <http://www.cs.ubc.ca/~murphyk/Bayes/rabiner.pdf>

[13] “Padmaja Sharma, Utpal Sharma, Jugal Kalita.”Named Entity Recognition: A Survey for the Indian Languages. ” . (LANGUAGE IN INDIA. Strength for Today and Bright Hope for Tomorrow .Volume 11: 5 May 2011 ISSN 1930-2940)AvailableAt:<http://www.languageinindia.com/may2011/v11i5may2011.pdf>

[14] Praveen Kumar P and Ravi Kiran V” A Hybrid Named Entity Recognition System for South Asian Languages”. Available at-<http://www.aclweb.org/anthology-new/I/I08/I08-5012.pdf>

[15] S. Pandian, K. A. Pavithra, and T. Geetha, “Hybrid Three-stage Named Entity Recognizer for Tamil,” INFOS2008, March Cairo-Egypt. Available at: http://infos2008.fci.cu.edu.eg/infos/NLP_08_P045-052.pdf

[16] 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.Available at: <http://cscjournals.org/csc/manuscript/Journals/IJCL/volume2/Issue1/IJCL-19.pdf>

[17] Sujan Kumar Saha, Sudeshna Sarkar, Pabitra Mitra “Gazetteer Preparation for Named Entity Recognition in Indian Languages”.

[18] Sujan Kumar Saha Sanjay Chatterji Sandipan Dandapat. “A Hybrid Approach for Named Entity Recognition in Indian Languages”

[19] S. Biswas, M. K. Mishra, Sitanath_biswas, S. Acharya, S. Mohanty “A Two Stage Language Independent Named Entity Recognition for Indian Languages” (IJCSIT) International Journal of

Computer Science and Information Technologies, Vol. 1 (4), 2010, 285-289.

[20] Vishal Gupta, Gurpreet Singh Lehal “Named Entity Recognition for Punjabi Language Text Summarization” International Journal of Computer Applications (0975 – 8887) Vpl.33 No.3, Nov.

2011 [21] NLTK Toolkit. Available at: <http://nltk.org/>

[22] Scikit-learn tool. Available at: <http://scikit-learn.org/stable/>

[23] Stanford Named Entity Recognizer. Available at: <http://nlp.stanford.edu/software/CRF-NER.shtm>.

MACHINE LEARNING ALGORITHMS FOR MYANMAR NEWS CLASSIFICATION

Khin Thandar Nwet

**Department of Computer Science, University of Information Technology, Yangon,
Myanmar**

ABSTRACT

Text classification is a very important research area in machine learning. Artificial Intelligence is

reshaping text classification techniques to better acquire knowledge. In spite of the growth and spread of AI in text mining research for various languages such as English, Japanese, Chinese, etc., its role with respect to Myanmar text is not well understood yet. The aim of this paper is comparative study of machine learning algorithms such as Naïve Bayes (NB), k-nearest neighbours (KNN), support vector machine (SVM) algorithms for Myanmar Language News classification. There is no comparative study of machine learning algorithms in Myanmar News. The news is classified into one of four categories (political, Business, Entertainment and Sport). Dataset is collected from 12,000 documents belongs to 4 categories. Well-known algorithms are applied on collected Myanmar language News dataset from websites. The goal of text classification is to classify documents into a certain number of pre-defined categories. News corpus is used for training and testing purpose of the classifier. Feature selection method, chi square algorithm achieves comparable performance across a number of classifiers. In this paper, the experimental results also show support vector machine is better accuracy to other classification algorithms employed in this research. Due to Myanmar Language is complex, it is more important to study and understand the nature of data before proceeding into mining.

KEYWORDS

Text Classification, Machine Learning, Feature Extraction

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

Full Text: <https://aircconline.com/ijnlc/V8N4/8419ijnlc02.pdf>

REFERENCES

- [1] Y. Ko and J. Seo, “Automatic Text Categorization by Unsupervised Learning”, 2000.
- [2] S M Kamruzzaman and C. Mofizur Rahman, “Text Categorization using Association Rule and Naïve Bayes Classifier”, 2001 .
- [3] E. Frank and R. R. Bouckaert, “Naïve Bayes for Text Classification with Unbalanced Classes”.
- [4] C. Chan, A. Sun and Ee-Peng Lim, “Automated Online News Classification with Personalization”, December 2001.
- [5] M. IKONOMAKIS, S. KOTSIANTIS and V. TAMPAKAS, “Text Classification Using

Machine Learning Techniques”, August 2005

[6] S.Niharika , V.Sneha Latha and D.R.Lavanya, “A Survey on Text Categorization”, 2012.

[7] H. Shimodaira, “Text Classification Using Naive Bayes”, February 2015.

[8] Anuradha Purohit, Deepika Atre, Payal Jaswani and Priyanshi Asawara, “Text Classification in Data Mining”, June 2015.

[9] Ahmed Faraz, “An Elaboration of Text Categorization And Automatic Text Classification Through Mathematical and Graphical Modeling”, June 2015.

[10] B. Trstenjak, S. Mikac, D. Donko, “KNN with TF-IDF Based Framework for Text Categorization”, 24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2013.

[11] K. Mikawa, T. Ishidat and M.Goto, “A Proposal of Extended Cosine Measure for Distance Metric Learning in Text Classification”, 2011.

[12] Y. Yang and X. Liu, “A re-examination of text categorization methods”. In proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR’99),42-49, 1999

[13] R. Al-Shalabi, G. Kanaan and Manaf H. Gharaibeh, “Arabic Text Categorization Using kNN Algorithm”, The International Arab Journal of Information Technology, Vol 4, P 5-7,2015.

[14] Meenakshi and Swati Singla, “Review Paper on Text Categorization Techniques”, SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE) – EFES April 2015

[15] K. T. D. Nwet, A.H.Khine and K.M.Soe “Automatic Myanmar News Classification”, Fifteenth International Conference On Computer Applications, 2017.

[16] M. Thangaraj and M. Sivakami, “Text classification Techniques: A Literature Review”, Journal of Information, Knowledge and Management, Vol 13, 2018

[17] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. Journal of Machine Learning Research, 12(Oct):2825{2830, 2011.