# November 2026:Top10 Download 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]

# Named Entity Recognition using Hidden Markov Model (HMM)

Sudha Morwal 1 , Nusrat Jahan 2 and Deepti Chopra 3

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)

**Full Text**: **https://airccse.org/journal/ijnlc/vol1.html**

**Volume Url:** **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://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.

# AN IMPROVED APRIORI ALGORITHM FOR ASSOCIATION RULES

Mohammed Al-Maolegi1 , Bassam Arkok2

Computer Science, Jordan University of Science and Technology, Irbid, Jordan

## Abstract

There are several mining algorithms of association rules. One of the most popular algorithms is Apriori that is used to extract frequent itemsets from large database and getting the association rule for discovering the knowledge. Based on this algorithm, this paper indicates the limitation of the original Apriori algorithm of wasting time for scanning the whole database searching on the frequent itemsets, and presents an improvement on Apriori by reducing that wasted time depending on scanning only some transactions. The paper shows by experimental results with several groups of transactions, and with several values of minimum support that applied on the original Apriori and our implemented improved Apriori that our improved Apriori reduces the time consumed by 67.38% in comparison with the original Apriori, and makes the Apriori algorithm more efficient and less time consuming.

## Keywords

Apriori, Improved Apriori, Frequent itemset, Support, Candidate itemset, Time consuming.

**Full Text**: https://airccse.org/journal/ijnlc/papers/3114ijnlc03.pdf

**Volume Url:** https://airccse.org/journal/ijnlc/vol3.html

## REFERENCES


[1] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," Knowledge and Information Systems, vol. 14, no. 1, pp. 1–37, Dec. 2007.

[2] S. Rao, R. Gupta, "Implementing Improved Algorithm Over APRIORI Data Mining Association Rule Algorithm", International Journal of Computer Science And Technology, pp. 489-493, Mar. 2012

[3] H. H. O. Nasereddin, "Stream data mining," International Journal of Web Applications, vol. 1, no. 4, pp. 183–190, 2009.

[4] F. Crespo and R. Weber, "A methodology for dynamic data mining based on fuzzy clustering," Fuzzy Sets and Systems, vol. 150, no. 2, pp. 267–284, Mar. 2005.

[5] R. Srikant, "Fast algorithms for mining association rules and sequential patterns," UNIVERSITY OF WISCONSIN, 1996.

[6] J. Han, M. Kamber,"Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, Book, 2000.

[7] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI magazine, vol. 17, no. 3, p. 37, 1996.

[8] F. H. AL-Zawaidah, Y. H. Jbara, and A. L. Marwan, "An Improved Algorithm for Mining Association Rules in Large Databases," Vol. 1, No. 7, 311-316, 2011

[9] T. C. Corporation, "Introduction to Data Miningand Knowledge Discovery", Two Crows Corporation, Book, 1999.

[10] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in ACM SIGMOD Record, vol. 22, pp. 207–216, 1993

[11] M. Halkidi, "Quality assessment and uncertainty handling in data mining process," in Proc, EDBT Conference, Konstanz, Germany, 2000.

# Rag-Fusion: A New Take on Retrieval Augmented Generation

Zackary Rackauckas

Power and Sensor Systems, Infineon Technologies, San Jose, California

## Abstract

Infineon has identified a need for engineers, account managers, and customers to rapidly obtain product information. This problem is traditionally addressed with retrieval-augmented generation (RAG) chatbots, but in this study, I evaluated the use of the newly popularized RAG-Fusion method. RAG-Fusion combines RAG and reciprocal rank fusion (RRF) by generating multiple queries, reranking them with reciprocal scores and fusing the documents and scores. Through manually evaluating answers on accuracy, relevance, and comprehensiveness, I found that RAG-Fusion was able to provide accurate and comprehensive answers due to the generated queries contextualizing the original query from various perspectives. However, some answers strayed off topic when the generated queries' relevance to the original query is insufficient. This research marks significant progress in artificial intelligence (AI) and natural language processing (NLP) applications and demonstrates transformations in a global and multiindustry context.

## Keywords

Chatbot, Retrieval-augmented Generation, Reciprocal Rank Fusion, Natural Language Processing

**Full Text**: https://aircconline.com/ijnlc/V13N1/13124ijnlc03.pdf

**Volume Url:** https://airccse.org/journal/ijnlc/vol13.html

## REFERENCES


[1] w. yu, "retrieval-augmented generation across heterogeneous knowledge," (2022)proceedings of the 2022 conference of the north american chapter of the association for computational linguistics: human language technologies: student research workshop, pp. 52-58, doi: 10.18653/v1/2022.naaclsrw.7

[2] h. naveed, u. a. khan, s. qiu, m. saqib, s. anwar, m. usman, et al., (2023) "a comprehensive overview of large language models," doi: 10.48550/arxiv.2307.06435.

[3] z. levonian, c. li, w. zhu, a. gade, o. henkel, m. postle, et al., (2023) "retrieval-augmented generation to improve math question-answering: trade-offs between groundedness and human preference," neurips'23 workshop on generative ai for education, doi: 10.48550/arxiv.2310.03184.

[4] p. lewis, e. perez, a. piktus, f. petroni, v. karpukhin, n. goyal, et al., (2020) "retrieval-augmented generation for knowledge-intensive nlp tasks," neurips,doi: 10.48550/arxiv.2005.11401.

[5] k. shuster, s. poff, m. chen, d. kiela, j. weston, (2021) "retrieval augmentation reduces hallucination in conversation,"doi: 10.48550/arxiv.2104.07567

[6] m. glass, g. rossiello, m. f. m. chowdhury, a. r. naik, p. cai, and a. gliozzo, (2022) "re2g: retrieve, rerank, generate," naacl, doi: 10.48550/arxiv.2207.06300

[7] g. v. cormack, c. l. a. clarke, and s. buettcher, (2009) "reciprocal rank fusion outperforms condorcet and individual rank learning methods," sigir '09: proceedings of the 32nd international acm sigir conference on research and development in information retrieval, pp. 758-759, doi: 10.1145/1571941.1572114.

[8] s. bruch, s. gai, and a. ingber, (2023) "an analysis of fusion functions for hybrid retrieval," acm transactions on information systems, vol. 42, no. 20, pp. 1-35, doi: 10.1145/3596512.

[9] a. h. raudaschl, (2023) "forget rag, the future is rag-fusion,"towards data science.

[10] "infineon developer community," [online]. available: https://community.infineon.com/.

[11] "xensiv™ – sensing the worldsensor solutions for automotive, industrial, consumer and iot applications," [online].available: https://www.infineon.com/cms/en/product/sensor/memsmicrophones/.

[12] "power mosfet - infineon technologies," [online]. available: https://www.infineon.com/cms/en/product/power/mosfet/.

[13] a. h. raudaschl, github - rag-fusion: the next frontier of search technology. [online]. available: https://github.com/raudaschl/rag-fusion.

[14] camara, and f. r. barrera, github – ragelo. [online] available: https://github.com/zetaalphavector/ragelo.

# Sentiment Analysis for Modern Standard Arabic and Colloquial

Hossam S. Ibrahim 1 , Sherif M. Abdou2 and Mervat Gheith

1 Computer Science Department, Institute of statistical studies and research (ISSR), Cairo University, EGYPT

2 Information Technology Department, Faculty of Computers and information Cairo University, EGYPT

## Abstract

The rise of social media such as blogs and social networks has fueled interest in sentiment analysis. With the proliferation of reviews, ratings, recommendations and other forms of online expression, online opinion has turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities and manage their reputations, therefore many are now looking to the field of sentiment analysis. In this paper, we present a feature-based sentence level approach for Arabic sentiment analysis. Our approach is using Arabic idioms/saying phrases lexicon as a key importance for improving the detection of the sentiment polarity in Arabic sentences as well as a number of novels and rich set of linguistically motivated features (contextual Intensifiers, contextual Shifter and negation handling), syntactic features for conflicting phrases which enhance the sentiment classification accuracy. Furthermore, we introduce an automatic expandable wide coverage polarity lexicon of Arabic sentiment words. The lexicon is built with gold-standard sentiment words as a seed which is manually collected and annotated and it expands and detects the sentiment orientation automatically of new sentiment words using synset aggregation technique and free online Arabic lexicons and thesauruses. Our data focus on modern standard Arabic (MSA) and Egyptian dialectal Arabic tweets and microblogs (hotel reservation, product reviews, etc.). The experimental results using our resources and techniques with SVM classifier indicate high performance levels, with accuracies of over 95%..

## Keywords

Sentiment Analysis, opinion mining, social network, sentiment lexicon, modern standard Arabic, colloquial, natural language processing

**Full Text**: **https://airccse.org/journal/ijnlc/papers/4215ijnlc07.pdf**

**Volume Url: https://airccse.org/journal/ijnlc/vol4.html**

## REFERENCES


[1] A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," in Collaboration Technologies and Systems (CTS) International Conference, Denver, CO, USA, 2012, pp. 546-550.

[2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002, pp. 79–86.

[3] D. Davidiv, O. Tsur, and A. Rappoport, "Enhanced Sentiment Learning Using Twitter Hash-tags and Smileys," in Proceedings of the 23rd International Conference on Computational Linguistics (Coling2010), Beijing, China, 2010, pp. 241–249.

[4] L. Barbosa and J. Feng, "Robust Sentiment Detection on Twitter from Biased and Noisy Data " in Proceedings of the 23rd International Conference on Computational Linguistics (Coling), 2010.

[5] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," in Proceedings of the 40th Annual Meeting on Association for Computational Linguistics ACL '02, Stroudsburg, PA, USA, 2002, pp. 417-424.

[6] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in Proceedings of the Joint ACL / EACL Conference, 1997, pp. 174–181.

[7] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval vol. 2, pp. 1–135, 2008.

[8] M. Hu and B. Liu, "Mining and summarizing customer reviews " in Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2004, pp. 168–177.

[9] B. Liu, "Sentiment Analysis and Subjectivity," in Handbook of Natural Language Processing, Second ed: CRC Press, Taylor and Francis Group, 2010.

[10] P. Alexander and P. Patrick, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining " in Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association ELRA, Valletta, Malta, 2010.

[11] C. Scheible and H. Schütze, "Bootstrapping Sentiment Labels For Unannotated Documents With Polarity PageRank," in Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012), Istambol-Turki, 2012.

[12] C. Manning and D. Klein, "Optimization, maxent models, and conditional estimation without magic," in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, 2003, p. 8.

[13] A. Abbasi, H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature

Selection for Opinion Classification in Web Forums," ACM Transactions on Information Systems, vol. 26, 2008.

[14] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003.

[15] E. Riloff, J. Wiebe, and T. Wilson, "Learning subjective nouns using extraction pattern bootstrapping," in Proceedings of the Conference on Natural Language Learning (CoNLL), 2003, pp. 25–32.

[16] M. Abdul-Mageed and M. Diab, "Subjectivity and Sentiment Annotation of Modern Standard Arabic Newswire," in Proceedings of the Fifth Law Workshop (LAW V), Association for Computational Linguistics, Portland, Oregon, 2011, pp. 110–118. [17] M. Abdul-Mageed, M. Diab, and M. Korayem, "Subjectivity and sentiment analysis of modern standard Arabic," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011.

[17] C. Scheible and H. Schütze, "Bootstrapping Sentiment Labels For Unannotated Documents With Polarity PageRank," in Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012), Istambol-Turki, 2012.

[18] C. Manning and D. Klein, "Optimization, maxent models, and conditional estimation without magic," in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, 2003, p. 8.

[19] A. Abbasi, H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," ACM Transactions on Information Systems, vol. 26, 2008.

[20] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003.

[21] E. Riloff, J. Wiebe, and T. Wilson, "Learning subjective nouns using extraction pattern bootstrapping," in Proceedings of the Conference on Natural Language Learning (CoNLL), 2003, pp. 25–32.

[22] M. Abdul-Mageed and M. Diab, "Subjectivity and Sentiment Annotation of Modern Standard Arabic Newswire," in Proceedings of the Fifth Law Workshop (LAW V), Association for Computational Linguistics, Portland, Oregon, 2011, pp. 110–118.

[23] M. Abdul-Mageed, M. Diab, and M. Korayem, "Subjectivity and sentiment analysis of modern standard Arabic," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011.

# OPINION MINING AND ANALYSIS: A SURVEY

Arti Buche 1, Dr. M. B. Chandak 2, Akshay Zadgaonkar 3

Computer Science Department, RCOEM , India

Zadgaonkar Software Pvt. Ltd. India

## Abstract

The current research is focusing on the area of Opinion Mining also called as sentiment analysis due to sheer volume of opinion rich web resources such as discussion forums, review sites and blogs are available in digital form. One important problem in sentiment analysis of product reviews is to produce summary of opinions based on product features. We have surveyed and analyzed in this paper, various techniques that have been developed for the key tasks of opinion mining. We have provided an overall picture of what is involved in developing a software system for opinion mining on the basis of our survey and analysis.

## Keywords

Opinion Mining, Sentiment Analysis, product reviews, machine learning

**Full Text**: **https://airccse.org/journal/ijnlc/papers/2313ijnlc04.pdf**

**Volume Url: https://airccse.org/journal/ijnlc/vol2.html**

## REFERENCES


[1] G.Vinodhini and RM.Chandrasekaran, Sentiment Analysis and Opinion Mining: A Survey, Volume 2, Issue 6, June 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering

[2] Zhongwu Zhai, Bing Liu, Hua Xu and Hua Xu, Clustering Product Features for Opinion Mining, WSDM'11, February 9–12, 2011, Hong Kong, China. Copyright 2011 ACM 978-1-4503-0493- 1/11/02...$10.00

[3] Singh and Vivek Kumar, A clustering and opinion mining approach to socio-political analysis of the blogosphere, Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference.

[4] Alexander Pak and Patrick Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining,

[5] Bing Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.

[6] V. S. Jagtap and Karishma Pawar, Analysis of different approaches to Sentence-Level Sentiment Classification, International Journal of Scientific Engineering and Technology (ISSN : 2277-1581) Volume 2 Issue 3, PP : 164-170 1 April 2013

[7] K. Ming Leung, Naive Bayesian Classifier,Copyright c 2007 mleung@poly.edu.

[8] Chuong B Do & Serafim Batzoglou, What is the expectation maximization algorithm?, 2008 Nature Publishing Group http://www.nature.com/naturebiotechnology

[9] Kamal Nigam, Andrew Kachites Mccallum, Sebastian Thrun and Tom Mitchell, Text Classification from Labeled and Unlabeled Documents using EM, Machine Learning, 39, 103–134, 2000.2000 Kluwer Academic Publishers. Printed in The Netherlands.

[10] Ainur Yessenalina, Yisong Yue and Claire Cardie, Multi-level Structured Models for Documentlevel Sentiment Classification, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 1046–1056, MIT, Massachusetts, USA, 9-11 October 2010. c 2010 Association for Computational Linguistics

[11] Sunghwan Sohn,1,* Manabu Torii,2,* Dingcheng Li,1 Kavishwar Wagholikar,1 Stephen Wu,1 and Hongfang Liu1, A Hybrid Approach to Sentiment Sentence Classification in Suicide Notes, Biomed Inform Insights. 2012; 5(Suppl. 1): 43–50. Published online 2012 January 30. doi: 10.4137/BII.S8961

[12] YuanbinWu, Qi Zhang, Xuanjing Huang, LideWu, Phrase Dependency Parsing for Opinion Mining, Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 1533–1541, Singapore, 6-7 August 2009. c 2009 ACL and AFNLP

[13] Zhai Z, Liu B, Xu H, and Jia P, Grouping Product Features Using Semi-supervised Learning with Soft-Constraints, in Proceedings of COLING. 2010

[14] Khairullah Khan and Baharum B. Baharudin, Identifying Product Features from Customer Reviews using Lexical Concordance, Research Journal of Applied Sciences Engineering and Technology, 2012

[15] Boris Kraychev and Ivan Koychev, Computationally Effective Algorithm for Information Extraction and Online Review Mining, 2010

[16] Akshi Kumar and Teeja Mary Sebastian, Sentiment Analysis: A Perspective on its Past, Present and Future, 2012

[17] Fellbaum C, WordNet: An electronic lexical database. 1998: MIT press Cambridge, MA

[18] Lin D. Automatic retreival and clustering of similar words, 1998: Proceedings of ACL 768-774

[19] Dae Hoon Park and Wei Peng, Generate Adjective Sentiment Dictionary for Social Media Sentiment Analysis Using Constrained Nonnegative Matrix Factorization, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media 2011

[20] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)

[21] Ana-Maria Popescu and Oren Etzioni, Extracting Product Features and Opinions from Reviews

[22] David Vadas and James R. Curran, Adding Noun Phrase Structure to the Penn Treebank, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 240–247 Prague, Czech Republic, June 2007. c 2007 Association for Computational Linguistics

[23] David Vadas and James R. Curran, Parsing Noun Phrases in the Penn Treebank, 2011 Association for Computational Linguistics

# FAKE NEWS DETECTION WITH SEMANTIC FEATURES AND TEXT MINING

Pranav Bharadwaj1 and Zongru Shao2

1 South Brunswick High School, Monmouth Junction, NJ 2 Spectronn , USA

## Abstract

Nearly 70% of people are concerned about the propagation of fake news. This paper aims to detect fake news in online articles through the use of semantic features and various machine learning techniques. In this research, we investigated recurrent neural networks vs. the naive bayes classifier and random forest classifiers using five groups of linguistic features. Evaluated with real or fake dataset from kaggle.com, the best performing model achieved an accuracy of 95.66% using bigram features with the random forest classifier. The fact that bigrams outperform unigrams, trigrams, and quadgrams show that word pairs as opposed to single words or phrases best indicate the authenticity of news.

## Keywords

Text Mining, Fake News, Machine Learning, Semantic Features, Natural Language Processing (NLP)

**Full Text**: **https://aircconline.com/ijnlc/V8N3/8319ijnlc02.pdf**

**Volume Url: https://airccse.org/journal/ijnlc/vol8.html**

## REFERENCES


[1] Babayan, Davit. “Bitcoin Bulls Spreading Fake News About Golden Asteroid: Peter Schiff.” NewsBTC, NewsBTC, 29 June 2019, www.newsbtc.com/2019/06/29/bitcoin-bulls-spreading- fakenews-about-golden-asteroid-peter-schiff/.

[2] Breiman, L. (2001). Random forests machine learning. 45: 5–32. View Article PubMed/NCBI Google Scholar.

[3] Handley, Lucy. “Nearly 70 Percent of People Are Worried about Fake News as a 'Weapon,' Survey Says.” CNBC, CNBC, 22 Jan. 2018, www.cnbc.com/2018/01/22/nearly-70-percent-ofpeople-are-worried-about-fake-news-as-a-weapon-survey-says.html.

[4] Kowalewski, J. (2017). The impact of fake news: Politics. Lexology.

[5] Mitchell, Amy, et al. “Many Americans Say Made-Up News Is a Critical Problem That Needs To Be Fixed.” Pew Research Center's Journalism Project, Pew Research Center, 28 June 2019, www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-thatneeds-to-be-fixed/.

[6] Mikolov, T., M. Karafia´t, L. Burget, J. C˘ernocky`, and S. Khudanpur (2010). Recurrent neural network based language model. In Eleventh annual conference of the international speech communication association.

[7] Patel, S. (2017). Chapter 1 : Supervised learning and naive bayes classification - part 1 (theory). Medium.

[8] Pennington, J., R. Socher, and C. Manning (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543.

[9] Romano, A. (2018). Mark zuckerberg lays out facebooks 3-pronged approach to fake news.

[10] Rubin, V. L., Y. Chen, and N. J. Conroy (2015). Deception detection for news: three types of fakes. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, pp. 83. American Society for Information Science.

[11] Shu, K., A. Sliva, S. Wang, J. Tang, and H. Liu (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter 19(1), 22–36.

[12] Soares, Isa, and Florence Davey-Atlee. “The Fake News Machine: Inside a Town Gearing up for 2020.” CNNMoney, Cable News Network, money.cnn.com/interactive/media/the-macedonia- story/.

[13] Tacchini, E., G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro (2017). Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506.

# AN AUTOMATED MULTIPLE-CHOICE QUESTION GENERATION USING NATURAL LANGUAGE PROCESSING TECHNIQUES

Chidinma A. Nwafor and Ikechukwu E. Onyenwe

Department of Computer Science, Nnamdi Azikiwe University Awka,
Anambra State, Nigeria

## Abstract

Automatic multiple-choice question generation (MCQG) is a useful yet challenging task in Natural Language Processing (NLP). It is the task of automatic generation of correct and relevant questions from textual data. Despite its usefulness, manually creating sizeable, meaningful and relevant questions is a time-consuming and challenging task for teachers. In this paper, we present an NLP-based system for automatic MCQG for Computer-Based Testing Examination (CBTE).We used NLP technique to extract keywords that are important words in a given lesson material. To validate that the system is not perverse, five lesson materials were used to check the effectiveness and efficiency of the system. The manually extracted keywords by the teacher were compared to the auto-generated keywords and the result shows that the system was capable of extracting keywords from lesson materials in setting examinable questions. This outcome is presented in a user-friendly interface for easy accessibility.

## Keywords

Natural Language Processing,NLP, Computer-Based Test Examination, CBTE, Automatic Question Generation, AOG, Document, Tokenization, Multiple-Choice Question, MCQ

**Full Text**: **https://aircconline.com/ijnlc/V10N2/10221ijnlc01.pdf**

**Volume Url: https://airccse.org/journal/ijnlc/vol10.html**

## REFERENCES


[1] Mannem Prashanth, Rashmi Prasad, and Aravind Joshi. (2010). Question generation from paragraphs at UPenn: QGSTEC system description. In Proceedings of QG2010: The Third Workshop on Question Generation, pages 84–91.

[2] Colleen E Crangle and Joyce Brothers Kart. (2015). A questions-based investigation of consumer mental Health information. PeerJ, 3:e867.

[3] Gall MD. (1970). The use of questions in teaching. Review of Educational Research, 40: 707 720.

[4] Swart (2010). Evaluation of Final Examination Papers in Engineering: A Case Study Using Bloom's Taxonomy. IEEE Transactions on Education, (May 2010) Vol. 53, No.2 257-264.

[5] Aldabe, I., De Lacalle, M. L., Maritxalar, M., Martinez, E., and Uria, L. (2006). Arikiturri: an Automatic question generator based on corpora and nlp techniques. In International Conference on Intelligent Tutoring Systems (pp. 584-594). Springer, Berlin, Heidelberg

[6] Folajimi, Y. O., and Omojola, O. E. (2013). Natural language processing techniques for automatic test Questions generation using discourse connectives. Journal of Computer Science and Its Application, 20(2), 6076.

[7] Aditya S. (2018). Using Natural Language Processing for Smart Question Generation. An Intel Developer Program Documentation. https://software.intel.com/en-us/articles/using-natural language processing forsmart-question-generation.

[8] Susanti, Y., Tokunaga, T., Nishikawa, H., and Obari, H. 2018. Automatic distractor generation for multiple Choice English vocabulary questions. Research and Practice in Technology Enhanced Learning, 13(1), 15.

[9] Das, B., and Majumder, M. (2017). Factual open cloze question generation for assessment of learne's knowledge. International Journal of Educational Technology in Higher Education, 14(1), 24.

# HEVAL: YET ANOTHER HUMAN EVALUATION METRIC

Nisheeth Joshi1 , Iti Mathur2 , Hemant Darbari3 and Ajai Kumar4

1,2Department of Computer Science, Banasthali University, India

3,4Center for Development of Advanced Computing, Pune, Maharashtra, India

## Abstract

Machine translation evaluation is a very important activity in machine translation development. Automatic evaluation metrics proposed in literature are inadequate as they require one or more human reference translations to compare them with output produced by machine translation. This does not always give accurate results as a text can have several different translations. Human evaluation metrics, on the other hand, lacks inter-annotator agreement and repeatability. In this paper we have proposed a new human evaluation metric which addresses these issues. Moreover this metric also provides solid grounds for making sound assumptions on the quality of the text produced by a machine translation.

## Keywords

Machine Translation Evaluation, Human Evaluation, Indian Languages, Subjective Evaluation, Semantic Adequacy.

**Full Text**: **https://airccse.org/journal/ijnlc/papers/2513ijnlc02.pdf**

**Volume Url: https://airccse.org/journal/ijnlc/vol2.html**

## REFERENCES


[1] Ananthakrishnan, R., Bhattacharyya, P., Sasikumar, M., Shah, & Ritesh M. (2007). "Some Issues in Automatic Evaluation of English-Hindi MT: More Blues for BLEU". In Proceedings of 5th International Conference on Natural Language Processing, Macmillan Publishers, India.

[2] Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2001). "Bleu: a method for automatic evaluation of machine translation", RC22176 Technical Report, IBM T.J. Watson Research Center.

[3] Chatterjee, N., Johnson, A., & Krishna, M. (2007). "Some Improvements over the BLEU Metric for Measuring Translation Quality for Hindi". In Proceedings of Internation Conference on Computing: Theory and Applications.

[4] Joshi, N., Darbari, H., & Mathur, I. (2012). "Human and Automatic Evaluation of English to Hindi Machine Translation Systems". In Advances in Computer Science, Engineering & Applications, pp. 423-432. Springer Berlin Heidelberg.

[5] Denkowski, M., & Lavie, A. (2011). "Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems". In Proceedings of the EMNLP 2011 Workshop on Statistical Machine Translation.

[6] ALPAC (1966). "Languages and Machines: Computers in Translation and Linguistics" (Technical Report). Automatic Language Processing Advisory Committee (ALPAC), Division of Behavioral Sciences, National Academy of Sciences, National Research Council.

[7] Slype, G.V. (1979). "Systran: evalaution of the 1978 version of systran English-French automatic system of the Commission of the European Communities", The Incorporated Linguist, Vol. 18, pp 86-89.

[8] Falkedal, K. (1991). "Evaluation Methods for Machine Translation Systems". An Historical overview and Critical Account, Technical Report, ISSCO, Universite de Geneve.

[9] Church, K. W., & Hovy, E. H. (1993). "Good Applications for Crummy Machine Translation". Machine Translation, Vol. 8(4), pp 239–258.

[10]White, J. (1993), "Evaluation of Machine Translation", In Proc. of Human Language Technology Workshop, Morgan Kaufmann, San Francisco, pp. 206-210.

[11]EAGLES Evaluation Working Group (1995), "Evaluation of natural language processing systems", Technical Report EAG-EWG-PR.2, ILC-CNR, Pisa. URL: http://isscowww.unige.sh/weg95/ewg95.html.

[12]Eck, M., & Hori., C. (2005). "Overview of the IWSLT 2005 Evaluation Campaign. In Proceedings of IWSLT 2005.

[13]Vanni, M., & Miller, K. (2002). "Scaling the ISLE Framework: Use of Existing Corpus Resources for Validation of MT Evaluation Metrics across Languages". In Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC), pp 1254–1262.

[14]Gates, D., Lavie, A., Levin, L., Waibel, A., Gavaldà, M., Mayfield, L., Woszczyna M., & Zhan P. (1996). "End-to-End Evaluation in JANUS: a Speech-to-Speech Translation System". In Proceedings of the European Conference on Artificial Intelligence, Workshop on Dialogue Processing in Spoken Language, Budapest, Hungary.

[15]Nubel, R. (1997). "End-to-End Evaluation in VERBMOBIL I". In Proceedings of MT Summit VI, pp 232–239, San Diego, CA.

[16]Wilks, Y. (2008). "Machine Translation: Its Scope and Limits". Springer Verlag, New York, NY.

[17]Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., & Schroeder, J. (2007). "(Meta-) Evaluation of Machine Translation". In Proceedings of the ACL Workshop on Statistical Machine Translation, pp 136–158.

[18]Doherty, S., O'Brien, S., & Carl, M. (2010). "Eye Tracking as an MT evaluation technique". Machine Transaltion, Vol 24 (1), pp 1-13.

[19]Koehn, P., Hoang, H., Birch, A., & Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A. and Herbst E., (2007). "Moses: Open source toolkit for statistical machine translation". In Proceedings of ACL: demonstration sesion.

[20]Joshi N., Mathur I., & Mathur S. (2011). "Translation Memory for Indian Languages: An Aid for Human Translators", In Proceedings of 2nd International Conference and Workshop in Emerging Trends in Technology. ACM

[21]Joshi N., Mathur I. (2011). “Design of English-Hindi Translation Memory for Efficient Translation”, In Proceedings of National Conference on Recent Advances in Computer Engineering. Organized by Poornima Group of Institutions, Jaipur.

[22] Joel, M., Mihalcea, R., & Pedersen, T. (2005). "Word Alignment for Languages with Scarce Resources", In Proceedings of the ACL Workshop on Building and Exploiting Parallel Texts: Data Driven Machine Translation and Beyond, Ann Arbor, MI, June 2005

# 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

2Principal, 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**: **https://airccse.org/journal/ijnlc/papers/2513ijnlc04.pdf**

**Volume Url: https://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)

# Hybrid Part-Of-Speech Tagger for Non-Vocalized Arabic Text

Meryeme Hadni1 , Said Alaoui Ouatik1 , Abdelmonaime Lachkar2 and Mohammed Meknassi1

1FSDM, Sidi Mohamed Ben Abdellah University (USMBA), Morocco
2E.N.S.A, Sidi Mohamed Ben Abdellah University (USMBA), Morocco

## Abstract

Part of speech tagging (POS tagging) has a crucial role in different fields of natural language processing (NLP) including Speech Recognition, Natural Language Parsing, Information Retrieval and Multi Words Term Extraction. This paper proposes an efficient and accurate POS Tagging technique for Arabic language using hybrid approach. Due to the ambiguity issue, Arabic Rule-Based method suffers from misclassified and unanalyzed words. To overcome these two problems, we propose a Hidden Markov Model (HMM) integrated with Arabic Rule-Based method. Our POS tagger generates a set of three POS tags: Noun, Verb, and Particle. The proposed technique uses the different contextual information of the words with a variety of the features which are helpful to predict the various POS classes. To evaluate its accuracy, the proposed method has been trained and tested with two corpora: the Holy Quran Corpus and Kalimat Corpus for undiacritized Classical Arabic language. The experiment results demonstrate the efficiency of our method for Arabic POS Tagging. In fact, the obtained accuracies rates are 97.6%, 96.8% and 94.4% for respectively our Hybrid Tagger, HMM Tagger and for the Rule-Based Tagger with Holy Quran Corpus. And for Kalimat Corpus we obtained 94.60%, 97.40% and 98% for respectively Rule-Based Tagger, HMM Tagger and our Hybrid Tagger.

## Keywords

Part-Of-Speech Tagger, Natural Language Applications, Natural Language Parsing, Hidden Markov Model, Multi Words Term Extraction, Speech Recognition.

**Full Text**: **https://airccse.org/journal/ijnlc/papers/2613ijnlc01.pdf**

**Volume Url: https://airccse.org/journal/ijnlc/vol2.html**

## REFERENCE


[1] Lee, S.hyun. & Kim Mi Na, (2008) “This is my paper”, ABC Transactions on ECE, Vol. 10, No. 5, pp120-122. [2] Gizem, Aksahya & Ayese, Ozcan (2009) Comunications & Networks, Network Books, ABC Publishers. [1] http://en.wikipedia.org/wiki/Part-of-speech_tagging.

[2] L.Van Guilder, (1995) “Automated Part of Speech Tagging: A Brief Overview” Handout for LING361, Georgetown University.

[3] H. Halteren, J.Zavrel & Walter Daelemans (2001).Improving Accuracy in NLP Through Combination of Machine Learning Systems. Computational Linguistics. 27(2): 199–229.

[4] DeRose & J.Steven (1990) "Stochastic Methods for Resolution of Grammatical Category Ambiguity in Inflected and Uninflected Languages." PhD.Dissertation. Providence, RI: Brown University Department of Cognitive and Linguistic Sciences.

[5] N. kumar Kumar, Anikel Dalal &Uma Sawant (2006)”hindi part of speech tagging and chunking”, NLPAI machine learning contest.

[6] M. Mohseni, H. Motalebi, B. Minaei-bidgoli & M. Shokrollahi-far (2008) “A farsi part-of-speech tagger based on markov”. In the proceedings of ACM symposium on Applied computing, Brazil.

[7] S. Jabbari &B. Allison(2007)“Persian Part of Speech Tagging”, In the Proceedings of Workshop on Computational Approaches to Arabic Script-Based Languages (CAASL-2), USA.

[8] E. Brill (1995) “Transformation-Based Error-Driven Learning and Natural Language Processing: A case Study in Part of Speech Tagging”, Computational Linguistics, USA.

[9] M. Hepple (2000), ”Independence and Commitment: Assumptions for Rapid Training and Execution of Rule-based Part of-Speech Taggers”, In Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL). Hong Kong.

[10] T. Brants (200),“TNT – a Statistical Part-of-Speech Tagger”, In the Proceedings of 6th conference on applied natural language processing (ANLP), USA.

[11] K. Megerdoomian (2004), “Developing a Persian part-of speech tagger”, In the Proceedings of first Workshop on Persian Language and computer, Iran .

[12] Khoja, S.( 2001) “ APT: Arabic part-of-speech tagger”. Proceeding of the Student Workshop at the 2nd Meeting of the NAACL, (NAACL’01), Carnegie Mellon University, Pennsylvania, pp: 1- 6. http://zeus.cs.pacificu.edu/shereen/NAACL.pdf

[13] Freeman A (2001), “Brill’s POS tagger and a morphology parser for Arabic”, In ACL’01 Workshop on Arabic language processing.

[14] Maamouri M, Cieri C. (2002). “Resources for Arabic Natural Language Processing at the LDC”, Proceedings of the International Symposium on the Processing of Arabic,Tunisia,

pp.125-146.

[15] Diab M., Hacioglu K. and Jurafsky D. (2004), “Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks”. proc. of HLTNAACL’04: 149–152.

[16] Banko M, Moore R. C. (2004). “Part of Speech Tagging in Context”, Proc of the 20th international conference on Computational Linguistics, Switzerland.

[17] Tlili-Guiassa Y. (2006) “Hybrid Method for Tagging Arabic Text”. Journal of Computer Science 2 (3): 245-248.

[18] L. Young-Suk, K. Papineni & S. Roukos ( 2003), “Language Model Based Arabic Word Segmentation,” in Proceedings of the Annual Meeting on Association for Computational Linguistics, Japan, pp. 399- 406.

[19] A.T Al-Taani & S. Abu-Al-Rub (2009),”A rule-based approaches for tagging non-vocalized Arabic words”. The International Arab Journal of Information Technology, Volume6 (3): 320-328.

[20] T. Brants (2000),” TnT: A statistical part of speech tagger”, Proceedings of the 6th Conference on Applied Natural Language Processing, Apr. 29- May 04, Association for Computational Linguistics Morristown, New Jersey, USA., pp: 224-231. [21] NLTK, Natural Language Toolkit. http://www.nltk.org/Home

[22] Quranic Arabic Corpus: http://corpus.quran.com

[23] Quran Tagset: http://corpus.quran.com/documentation/tagset.jsp

[24] N. Habash & O. Rambow (2005), “Arabic Tokenization, Part-of-Speech Tagging and Morphological Disambiguation in One Fell Swoop,” in Proceedings of the Annual Meeting on Association for Computational Linguistics, Michigan, pp. 573-580.

[25] http://sibawayh.emi.ac.ma/web/s/?q=node/79 [26] http://bit.ly/16jO3Ks

[27] http://www.alwatan.com/

[28] F. Al Shamsi & A.Guessoum(2006),” A Hidden Markov Model–Based POS Tagger for Arabic”, 8es Journées internationales d’Analyse statistique des Données Textuelles (JADT).

[29] M. Albared & O.Nazlia(2010),” Automatic Part of Speech Tagging for Arabic: An Experiment Using Bigram Hidden Markov Model “,Springer-Verlag Berlin Heidelberg, LNAI 6401, pp. 361– 370.

[30] Y.O. Mohamed Elhadj(2009),” Statistical Part-of-Speech Tagger for Traditional Arabic Texts”, Journal of Computer Science 5 (11): 794-800.