

## BIBLIOGRAPHY

Abdelali, A., Cowie, J., Farwell, D. and Ogden, D. (2004) UCLIR: a Multilingual Information Retrieval Tool, *Inteligencia Artificial*, Vol. 8, Issue 22, Pp. 103-110.

Abraham, S.A. (2003) Chera, Chola, Pandya: Using Archaeological Evidence to Identify the Tamil Kingdoms of Early Historic South India, *Asian Perspectives*, Vol. 42, No. 2, Pp. 207-223.

Abusalah, M., Tait, J. and Oakes, M. (2005) Literature Review of Cross Language Information Retrieval. *World Academy of Science, Engineering and Technology*, Vol. 4, Pp. 175–177.

Adriani, M. (2002) English-Dutch CLOR using query translation techniques, *Evaluation of Cross-Language Information Retrieval Systems*, Lecture notes in Computer Science, Vol. 2406, Pp. 219-225.

Afraz, M. and Sobha, L (2008) English to Dravidian Language Machine Transliteration: A Statistical Approach Based on N-grams, *Proceedings of International Seminar on Malayalam and Globalization*, Trivandrum, Kerala.

Aggarwal, R.K. and Dave, M. (2011) Using Gaussian Mixtures for Hindi Speech Recognition System, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 4, No. 4, Pp. 157-170.

Aghbari, Z. A. (2005) Array-index: A plug & search K nearest neighbor's method for high-dimensional data. *Data & Knowledge Engineering*, 52(3), 33–352.

Agrawal, R., Imirlinksi, T. and Swami, A. (1993) Mining association rules between sets of items in large databases, *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, Pp. 207-216.

Akila, A. and Chandra, E. (2013) Isolated Tamil Word Speech Recognition System Using HTK, *International Journal of Computer Science Research and Application*, Vol. 03, Issue 02, Pp. 30-38.

Alexandridis, A. and Zapanis, A. (2013) Wavelet neural networks: A practical guide, *Neural networks*, Vol.42, Pp.1-27.

Anand Kumar, M., Dhanalakshmi, V., Soman, K.P. and Rajendran, S. (2010) A Sequence Labeling Approach to Morphological Analyzer for Tamil Language, *International Journal on Computer Science and Engineering*, Vol.02, No. 06, Pp.1944-1951.

Anasyali, M.F. and Ersoy, O.K. (2011) Comparison of single and ensemble classifiers in terms of accuracy and execution time, *International Symposium on Innovations in Intelligent Systems and Applications*, Pp. 470-474.

Antonie, M.L. and Zaiane, O.R. (2002) Text document categorization by term association, *Proceedings of IEEE International Conference on Data Mining, ICDM 2002*, Pp.19- 26.

Anusuya, M.A. and Katti, S.K. (2010) Kannada speech Recognition using discrete wavelet Transform, PCA International Conference on Computer Applications, Pp. 24-27.

Anusuya, M.A. and Katti, S.K. (2012) Speaker Independent Kannada Speech Recognition using Vector quantization, IJCA Proceedings on National Conference on Advancement in Electronics and Telecommunication Engineering, New York, USA, Pp. 32-35.

Avci, E. and Akpolat, Z.H. (2006) Speech recognition using a wavelet packet adaptive network based fuzzy inference system, Expert Systems with Applications, Vol. 31, Pp. 495-503.

Balajapally, P., Pydimarri, P., Ganapathiraju, M., Balakrishnan, N. and Reddy, R. (2006) Multilingual Book Reader: Transliteration, Word-to-Word Translation and Full-text Translation, 13th Biennial Conference and Exhibition Conference of Victorian Association for Library Automation, Melbourne, Australia, Pp. 8-10.

Balamurugan, M.T. and Balaji, M. (2006) SOPC-Based Speech-to-Text Conversion, Nios II Embedded Processor Design Contest Outstanding Designs, Pp. 83-108

Bandyopadhyay, S. (2004) ANUBAAD - The Translator from English to Indian Languages, Proceedings of the VII<sup>th</sup> State Science and Technology Congress, Calcutta, India, Pp.43-51.

Banerjee, P., Garg, G., Mitra, P. and Basu, A. (2008) Application of triphone clustering in acoustic modeling for continuous speech recognition in Bengali, International Conference on Pattern Recognition, Pp. 1-4.

Barnard, E., Davel, M.H. and Van Huyssteen, G.B. (2010) Speech Technology for Information Access: A South African Case Study, AAI Spring Symposium Series, Artificial Intelligence for Development, Pp. 8-13.

Barnett, J., Anderson, S., Broglio, J., Singh, M., Hudson, R. and Kuo, S. W. (1997) Experiments in spoken queries for document retrieval, Proceedings of Eurospeech97, Pp. 1323-1326.

Baskaran, S. and Vaidehi, V. (2004) International journal of Dravidian linguistics, Vol.33, No.1, Pp.13-28.

Basu, A., Watters, C. and Shepherd, M. (2002) Support Vector Machines for Text Categorization, Dalhousie University, Halifax, Nova Scotia, Canada, Proceedings of the 36th Hawaii International Conference on System Sciences, Pp.43-49.

Bauer, E. and Kohavi, R. (1999) An empirical comparison of voting classification algorithms: Bagging, boosting and variants, Machine Learning, Vol. 36, Issues 1-2, Pp. 105-39.

Baxendale, P.B. (1958) Machine-made Index for Technical Literature: An Experiment. IBM, Journal of Research and Development, Vol.2, No.4, Pp.354-361.

Ben, G. and Nelson, M. (1999) Speech and Audio Signal Processing: Processing and Perception of Speech and Music, New York: John Wiley and Sons, Pp. 1-560.

- Benyassine, A. and Akansu, A.N. (1995) Performance Analysis and Optimal Structuring of Subchannels for Discrete Multitone Transceivers, Proc. IEEE Proc. IEEE International Symposium on Circuits and Systems (ISCAS), Pp. 1456-1459.
- Bharathi, B., Deepalakshmi, V. and Nelson, I. (2006) A neural network based speech recognition system for isolated tamil words, Proceedings of International Conference on Neural Networks and Artificial Intelligence, Belarus, Pp. 1-5.
- Bhattacharjee, U. (2013) Recognition of the Tonal Words of BODO Language, International Journal of Recent Technology and Engineering (IJRTE), Vol. 1, Issue 6, Pp. 114-118.
- Biddulph, D.K. and Balashek, S. (1952) Automatic recognition of spoken digits, Journal of the Acoustic Society of America, Vol.24, No.6, Pp. 637-642.
- Bijalwan, V., Kumar, V., Kumari, P. and Pascual, J. (2014) KNN based Machine Learning Approach for Text and Document Mining, International Journal of Database Theory and Application, Vol. 7, No. 1, Pp.61-70.
- Bookstein, A. and Swanson, D.R. (2007) A decision theoretic foundation for indexing, Journal of the Association for Information Science and Technology, Vol. 26, Pp. 45-50.
- Borowicz, A. (2015) A signal subspace approach to spatio-temporal prediction for multichannel speech enhancement, EURASIP Journal on Audio, Speech and Music Processing, Vol. 5, Pp. 1-12.
- Bramer, M. (2013) Ensemble Classification, Principles of Data Mining, Undergraduate Topics in Computer Science, Pp. 209-220.
- Bruce, A., Donoho, D. and YeGao, H. (1996) Wavelet Analysis, IEEE Spectrum, Pp.27-35.
- Calvin, M.N. (1950) The theory digital handling non-numerical information and its implications to machine economics, Zator Technical Bulletin No. 48, Pp. 1-34.
- Chandra, E. and Sujiya, S. (2014) Tamil speech recognition using hybrid technique of EWTLBO and HMM, International Journal of Computer Science and Information Technologies, Vol. 5, No. 5, Pp. 6664-6669.
- Chandrasekar, M. and Ponnaivaikko, M. (2008) Tamil speech recognition: A complete Model, Electronic Journal on Technical Acoustics, Article No. 20, <http://www.ejta.org/en/chandrasekar2>, Last Accessed During August, 2015.
- Chellamuthu, K.C. (2002) Russian language to Tamil machine translation system, INFITT, (TI2002), Pp. 74-83.
- Chen, J. (2006) A lexical knowledge base approach for English–Chinese Cross–language information retrieval, Journal of the American Society for Information Science and Technology, Vol. 57, No. 2, Pp. 233–243.

- Chen, S.F. and Goodman, J. (1996) An Empirical Study of Smoothing Techniques for Language Modeling, Proceedings of the 34th Annual Meeting on Association for Computational Linguistics, ACL '96, Pp. 310–318.
- Chen, Y. and Lin, C. (2006) feature extraction, Springer, Berlin Heidelberg, Pp.315-324.
- Chris, D. (1996) Pick the Right Input Device for Your POC Systems, Health Management Technology, Vol. 17, No.6, Pp. 24–27.
- Clarke, D.C. and Wall, R.E. (1965) An Economical Program for Limited Parsing of English, Proceedings of the November 30–December 1, 1965, Fall Joint Computer Conference, Part I, AFIPS '65 (Fall, part I), Pp. 307–316.
- Coifman, R.R. and Wickerhauser, M.V. (1992) Entropy-Based Algorithms for Best Basis Selection, IEEE Transactions on Information Theory, Vol.38, No.2, Pp.713–718.
- Cooper, W.S., Chen, A. and Gey, F.C. (1994) Full Text Retrieval Based on Probabilistic Equations with Coefficients Fitted by Logistic Regression, Proceedings of the Second Text Retrieval Conference (TREC-2), Gaithersburg, MD, Pp.57-66.
- Cormack, G.V. and Lynam, T.R. (2005) Spam corpus creation for TREC, CEAS, Second Conference on Email and Anti-Spam, Stanford University, California, Pp. 1-8.
- Cortes, C. and Vapnik, V. (1995) Support Vector Networks, Machine Learning, Vol. 20, Pp. 273-297.
- Cover, T. and Hart, P. (2006) Nearest neighbor pattern classification. IEEE Transactions on Information Theory, Vol. 13, No. 1, Pp. 21-27.
- Crestani, F. (2000) Word recognition errors and relevance feedback in spoken query processing, Proceedings of the Fourth International Conference on Flexible Query Answering Systems, Pp. 267–281.
- Croft, W.B. (1993) Knowledge-based and statistical approaches to text retrieval, IEEE Expert, Vol. 8, No. 2, Pp. 8-12.
- Cummins, R., Paik, J.H. and Lv, Y. (2015) A Polya Urn Document Language Model for Improved Information Retrieval, Computer Science, Information Retrieval, arXiv:1502.00804, Pp. 1-37.
- Dagan, I. and Itai, A. (1994) Word Sense Disambiguation using a Second Language Monolingual Corpus. Computational Linguistics, Pp.563-596.
- Daqrouq, K., Al-Qawasmi, A.R., AlAzzawi, K.Y. and Hilal, A.T. (2011) Discrete Wavelet Transform & Linear Prediction Coding Based Method for Speech Recognition via Neural Network, Discrete Wavelet Transforms - Biomedical Applications, Prof. HannuOlkkonen (Ed.), <http://cdn.intechopen.com/pdfs-wm/19512.pdf>, Last Accessed During January, 2015.

Davis, M., and Dunning, T. (1995) A TREC evaluation of query translation methods for multilingual text retrieval, Proceedings of TREC Conference, Pp. 483–497.

Definitive Editions of Ancient Tamil Works (2008) Classical Tamil, Centre of Excellence for Classical Tamil, Government of India, <http://web.archive.org/web/20100707000020/http://www.ciil-classicaltamil.org/project1.html>, Last Accessed During August, 2015.

Deka, M.K., Nath, C.K., Sarma, S.K. and Talukdar, P.H. (2011) An Approach to Noise Robust Speech Recognition using LPC-Cepstral Coefficient and MLP based Artificial Neural Network with respect to Assamese and Bodo Language, International Symposium on Devices MEMS, Intelligent Systems & Communication (ISDMISC), Pp. 23-26.

Deng, L. and Li, X. (2013) Machine Learning Paradigms for Speech Recognition: An Overview, IEEE Transactions ON Audio, Speech and Language Processing, Vol. 21, No. 5, Pp. 1-30.

Dhanalakshmi, V., Anandkumar, Shivapratap, G., Soman, K.P. and Rajendran, S. (2009) Tamil POS tagging using linear programming, International journal of recent trends in engineering, Vol.1, No.2, Pp.166-169.

Dharun, V.S. and Karnan, M. (2012) Voice and Speech Recognition for Tamil Words and Numerals, IJMER, Vol. 2, No.5, Pp. 3406-3414.

Dietterich, T. G. (2000) Ensemble methods in machine learning, Kittler, J. and Roli, F. (Eds.), Proc. of 1st International Workshop on Multiple Classifier Systems (MCS 2001), Springer-Verlag, Berlin / Heidelberg.

Dillon, M. and McDonald, L.K. (1983) Fully Automatic Book Indexing. Journal of Documentation, Vol. 39, No. 3, Pp. 135–154.

Donoho, D.L. (1992) De-noising by Soft-Thresholding, IEEE Transaction on Information Theory, Vol.41, No.3, Pp.613-627.

Donoho, D.L. and Johnstone, I.M. (1995) Adapting to unknown smoothness via wavelet shrinkage, Journal of the American Statistical Association, Vol. 90, No. 432, Pp. 1200-1224.

Efthimiadis, E.N. (1996) Query expansion, Annual Review of Information Science and Technology, Vol. 31, Pp. 121-187.

Eirinaki, M. (2009) Web Mining: A Roadmap, Athens University of Economics and Business, <http://www.dbnet.aueb.gr/index.php/corporate/content/download/341/1381/file/NEMIS.pdf>.

Ekbal, A. and Saha, S. (2013) Simulation Annealing based classifier ensemble techniques: Application to part of speech tagging, An International Journal on Multi-sensor, Multi-Source, Information fusion, Vol.14, Issue 3, Pp. 288-300.

Ertoz, L., Michael, S. and Kumar, V. (2003) Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data, SDM. SIAM, Pp. 1-6.

- Evans, D.A. and Zhai, C. (1996) Noun-Phrase Analysis in Unrestricted Text for Information Retrieval, Proceedings of the 34th annual meeting on Association for Computational Linguistics, Association for Computational Linguistics, Pp. 17–24.
- Fagan, J.L. (1987) Experiments in Automatic Phrase Indexing for Document Retrieval: A Comparison of Syntactic and non-Syntactic Methods. PhD thesis, Cornell University.
- Ferrandez, A. (2011) Lexical and Syntactic knowledge for Information Retrieval, *Information Processing & Management*, Vol. 47, Issue 5, Pp. 692–705.
- Forestiero, A., Pizzuti, C. and Spezzano, G. (2013) A single pass algorithm for clustering evolving data streams based on swarm intelligence, *Data Min. Knowl. Discov.*, Vol. 26, No. 1, Pp. 1-26.
- Freund, Y. and Shapire, R. (1995) A decision-theoretic generalization of on-line learning and an application to boosting, Proceedings of the second European Conference on Computational Learning Theory, San Francisco, Springer-Verlag., Pp. 23–37.
- Freund, Y. and Shapire, R. (1996) Experiments with a new boosting algorithm. Proceedings of the 13th International Conference on Machine Learning, San Francisco, Morgan Kaufmann, Pp. 148–156.
- Fujii, A., Itou, K. and Ishikawa, T. (2002a) Speech-driven text retrieval : Using target IR collections for statistical language model adaptation in speech recognition, A.R. Coden, E.W. Brown and S. Srinivasan (Eds.), *IR Techniques, LNCS 2273*, Springer-Verlag Berlin Heidelberg, Pp. 94-104.
- Fujii, A., Itou, K. and Ishikawa, T. (2002b) A Method for Open-Vocabulary Speech-Driven Text Retrieval, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Pp. 188-195.
- Gabrea, L.A. and Gargour, C.S. (2004) Wavelet based speech enhancement using two different threshold-based denoising algorithms , *IEEE Xplore Digital Library, Canadian Conference on Electrical and Computer Engineering*, Vol. 1, Pp.315-318.
- Ganesan, K. and Siva, G. (2007) Multilingual Querying and information processing, *Information Technology Journal*, Pp. 751-755.
- Garofolo, J.S., Voorhees, E.M., Stanford, V.M. and Jones, K.S. (1997) TREC-6 1997 spoken document retrieval track overview and results, Proceedings of the 6th Text Retrieval Conference, Pp. 83–91.
- Gatian, A.W. (1994) Is user satisfaction a valid measure of system effectiveness?, *Information and Management*, Vol.26, No.3, Pp119-131.
- Gawali, B.W., Gaikwad, S., Yannawar, P. and Mehrotra, S.C. (2011) Marathi Isolated Word Recognition System using MFCC and DTW Features, *ACEEE Int. J. on Information Technology*, Vol. 01, No. 01, Pp. 21-24.

- Gawande, P.S. and Suryawanshi, A. (2015) Improving Web Page Classification by Vector Space Model, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 2, Pp. 1252-1257.
- George, L.H. (2015) Statement on the Status of Tamil as a Classical Language, University of California Berkeley Department of South Asian Studies – Tamil , [http://www.oocities.org/bhel\\_ywca/Tamil2.htm](http://www.oocities.org/bhel_ywca/Tamil2.htm), Last Accessed During August, 2015.
- Germann, U. (2001) Building a statistical machine translation system from scratch: how much bang for the buck can we expect?, Proceedings of the workshop on Data-driven methods in machine translation, Pp. 1–8.
- Gey, F.C. (2002) Prospects for Machine Translation of the Tamil Language, Proceedings of Tamil Internet, Pp. 221-227.
- Giannakopoulos, T. (2010) <http://www.mathworks.com/matlabcentral/fileexchange/28826-silence-removal-in-speech-signals>, Last Access During August, 2015.
- Gluck, M. (1996) Exploring the relationship between user satisfaction and relevance in information systems, Information Processing and Management, Vol.32 No.1, Pp.89-104.
- Gnanathesigar, H. (2012) Tamil Speech Recognition using semi continuous models, International Journal of Scientific and Research Publications, Vol. 2, Issue 6, Pp. 1-5.
- Gnatyshak, D.V. (2015) A single-pass triclustering algorithm, Automatic Documentation and Mathematical Linguistics, Vol. 49, Issue 1, Pp. 27-41.
- Godambe, T. and Samudravijaya, K. (2011) Speech Data Acquisition for voice based Agricultural Information Retrieval, Proceedings of 39th All India DLA Conference, Patiala, India, Pp. 1-8.
- Goenka, K., Arpinar, I. B. and Mustafa, N. (2010) Mobile Web Search Personalization using Ontological User profile, Proceedings of 48<sup>th</sup> ACM SE, Pp. 13-21.
- Goto, I., Utiyama, M., Sumita, E. and Kurohashi, S. (2015) Preordering using a Target-Language Parser via Cross-Language Syntactic Projection for Statistical Machine Translation, ACM Trans. Asian Low-Resource Language Information Processing, Vol. 14, No. 3, Article 13, Pp. 1-23.
- Guo, Y., Jiao, L., Wang, S., Wang, S., Liu, F., Rong, K. and Xiong, T. (2015) A novel dynamic rough subspace based selective ensemble, Pattern Recognition, Vol. 48, No. 5, Pp. 1638-1652.
- Gupta, R. (2006) Speech Recognition for Hindi, M. Tech Thesis, Dept. Computer Science and Eng. IIT Bombay, Bombay.
- Gupta, S. and Garg, D. (2011) Comparison of semantic and syntactic information retrieval system on the basis of precision and recall, International Journal of Data Engineering, Vol. 2, Issue 3, Pp. 93-101.

Harshawardhan, R., Augustine, M.S. and Soman, K.P. (2011) Phrase based English – Tamil Translation System by Concept Labeling using Translation Memory, International Journal of Computer Applications (0975 – 8887), Vol. 20, No.3, Pp. 1-6.

Hassan, F., Uzzaman, N., and Khan, M. (2007) Advances and Innovations in systems, computing science and software engineering, Springer, Netherlands, Pp.121-126.

Hassan, M. R., Nath, B. and Bhuiyan, A. (2003) Bengali Phoneme Recognition: A New Approach, Proc. 6th ICCIT Dhaka (www.iccit.org), Pp. 365-369.

Hastie, T., Tibshirani, R., and Friedman, J. (2001) The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer-Verlag, New York.

He, B., Huang, J.X. and Zhou, X. (2011) Modeling term proximity for probabilistic information retrieval models, Information Science, Vol. 181, No. 14, Pp. 3017-3031.

Hegde, R.M., Murthy, H.A. and Gadde, V.R.R. (2004) Continuous speech recognition using joint features derived from the modified group delay function and MFCC, Proceedings of ICSLP, Jeju, Korea, Pp. 905–908.

Hegde, S., Achary K.K. and Surendra, S. (2012) Isolated Word Recognition for Kannada Language Using Support Vector Machine, Wireless Networks and Computational Intelligence Communications in Computer and Information Science, Vol. 292, Pp. 262-269.

Hegde, S., Achary, K.K. and Shetty, S. (2013) Analysis of Isolated Word Recognition for Kannada Language using Pattern Recognition Approach, International Journal of Information Processing , Vol.7, No.1, Pp.73-80.

Hiemstra, D. (2001) Using language models for information retrieval, CTIT Ph.D. Thesis Series No. 01-32, Centre for Telematics and Information Technology, Netherlands.

Hillman, D.J. (1968) Negotiation of Inquiries in an On-line Retrieval System, Information Storage and Retrieval, Vol.4, No.2, Pp.219–238.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B.(2012) Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups, IEEE Signal Processing Magazine, Vol. 29, Issue 6, Pp. 82-97.

Hirst, G. (1987) Semantic Interpretation and the Resolution of Ambiguity, Cambridge: Cambridge University Press.

Hoque, A.K.M.M. (2006) Bengali Segmented Speech Recognition System, Undergraduate Thesis Report, Computer Science and Engineering, BRAC University.

[http://en.wikipedia.org/wiki/Okapi\\_BM25](http://en.wikipedia.org/wiki/Okapi_BM25), Last Accessed During August, 2015.

[http://en.wikipedia.org/wiki/Tamil\\_language](http://en.wikipedia.org/wiki/Tamil_language), Last Accessed During August, 2015.

<http://global-reach.biz/globstats/index.php3>, Last Accessed During August, 2015.

<http://net.pku.edu.cn/~webg/cwt>, Last Accessed During August, 2015.

<http://nlp.stanford.edu/IR-book/html/htmledition/language-models-for-information-retrieval-1.html>, Last Accessed During August 2015.

<http://pib.nic.in/newsite/erelease.aspx?relid=87874>, Last Accessed During August, 2015.

<http://research.nii.ac.jp/ntcir>, Last Accessed During August, 2015.

<http://tdil.mit.gov.in>, Last Accessed During August, 2015.

<http://trak.in/tags/business/2014/11/19/india-300m-internet-users-2014>, Last Accessed During August, 2015.

<http://trec.nist.gov>, Last Accessed During August, 2015.

<http://voicefinger.cozendey.com>, Last Accessed During August, 2015.

<http://www.amazon.com/gp/product/B00I3JF87U>, Last Accessed During August, 2015.

<http://www.azhagi.com/az-tamil-modern.html>, Last accessed during August 2015.

<http://www.clef-campaign.org>, Last Accessed During August, 2015.

<http://www.e-speaking.com>, Last Accessed During August, 2015.

[http://www.euromatrix.net/deliverables/Euromatrix\\_D1.3\\_Revised.pdf](http://www.euromatrix.net/deliverables/Euromatrix_D1.3_Revised.pdf), Last Accessed during August, 2015.

<http://www.isi.edu>, Last Accessed During August, 2015.

<http://www.isical.ac.in/~clia>, Last Accessed During August, 2015.

<http://www.isical.ac.in/~clia/2011>, Last Accessed During August, 2015.

<http://www.isical.ac.in/~fire/data.html>, Last Accessed During August, 2015.

<http://www.nexus.carleton.ca/~kekoura/history.html>, Last Accessed During August 2015.

<http://www.tazti.com/index.php>, Last Accessed During August, 2015.

<http://www.thehindu.com/2005/11/22/stories/2005112215970400.htm>(2005) Students get glimpse of heritage, The Hindu, India, , Last Accessed During August, 2015.

<http://www.thehindu.com/todays-paper/tamil-brahmi-script-in-egypt/article1952611.ece> (2007) Tamil Brahmi script in Egypt, The Hindu. India, Last Accessed During August, 2015.

[https://en.wikipedia.org/wiki/Binary\\_Independence\\_Model](https://en.wikipedia.org/wiki/Binary_Independence_Model), Last Accessed During August, 2015.

[https://en.wikipedia.org/wiki/Cross-language\\_information\\_retrieval](https://en.wikipedia.org/wiki/Cross-language_information_retrieval), Last Accessed During August, 2015.

[https://en.wikipedia.org/wiki/Document\\_retrieval](https://en.wikipedia.org/wiki/Document_retrieval), Last Accessed During August 2015.

[https://en.wikipedia.org/wiki/Ensembles\\_of\\_classifiers#Weaknesses](https://en.wikipedia.org/wiki/Ensembles_of_classifiers#Weaknesses), Last Accessed During August, 2015.

[https://en.wikipedia.org/wiki/Relevance\\_feedback](https://en.wikipedia.org/wiki/Relevance_feedback), Last Accessed During June, 2015.

[https://en.wikipedia.org/wiki/Wavelet\\_packet\\_decomposition](https://en.wikipedia.org/wiki/Wavelet_packet_decomposition), Last Accessed During August, 2015.

[https://www.clear.rice.edu/elec301/Projects01/dig\\_hear\\_aid/noisered.html](https://www.clear.rice.edu/elec301/Projects01/dig_hear_aid/noisered.html), Last Access During August, 2015.

Huang, M. and Cui Baotong. (2005) *Advances in natural computation*, Springer, Berlin Heidelberg, Pp.1-7.

Ionescu, M., Sterca, A. and Badarinza, I. (2014) *Syntactic Indexes for Text Retrieval*, Proceedings of the National Symposium ZAC2014, Pp. 19-31.

Iravatham, M. (2003) *Early Tamil Epigraphy from the Earliest Times to the Sixth Century A.D.*, Harvard Oriental Series, Vol. 62, Cambridge: Harvard University Press.

Iravatham, M. (2010) *An epigraphic perspective on the antiquity of Tamil*, *The Hindu, India*, <http://www.thehindu.com/opinion/op-ed/an-epigraphic-perspective-on-the-antiquity-of-tamil/article482654.ece>, Last Accessed During August, 2015.

Itou, K., Fujii, A. and Ishikawa, T. (2001) *Language modelling for multi-domain speech-driven text retrieval*, *IEEE Automatic Speech Recognition and Understanding Workshop*, Pp. 1-4.

Jagarlamudi, J. and Kumaran, A. (2008) *Crosslingual Information Retrieval System For Indian Languages*, *Cross Language Extraction Forum(CLEF)*, Pp. 21-29.

Janarthanam, S., Sethuramalingam, S. and Nallasamy, U. (2008) *Named Entity Transliteration for Cross Language Information Retrieval using Compressed Word Format Algorithm*, *2nd International ACM Workshop Improving Non-English Web Searching (iNEWS-08)*, California.

Jansen, B.J., Spink, A. and Pedersen, J. (2005) *A Temporal Comparison of Altavista Web Searching*, *Journal of the American Society for Information Science and Technology*, Vol.56, No.6, Pp.559–570.

Jiang, S. (2006) *Efficient Classification Method for Large Dataset*, *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian*, Pp. 1190–1194.

- Jiang, S., Pang, G., Wu, M. and Kuang, L. (2012) An improved K-nearest-neighbour algorithm for text categorization, *Expert Systems with Applications*, Vol. 39, Pp. 1503-1509.
- Jin, Y.G., Shin, J.W. and Kim, N.S. (2014) Spectro-Temporal Filtering for Multichannel Speech Enhancement in Short-Time Fourier Transform Domain, *IEEE Signal Processing Letters*, Vol. 21, No. 3, Pp. 352-355.
- Johnson, M.T., Yuan, X. and Ren, Y. (2007) Speech Signal Enhancement through Adaptive Wavelet Thresholding, *Speech Communication*, Vol. 49, No. 2, Pp. 123-133.
- Johnson, S., Jourlin, P., Moore, G., Jones, K. S. and Woodland, P. (1999) The Cambridge University spoken document retrieval system, *Proceedings of ICASSP'99*, Pp. 49-52.
- Jones, G., Foote, J., Jones, K. S. and Young, S. (1996) Retrieving spoken documents by combining multiple index sources, *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Pp. 30-38.
- Jones, K.P. (1983) How do we Index? A Report of some ASLIB Informatics Group Activities, *Journal of Documentation*, Vol.39, No.1, Pp.1-23.
- Kalyanasundaram (2015) A brief history and features of the Tamil language, *Tamil Electronic Library*, <http://tamilelibrary.org/teli/tamil7.html>, Last Accessed During August 2015.
- Kalyani, N. and Sunitha, K.V.N. (2009) Syllable analysis to build a dictation system in Telugu language, *International Journal of Computer Science and Information Security*, Vol.6, No.3, Pp.171-176.
- Kamil, Z. (1973) *The Smile of Murugan on Tamil Literature of South India*, BRILL, Pp. 11-12.
- Kamil, Z. (1992) *Companion studies to the history of Tamil literature*, Leiden, Brill, Pp.12.
- Karpagavalli, S., Deepika, R., Kokila, P., Usharani, J., and Chandra, E. (2012b) Isolated Tamil Digits Speech Recognition using template-based and HMM-based approaches, *Global Trends in Information Systems and Software Applications, Communications in Computer and Information Science*, Vol. 270, Pp. 441-450.
- Karpagavalli, S., Usharani, J., Deepika, R., and Kokila, P. (2012a) Isolated Tamil Digits Speech Recognition using Vector Quantization, *International Journal of Engineering Research and Technology*, Vol. 1, Issue 4, Pp. 1-12.
- Keerthi, S. S. and DeCoste, D. (2005) A modified finite Newton method for fast solution of large scale linear SVMs, *JMLR*, Vol. 6, Pp.341-361.
- Kembellec, G., Saleh, I. and Sauvaget, V. (2009) Ontology Navigator: WEB 2.0 scalable ontology based CLIR portal to IT scientific corpus for researchers, *International Journal of Design Sciences and Technology*, Vol.16, No.2, Pp.105-119.

- Kolappan, B. (2014) Delay, howlers in Tamil Lexicon embarrass scholars, The Hindu, India, <http://www.thehindu.com/news/national/tamil-nadu/delay-howlers-in-tamil-lexicon-embarrass-scholars/article6138747.ece>, Last Accessed During August, 2015.
- Kraaij, W. and Spitters, M. (2003) Language Models for Topic Tracking. In W. B. Croft and J. Lafferty, editors, *Language Modeling for Information Retrieval*, Pp. 95–124.
- Krishnamurti, B. (2003) *The Dravidian languages*, Cambridge University Press, New York.
- Krishnan, M., Neophytou, C.P. and Prescott, G. (1994) Wavelet transform speech recognition using vector quantization, dynamic time warping and artificial neural networks, *International Conference On Spoken Language Processing*, Pp. 1191–1193.
- Krishnan, V.R.V., Jayakumar, A. and Anto, P.B. (2008) Speech Recognition of isolated Malayalam Words Using Wavelet features and Artificial Neural Network, *4th IEEE International Symposium on Electronic Design, Test and Applications*, Pp. 240 – 243.
- Kumar, A. and Das, S. (2013) Pre-Retrieval based Strategies for Cross Language News Story Search, *Proceedings of the 5th 2013 Forum on Information Retrieval Evaluation (FIRE '13)*, Prasenjit Majumder, Mandar Mitra, Madhulika Agrawal, and Parth Mehta (Eds.). ACM, New York, Article 11 , Pp. 1-10.
- Kumar, K. and Aggarwal, R.K. (2011) Hindi Speech Recognition System Using HTK, *International Journal of Computing and Business Research*, Vol. 2, Issue 2, Pp. 1-12.
- Kumar, R. (2010) Comparison of HMM and DTW for Isolated Word Recognition System for Punjabi Language, *International Journal of Soft Computing*, Vol.5, No.3, Pp.88-92.
- Kumar, V., Steinback, M. and Tan, P. (2004) Introduction to data mining, *Lecture Notes for Chapter 6*, Pp.1-88.
- Kumaran, A. and Kellner, T. (2007) A Generic Framework for Machine Transliteration, *30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007)*, Amsterdam, Netherlands, Association for Computing Machinery, Inc., Pp. 721-723.
- Kuncheva, L.I. (2004) *Combining pattern classifiers: methods and algorithms*, John Wiley & Sons, Hoboken, New Jersey.
- Kunpeng, Z., Xiaolong, W. and Yuanchao, L. (2009) A new query expansion method based on query logs mining, *International Journal on Asian Language Processing*, Vol. 19, No. 1, Pp. 1-12.
- Kupiec, J., Kimber, D. and Balasubramanian, V. (1994) Speech-based retrieval using semantic co-occurrence filtering, *Proceedings of the ARPA Human Language Technology Workshop*, Pp. 373–377.

- Kurian, C. (2014) A review on the progress of natural language processing in India, *International Journal of Advances in Engineering and Technology*, Vol. 7, Issue 5, Pp. 1420-1425.
- Kwok, K.L. (1996) Network Approach to Probabilistic Information Retrieval, *ACM Transactions on Information Systems (TOIS)*, Vol.13, No.3, Pp.324–353.
- Lafferty, J. and Zhai, C. (2001) Document Language Models, Query Models and Risk Minimization for Information Retrieval, *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New Orleans, Louisiana, United States, Pp.111-119.
- Lakshmi, A. and Hema, A.M. (2006) A Syllable Based Continuous Speech Recognition for Tamil, *INTERSPEECH -ICSLP*, Pennsylvania, Pp. 1878-1881.
- Larsen, B. and Aone, C. (1999) Fast and effective text mining using linear-time document clustering, *Proceedings of the 5th International Conference on ACM Special Interest Group on Knowledge Discovery and Data Mining*, Aug. 15-18, San Diego, CA, Pp. 16-22.
- Lavrenko, V. (2004) A Generative Theory of Relevance, PhD thesis, University of Massachusetts.
- Lee, K.S. and Croft, W.B. (2013) A deterministic resampling method using overlapping document clusters for pseudo-relevance feedback, *Information Processing and Management*, Vol. 49, Pp. 792-06.
- Lehtokangas, R., Keskustalo, H. and Jarvelin, K. (2006) Experiments with dictionary-based CLIR using graded relevance assessments: Improving effectiveness by pseudo-relevance feedback, *Information Retrieval*, Vol. 9, No. 4, Pp. 421-433.
- Lesk, M. (1986) Automatic Sense Disambiguation: How to tell a Pine Cone from an Ice Cream Cone, *Proceedings of the 1986 SIGDOC Conference*, ACM, New York, Pp.24-26.
- Li, B., Sugandh, N., Garcia, E.V. and Ram, A. (2007) Adapting Associative Classification to Text Categorization, In. *Proceedings of the 2007 ACM Symposium on Document Engineering*, Winnipeg, Canada, Pp. 205-208.
- Li, J., Deng, L., Gong, Y. and Haeb-Umbach, R. (2014) An overview of noise-robust automatic speech recognition, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, Vol. 22, Issue 4, Pp. 745-777.
- Li, R. and Hu, Y. (2004) A density-based method for reducing the amount of training data in kNN text classification. *Journal of Computer Research and Development*, 41(4), 539–545.
- Li, W., Han, J. and Pei J. (2001) CMAR: Accurate and efficient classification based on multiple class-association rule, *IEEE International Conference on Data Mining (ICDM'01)*, San Jose, CA, Pp. 369–376.
- Lindsey, B.G. (1996) *Mixture Models: Theory, Geometry and Applications*. IMS.

- Lioma, C. and Ounis, I. (2008) A syntactically-based query reformulation technique for information retrieval, *Information Processing and Management*, Vol. 44, Pp. 143–162.
- Liu, B., Hsu, W. and Ma, Y. (1999) Mining association rules with multiple minimum supports, *Proceedings of fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, California, Pp. 337-341.
- Liu, B., Hsu, W., and Ma, Y. (1998) Integrating Classification and association rule mining, *KDD '98*, New York, NY, Aug. 1998.
- Liu, C., Wang, H., McClean, S., Liu, J. and Wu, S. (2007) Syntactic Information Retrieval, *IEEE International Conference on Granular Computing*, Pp. 703-709.
- Lloyd, S.P. (1982) Least squares quantization in PCM. *IEEE Transactions on Information Theory*, Vol. 28, No.2, Pp.129–136.
- Lodato, G. (2002) Voice recognition technology from a user's perspective, [www.e-bility.com/articles/voice\\_recognition.php](http://www.e-bility.com/articles/voice_recognition.php), Last Accessed 18-12-2007.
- Loganathan, R. (2010) English-Tamil Machine Translation System, Master of Science by Research Thesis, Amrita Vishwa Vidyapeetham, Coimbatore.
- Luk, R.W.P. and Kwok, K.L. (2002) A Comparison of Chinese Document Indexing Strategies and Retrieval Models, *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol.1, No.3, Pp.225–268.
- Ma, J., Wu, F. and Li, C. (2014) The Topic Tracking Based on Semantic Similarity of Sememe's Lexical Chain, *2nd International Conference on Software Engineering, Knowledge Engineering and Information Engineering*, Pp. 118-121.
- Madhavan, K. (2010) Tamil saw its first book in 1578, *The Hindu*, India, <http://www.thehindu.com/news/national/tamil-nadu/tamil-saw-its-first-book-in-1578/article476102.ece>, Last Accessed During August, 2015.
- Madhavi, G., Balakrishnan, M. and Balakrishnan, R.N. (2005) One tool for many (Indian) languages, *Journal of Zhejiang University Science*, Vol.6, No.11, Pp. 1348-1353.
- Madhuresh, S., Prasanna, T.S., Sharad, K.S. and Shashinath, S. (2003) Developing Information Technology Solutions in Indian Languages: Pros and Cons., *First International CALIBER : Mapping Technology on Libraries and People*, Ahmedabad (India), Pp. 1-10.
- Maeda, T., Momouchi, Y. and Sawamura, H. (1980) An Automatic Method for Extracting Significant Phrases in Scientific or Technical Documents, *Information Processing & Management*, Vol.16, No.3, Pp.119–127.
- Maganti, H.K. and Matassoni, M. (2014) Auditory processing-based features for improving speech recognition in adverse acoustic conditions, *EURASIP Journal on Audio, Speech and Music Processing*, Vol. 21, Pp. 1-9.

- Maitah, W., Al-Rababaa, M. and Kannan, G. (2013) Improving the effectiveness of information retrieval system using adaptive genetic algorithm, *International Journal of Computer Science & Information Technology (IJCSIT)* Vol 5, No 5, Pp. 91-105.
- Mallat, S. (1998) *A Wavelet Tour of Signal Processing*, Academic, New York.
- Maloney, C. (1970) The Beginnings of Civilization in South India, *The Journal of Asian Studies*, Vol. 29, No. 3, Pp. 603-616.
- Mandal, S., Das, B , Mitra, P. (2010) Shruti-II: A vernacular speech recognition system in Bengali and an application for visually impaired community, *IEEE Students' Technology Symposium (TechSym)*, Pp. 229-233.
- Mangai, J.A., Kumar, V.S. and Balamurugan, S. (2012) A novel feature selection framework for automatic web page classification, *International Journal of Automation and Computing*, Springer-Verlag, Vol. 9, No.4, Pp.442 – 448.
- Mangai, J.A., Wagle, S.M. and Kumar, V.S. (2013) A novel web page classification model using an improved k nearest neighbor algorithm, *3rd International Conference on Intelligent Computational Systems*, Pp. 49-53.
- Maning, C.D., Raghavan, P. and Schtze, H. (2008) *An Introduction to Information Retrieval*, Cambridge University Press, Cambridge, UK.
- Manning, C.D. and Schutze, H. (1999) *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, Massachusetts, USA.
- Manning, C.D., Raghavan, P. and Schutz, H. (2009) *Introduction to Information Retrieval*, Cambridge University Press, <http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf>, Last Accessed During August, 2015.
- Maria Gabriela, F.D. (2005) *A Machine translation approach to cross-language text retrieval*, USA, Dissertation.com.
- Marin, E.C., Zamora, M. and Rodriguez, A.F. (2014) Improving DFR Models for Information Retrieval in Cognitive Contexts by Adjusting the Documents Length Normalization, *Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management*, Indonesia, Pp. 2402-2410.
- Mathur, R., Babita, Kansal, A. (2010) Domain specific speaker independent continuous speech recognizer using Julius, *Proceedings of ASCNT – 2010*, CDAC, Noida, India, Pp. 55 – 60.
- Mavaluru, D. (2014) *Grammar rule based cross-language information retrieval for Telegu*, A Thesis Submitted in Partial Fulfillment for the Award of Degree of Doctor of Philosophy in Computer Science, B.S. Abdur Rahman University, B.S. Abdur Rahman Institute of Science and Technology, India.

- Mishra, A.N., Biswas, A., Chandra, M. and Sharan, S.N. (2011) Robust Hindi connected digits recognition, *International Journal of Signal Processing, Image Processing and Pattern Recognition* Vol. 4, No. 2, Pp. 12-16.
- Mohamed, A. and, Nair, K.N.R. (2010) Continuous Malayalam speech recognition using Hidden Markov Models, *Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India*, Article No. 29, Pp. 1-4.
- Mohanty, S. and Swain, B.K. (2011) Continuous Oriya Digit Recognition using Bakis Model of HMM, *International Journal of Computer Information Systems*, Vol. 2, No. 1, Pp. 11-15.
- Mohanty, S., Swain, B. K. (2010) Markov Model Based Oriya Isolated Speech Recognizer-An Emerging Solution for Visually Impaired Students in School and Public Examination, *Special Issue of IJCCT*, Vol. 2, Issue 2, 3, 4, *International Conference On Communication Technology*, Pp. 107-111.
- Murthy, B.K. and Deshpande, W.R. (1998) Language technology in India: past, present and future, <http://www.cicc.or.jp/english/hyoujyunka/mlit3/7-12.html>, Last Accessed During August 2015.
- Nayeemulla. A.K. and Yegnanarayana, B. (2001) Development of Speech Recognition System for Tamil for Small Restricted Task, *Proceedings of National Conference on Communication*, Pp. 1-8.
- Neeba, N.V and Jawahar, C.V. (2009) Empirical evaluation of character classification schemes, *Seventh International Conference on Advances in Pattern Recognition (ICAPR)*, Pp. 310–313.
- Nehe, S.N. and Holambe, S.R. (2012) DWT and LPC based feature extraction method for isolated word recognition, *Eurasip Journal on Audio, Speech and Music Processing*, Vol. 7, Pp. 1-7.
- Ng, H.T and Lee H.B. (1996) Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-Based Approach, *Association of Computational Linguistics (ACL)*, Pp. 40-47.
- Ngai, E., Poon, J. And Chan, Y. (2007) Empirical examination of the adoption of WebCT using TAM, *Journal of Computers & Education*, Vol. 48, Pp. 250-267.
- Nie, J.Y. (2010) *Synthesis Lectures on Human Language Technologies*, Vol. 3, No. 1, Pp. 1-125.
- Nofal, M. and Ahmad, S.B. (2010) Classification based on association-rule mining techniques: a general survey and empirical comparative evaluation, *Ubiquitous Computing and Communication Journal*, Vol. 5, No. 3, Pp. 9-17.
- Oard, D. (1997) Alternative approaches for cross-language text retrieval, *AAAI Spring Symposium on Cross Language Text and Speech Retrieval*, Pp. 1-8.
- Oard, D.W. and Diekema, A.R. (1999) Cross-language information retrieval, *Annual Review of Information Science and Technology*, Vol. 33, Pp. 223–256.

- Oard, D.W. and Dorr, B.J. (1996) A survey of multilingual text retrieval, Technical Report. University of Maryland at College Park, College Park, MD, USA.
- Ogden, D., Cowie, J., Davis, M., Ludovik, E., Molina-Salado, H. and Shin, H. (1999) Getting Information from Documents You Cannot Read: An Interactive Cross-Language Text Retrieval and Summarization System, Joint ACM Digital Library/SIGIR Workshop on Multilingual Information Discovery and Access (MIDAS).
- Okpor, M.D. (2014) Machine Translation approaches: Issues and challenges, International Journal of Computer science Issues, Vol.11, Issue 5, No.2, Pp. 159-165.
- Park, D.C. (2010) Image Classification Using Partitioned-Feature based Classifier Model, International Conference on Computer Systems and Applications (AICCSA), IEEE/ACS, Pp.1-6.
- Patel, H.N. and Virparia P.V. (2011) A Small Vocabulary Speech Recognition for Gujarati, International Journal of Advanced Research in Computer Science, Vol. 2, Issue 1, Pp. 208-210.
- Piternick, A.B. (1984) Searching Vocabularies: A Developing Category of Online Search Tools, Online Review, Vol.8, No.5, Pp.441-449.
- Plauche, M., Nallasamy, U., Pal, J., Wooters, C. and Ramachandran, D. (2006) Speech recognition for illiterate access to information and technology, International Conference on Information and Communication Technologies and Development, Pp. 83-92.
- Pohlmann, R. and Kraaij, W. (1997) The Effect of Syntactic Phrase Indexing on Retrieval Performance for Dutch Texts, Proceedings of RIAO, Pp. 176-187.
- Prathima, M. (2009) Text categorization based on Apriori algorithm's frequent itemsets, MS Thesis, University of Nevada, Las Vegas, Pp. 1-71.
- Pravin, P. and Jethva, H. (2013) Neural Network Based Gujarati Language Speech Recognition, International Journal of Computer Science and Management Research, Vol 2, Issue 5, Pp.2623-2627.
- Puspha, N., Revathi, R., Ramya, C. and Shahul, H.S. (2014) speech processing of Tamil language with Back Propagation Neural Network and Semi supervised training, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Special Issue 1, Pp. 2718-2723.
- Puurula, a. (2013) Cumulative progress in language models for information retrieval, Proceedings of Australasian Language Technology Association Workshop, Pp. 96-100.
- Quinlan, J.R. (1993) C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann, San Francisco.
- Quinlan, J.R. (1996) Bagging, boosting and c4.5, Proceedings of the 13th National Conference on Artificial Intelligence, Denver, CO, AAAI/MIT Press, Pp. 725-730.

- Rabiner, L. and Juang, B.H. (1993) Fundamentals of Speech Recognition (Prentice-Hall Inc., Englewood Cliffs, NJ).
- Rabiner, L. and Schafer, R.W. (2007) Introduction to Digital Speech Processing (Foundations and trends in Signal Processing), Now Publishers Inc., Netherlands.
- Rajput, N., Kumar, M. and Verma, A. (2004) A large-vocabulary continuous speech recognition system for Hindi, IBM Journal for Research and Development, Vol. 48, Issue 5.6, Pp. 703 - 715.
- Rao, P.R.K. and Devi, S.L. (2010) AU-KBC FIRE 2010 Submission – Cross-lingual Information retrieval track: Tamil-English, Working notes of Forum for information retrieval track: Tamil-English, Gujarat, India.
- Raza, A., Hussain, S., Sarfraz, H., Ullah, I. and Sarfraz, Z. (2009) An ASR System for Spontaneous Urdu Speech, Proceedings of O-COCOSDA, IEEE Explore, Pp. 1-6.
- Reddy, M.V. and Hanumanthappa, M. (2013) Indic Language Machine Translation Tool: English to Kannada/Telugu, Multimedia Processing, Communication and Computing Applications, Lecture Notes in Electrical Engineering, Vol. 213, Pp.35-49
- Ren, F. and Bracewell, D.B. (2009) Advanced Information Retrieval, Electronic Notes in Theoretical Computer Science, Vol.225, Pp.303–317.
- Renganathan, V. (2002) An interactive approach to development of English to Tamil machine translation system on the web, INFITT, (TI2002), Pp. 1-6.
- Rex, D.N. and Thomas B.M. (1978) Automatic Speech and Speaker Recognition, IEEE Press, New York.
- Reynolds, D.A. and Rose, R.C. (1995) Robust text-independent speaker identification using Gaussian mixture speaker models, IEEE Transactions on speech and Audio Processing, Vol. 3, No. 1, Pp. 72-83.
- Ripley, B.D. (2008) Pattern Recognition and Neural Networks, Cambridge University Press, First Edition.
- Rivas, A.R., Iglesias, E.L. and Borrajo, L. (2014) Study of Query Expansion Techniques and Their Application in the Biomedical Information Retrieval, The Scientific World Journal, Vol. 2014, Article ID 132158, Pp. 1-10.
- Robert, B. (1995) Operating Computers by Voice, Accent on Living, Vol.39, Pp.78.
- Robertson, S. and Zaragoza, H. (2009) The probabilistic relevance framework: BM25 and beyond, Foundations and Trends in Information Retrieval, Vol. 3, No. 4, Pp. 333–389.
- Robertson, S.E. and Jones, K.S. (1976) Relevance weighting of search terms, Journal of the American Society for Information Science, Vol. 27, No. 3, Pp. 129–146.

- Robertson, S.E. and Walker, S. (1994) Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval, Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin Ireland, Pp. 232–24.
- Robertson, S.E., Walker, S., Hancock-Beaulieu, M., Gull, A. and Lau, M. (1992) Okapi at TREC, D. K. Harman, editor, Proceedings of Text Retrieval Conference (TREC), Gaithersburg, MD, Pp.21-30, NIST Special Publication 500-207.
- Robertson, S.E., Walker, S., Jones, S., Hancock-Beaulieu, M.M and Gatford, M. (1994) Okapi at TREC-3, D. K. Harman, editor, Proceedings of Third Text Retrieval Conference (TREC), Gaithersburg, MD, Pp. 109–126, NIST Special Publication 500-225.
- Robin, R. (2012) Introduction to Tamil Language, Articles on Natural Language Processing, <http://language.worldofcomputing.net/tamil/introduction-to-tamil-language.html>, Last Accessed During August, 2015.
- Rosenberg, A.E., Lee, C.H. and Soong, F.K. (1994) Cepstral channel normalization techniques for hmm-based speaker verification, Proc ICSLP, Pp. 1835–1838.
- Saiprasad, P.S. and Girija, P.N. (2003) Speech Recognition of Isolated Telugu Vowels Using Neural Networks, Proceedings of the 1st Indian International Conference on Artificial Intelligence, India, Pp. 28-34.
- Saiva Manuscript in Pondicherry (2012) UNESCO, <http://www.unesco.org/new/en/communication-and-information/flagship-project-activities/memory-of-the-world/homepage>, Last Accessed During August, 2015.
- Salton, G. (1970) Automatic processing of foreign language documents, Journal of the American Society for Information Science, Vol.21, No.3, Pp. 187–194.
- Salton, G. and Michael J. (1983) Introduction to Modern Information Retrieval, McGraw Hill Book Co., New York.
- Sampath, K.M.P., Jayathilake, P.W.D.C., Ramanan, R., Fernando, S. and Sila, C.D. (2010) Speech recognition using neural networks, University of Moratuwa, Sri Lanka, <http://docslide.us/documents/speech-recognition-using-neural-network.html>, Last Accessed During August, 2015, Pp. 1-30.
- Samudravijaya, K. (2000) Computer Recognition of Spoken Hindi, Proceeding of International Conference of Speech, Music and Allied Signal Processing, Triruvananthapuram, Pp. 8-13.
- Saraswathi, S. and Geetha, T.V. (2004) Building language models for Tamil speech recognition system, Applied Computing, Lecture Notes in Computer Science, Vol. 3285, Pp 161-168.
- Saraswathi, S., Siddhiqaa, A. M., Kalaimagal, K. and Kalaiyarasi M. (2010) BiLingual Information Retrieval System for English and Tamil, Journal Of Computing, Vol.2, No.4, Pp.85-89.

Saravanan S., Menon A.G. and Soman K.P. (2010) English to Tamil Machine Translation System, Chemmozhi Maanaad, INFITT, Coimbatore.

Saravanan, K., Udupa, R. and Kumaran, A. (2011) Improving Cross-Language Information Retrieval by Transliteration Mining and Generation, Proceedings of Third International Workshop on Forum for Information Retrieval and Evaluation, Bombay, India.

Sarfraz, H., Hussain, S., Bokhari, R., Raza, A.A., Ullah, I., Sarfraz, Z., Pervez, S., Mustafa, A., Javed, I. and Parveen, R. (2010) Large Vocabulary Continuous Speech Recognition for Urdu, Proceedings of 8<sup>th</sup> International Conference on Frontiers of Information Technology, ACM, Islamabad, Article No. 1, Pp. 1-5.

Sarma, M.P. and Sarma, K.K. (2010) Speech Recognition of Assamese Numerals using combinations of LPC-features and heterogeneous ANNs, Proceedings of International Conference on Advances in Information and Communication Technologies (ICT 2010), Kochi, Kerala, India, Pp.8-11.

Sarma, M.P. and Sarma, K.K. (2011) Assamese Numeral Speech Recognition using Multiple Features and Co operative LVQ – Architectures, International Journal of Electrical and Electronics, Vol. 5, No. 1, Pp. 792-802.

Schapire, R.E., Freund, Y., Bartlett, P. and Lee, W.S. (1998) Boosting the margin: a new explanation for the effectiveness of voting methods, Ann. Stat., Vol. 26, No. 5, Pp. 1651–1686.

Scharenborg, O. (2007) Reaching Over the Gap: A Review of Efforts to Link Human and Automatic Speech Recognition Research, Speech Communication, Vol. 49, Issue 5, Pp. 336–347.

Schutte, K.T. (2009) Parts-based Models and Local Features for Automatic Speech recognition, PhD Thesis in Electrical Engineering and Computer Science, B.S., University of Illinois at Urbana-Champaign

Sculley, D. (2010) Web-scale k-means clustering, *Proceedings of the 19th international conference on World Wide Web*, ACM, Pp.1177–1178.

Selvam, M., and Natarajan, A.M. (2009) Improvement of rule-based morphological analysis and POS tagging in Tamil language via projection and induction technique, International journal of computers, Vol.3, Issue 4, Pp.357-367.

Sharma, N. and Niranjana, S. (2012) Optimization of Word Sense Disambiguation Using Clustering in Weka, International journal of computer technology and applications, Vol.3, Issue 4, Pp.1598-1604.

Sharma, P. (2012) Automatic identification of silence, voiced and unvoiced chunks in speech, A Thesis Submitted in Partial Fulfilment of the Requirements for the Award of the Degree of Master of Technology, School of Mathematics and Computer Applications, Thapar University, Patiala, India.

- Sharma, R. and Pyara, V.P. (2013) A Robust Denoising Algorithm for Sounds of Musical Instruments Using Wavelet Packet Transform, *Circuits and Systems*, Vol. , Pp. 459-465
- Sheridan, P., Wechsler, M. and Schauble, P. (1997) Crosslanguage speech retrieval: Establishing a baseline performance, *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Pp. 99–108.
- Shivakumar, C. (2012) A Simple approach for Speech Recognition: Kannada Digits Speech Recognition for Permanent Deaf Patients, Lap Lambert Academic Publishing.
- Shrawankar, U. and Thakare, V. (2010) Noise Estimation and Noise Removal Techniques for Speech Recognition in Adverse Environment, Zhongzhi Shi, Sunil Vadera, Agnar Aamodt, David Leake, *Intelligent Information Processing, IFIP Advances in Information and Communication Technology*, Springer, Vol.340, Pp.336-342.
- Siafarikas, M., Ganchev, T. and Fakotakis, N. (2004) Wavelet packet based speaker verification, *The Speaker and Language Recognition Workshop*. Toledo, Spain.
- Singhal, A. and Pereira, F. (1999) Document expansion for speech retrieval, *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Pp. 34–41.
- Sivaraman, G. and Samudravijaya, K. (2011) Hindi Speech Recognition and Online Speaker Adaptation, *International Conference on Technology Systems and Management*, Proceedings published by International Journal of Computer Applications, Pp. 27-30.
- Smeaton, A.F., O'Donnell, R. and Kellely, F. (1994) Indexing Structures Derived from Syntax in TREC-3: System Description, *The Third Text RETrieval Conference (TREC-3)*, TREC '94, Pp. 55–67.
- Sobha, L., Pralayankar, P. and Kavitha, V. (2009) Case Marking Pattern from Hindi to Tamil MT, *3rd National Conference on Recent Advances and Future Trends in IT (RAFIT)*, Punjab University, Patiala, Punjab.
- Sokolov, A., Hieber, F. and Riezler, S. (2014) Learning to translate queries for CLIR, *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (SIGIR '14)*, ACM, New York, Pp. 1179-1182.
- Sourabh, K. (2013) An Extensive Literature Review on CLIR and MT activities in India, *International Journal of Scientific & Engineering Research* Volume 4, Issue 2, Pp. 1-21.
- Srinivasan, S. and Petkovic, D. (2000) Phonetic confusion matrix based spoken document retrieval, *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Pp. 81–87.
- Stein, B. (1977) Circulation and the Historical Geography of Tamil Country, *The Journal of Asian Studies*, Vol. 37, No. 1, Pp. 7-26.

- Stephanie, S. (1997) Voice Recognition Systems: Part 2, Paraplegia News. Vol. 49, No. 9, Pp. 65–66.
- Straková, J. and Pecina, P. (2010) Czech Information Retrieval with Syntax-based Language Models, Proceedings of the Seventh International Conference on Language Resources and Evaluation, Pp. 1-4.
- Subramaniam, T.S (2011) Palani excavation triggers fresh debate, The Hindu, India, <http://www.thehindu.com/news/national/tamil-nadu/article2408091.ece>, Last Accessed During August, 2015.
- Subramanian, T.S. (2005) Rudimentary Tamil-Brahmi Script unearthed at Adichanallur, The Hindu, India, <http://www.thehindu.com/2005/02/17/stories/2005021704471300.htm>, Last Accessed During August, 2015.
- Subramanian, T.S. (2012) 2,200-year-old Tamil-Brahmi inscription found on Samanamalai, The Hindu, India, <http://www.thehindu.com/features/friday-review/history-and-culture/2200yearold-tamilbrahmi-inscription-found-on-samanamalai/article3220674.ece>, Last Accessed During August, 2015.
- Sukumar, A.R., Shah, A.F. and Anto, P.B. (2010) Isolated question words recognition from speech queries by using Artificial Neural Networks, proc. of IEEE 2nd International conference on Computing, Communication and Networking Technologies(ICCCNT), Pp. 1-4.
- Sumithra, A.M.G. and Thanushkodi, B.K. (2009) Performance evaluation of different thresholding methods in time adaptive wavelet based speech enhancement, IACSIT International Journal of Engineering and TEchnology, Vol.1, No. 5, Pp. 439-447.
- Sunitha, K.V.N. and Kalyani, N. (2012) Isolated Word Recognition using Morph Knowledge for Telugu Language, International Journal of Computer Applications, Vol.38, No.12, Pp.47-54.
- Sunny, S., David Peter, S. and Poulouse Jacob, K. (2012) Recognition of speech signals: an experimental comparison of linear predictive coding and discrete wavelet transforms, International Journal of Engineering Science & Technology, Vol. 4 Issue 4, Pp. 1594-1599.
- Sunny, S., Peter, D.S , and Jacob, K.P. (2013) A Comparative Study of Wavelet Based Featre Extraction Techniques in Recognizing Isolated Spoken Words, International Journal of Signal Processing Systems, Vol.1, No.1, Pp.49-53.
- Syama, R. and Idikkula, S.M. (2008) HMM Based Speech Recognition System for Malayalam, International Conference on Artificial Intelligence, Pp. 1-4.
- Tao, T., Wang, X., Mei, Q. and Zhai, C.X. (2006) Language model information retrieval with document expansion, Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL, Pp. 407–414.

Tazebay, M.V. and Akansu, A.N. (1995) Adaptive Subband Transforms in Time-frequency Excisers for DSSS Communications Systems, IEEE Trans. Signal Processing, Vol. 43, Pp. 2776-2782.

Thabtah, F.A. (2007) *A review of associative classification mining*, Knowledge Engineering Review, Vol. 22, No. 1, Pp. 37-65.

Thangarajan, R., Natarajan, A.M. and Selvam, M. (2008) Word and Triphone based Approaches in Continuous Speech Recognition for Tamil Language, WSEAS Transactions on Signal Processing, Vol. 4, Issue 3, Pp. 76-95.

Thangarajan, R., Natarajan, A.M. and Selvam, M. (2009) Syllable modelling in continuous speech recognition for Tamil language, International Journal of Speech Technology, Vol. 12, Pp.47 -57.

The I.A.S. Tamil Medical Manuscript Collection (2012) UNESCO, <http://www.unesco.org/new/en/communication-and-information/flagship-project-activities/memory-of-the-world/homepage>, Last Accessed During August, 2015.

Thenmozhi, D. and Aravindan, C. (2009) Tamil-English Cross Lingual Information Retrieval System for Agriculture Society, International Forum for Information Technology in Tamil Conference, Germany, Pp.173-178.

Thomas, L. (1998) Old Tamil, Steever, Sanford, The Dravidian Languages, London. Routledge, Pp. 75–99.

Thuan, N.D. (2010) Mining Time Pattern Association Rules in Temporal Database, Journal of Communication and Computer, Vol. 7, No. 3 (Serial No. 64), Pp. 50-56.

Trejo, V.C., Sidorov, G., Jimenez, S.M., Ibarra, M.M. and Martinez, R.C. (2015) Latent Dirichlet Allocation complement in the vector space model for Multi-Label Text Classification, International Journal of Combinatorial Optimization Problems and Informatics, Vol. 6, No. 1, Pp. 7-19.

Tseng, F.C. (2012) Mining frequent itemsets in large databases: The hierarchical partitioning approach, Expert Systems with Applications, Vol. 40, Issue 5, Pp. 1654-1661.

Ture, F. and Lin, J. (2014) Exploiting Representations from Statistical Machine Translation for Cross-Language Information Retrieval, ACM Transactions on Information Systems, Vol. 32, No. 4, Article 19, Pp. 1-32.

Udhyakumar, N., Swaminathan, R. and Ramakrishnan, S.K. (2004) Multilingual speech recognition for information retrieval in Indian context, Proceedings of the Student Research Workshop at HLT-NAACL, Pp. 1-6.

Usharani, N. and Girija, P.N. (2012) Error analysis to improve the speech recognition accuracy on Telugu language Sadhana, Indian Academy of Sciences, Vol. 37, Part 6, Pp. 747–761.

Utami, I.T., Bagus, S. and Kusman, S. (2014) Comparison of Single and Ensemble Classifiers of Support Vector Machine and Classification Tree, *Journal of Mathematical Sciences and Applications*, Vol. 2, Issue 2, Pp. 17-20.

Van Rijsbergen, C.J. (1979) *Information Retrieval*, Butterworths, 2<sup>nd</sup> Edition.

Vanopstal, K., Stichele, R.V., Laureys, G. and Buyschaert, J. (2010) Assessing the impact of English language skills and education level on PubMed searches by Dutch-speaking users, *Proceedings of the seventh International Conference on Language Resources and Evaluation*, European Language Resources Association, Malta, Pp. 2944-2948.

Vapnik, V. (1995) *The Nature of Statistical Learning Theory*. Springer, New York.

Varadarajan, M. (1988) *A history of Tamil literature*, Sahitya Akademi, New Delhi, Pp. 1-17.

VijaiBhaskar, P., Rama Mohan Rao, S. and Gopi, A. (2012) HTK Based Telugu Speech Recognition, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 2, Issue 12, Pp. 307-314.

Vijaya, M.S., Shivapratap, G. and Soman K.P. (2010) English to Tamil Transliteration using One Class Support Vector Machine, *International Journal of Applied Engineering Research*, Vol.5, No. 4, Pp.641-652.

Vimala, C. and Krishnaveni, M. (2012) Continuous Speech Recognition system for Tamil language using monophone-based Hidden Markov Model , *Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology CCSEIT '12* , Pp. 227-231.

Vimala, C. and Radha, V. (2012) Speaker Independent Isolated Speech Recognition System for Tamil Language using HMM, *Procedia Engineering*, Vol. 30, Pp. 1097-1102.

Voorhees, E.M. (2000) The TREC-8 question answering track report, *Proceedings of the Eighth Text Retrieval Conference (TREC-8)*, NIST Special Publication, Pp. 77-82.

Voorhees, E.M. and Harman, D. (1997) Overview of the Sixth Text Retrieval Conference (TREC-6), *Proceedings of TREC Conference*, Pp. 1–24.

Vries, A.D., Mamoulis, N. and Nes, N. (2002) Efficient k-NN search on vertically decomposed data, *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data*, Pp. 322–333.

Wang, S., Sun, J., Gao, B.J. and Ma, J. (2012) Adapting vector space model to ranking-based collaborative filtering, *Proceedings of the 21st ACM international conference on Information and knowledge management (CIKM '12)* ACM, New York, Pp. 1487-1491.

Wang, X. (2011) Chinese-English cross lingual information retrieval in biomedical using ontoloy-based query expansion, A Thesis submitted to the University of Manchester for the

Degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences, School of Computer Science.

Wechsler, M., Munteanu, E. and Schauble, P. (1998) New techniques for open-vocabulary spoken document retrieval, Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Pp. 20–27.

Weerasinghe, R. (2004) A statistical machine translation approach to Sinhala Tamil language translation, SCALLA, Pp. 1-7.

Weston, J. and Watkins, C. (1999) Multi-class support vector machines, Proceedings of ESANN99, M. Verleysen, Ed., Brussels, Belgium.

Whittaker, S., Hirschberg, J., Choi, J., Hindle, D., Pereira, F. and Singhal, A. (1999) SCAN: Designing and evaluating user interfaces to support retrieval from speech archives, Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Pp. 26–33.

[www4.ibm.com/software/speech/linux/dictation.html](http://www4.ibm.com/software/speech/linux/dictation.html) , Last Accessed During August, 2015.

[www.babeltech.com](http://www.babeltech.com) , Last Accessed During August, 2015..

[www.cse.ogi.edu/~omega/gnome/gvoice](http://www.cse.ogi.edu/~omega/gnome/gvoice), Last Accessed During August, 2015.

[www.isip.msstate.edu/project/speech](http://www.isip.msstate.edu/project/speech), Last Accessed During August, 2015.

[www.kieczka.de/daniel/linux/index.html](http://www.kieczka.de/daniel/linux/index.html), Last Accessed During August, 2015.

[www.nuance.com](http://www.nuance.com) ,Last Accessed During August, 2015.

[www.nuance.com/for-business/by-product/dragon/dragon-for-the-pc/dragon-professional](http://www.nuance.com/for-business/by-product/dragon/dragon-for-the-pc/dragon-professional), Last Accessed During August, 2015.

[www.softsound.com](http://www.softsound.com), Last Accessed During August, 2015.

[www.speech.cs.cmu.edu/sphinx/Sphinx.html](http://www.speech.cs.cmu.edu/sphinx/Sphinx.html) , Last Accessed During August, 2015.

[www.speech.kth.se/NICO/index.html](http://www.speech.kth.se/NICO/index.html) , Last Accessed During August, 2015.

Xinkai, W. (2011) Chinese-English Cross-Lingual Information Retrieval in Biomedicine Using Ontology-Based Query Expansion, Doctoral Thesis, University of Manchester

Xu, Z. and Akella, R. (2010) Improving probabilistic information retrieval by modelling burstiness of words, Information Process Management. Vol.46, No.2, Pp. 143-158.

Yang, H., Sloan, M. and Wang, J. (2014) Dynamic information retrieval modeling, Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14), ACM, New York, Pp. 1290-1290.

- Yarowsky, D. (1992) Word Sense Disambiguation using Statistical Models of Roget's Categories Trained on Large Corpora. Proceedings of COLING, Nantes, France, Pp. 454-460.
- Yarowsky, D. (1995) Unsupervised Word Sense Disambiguation rivalling Supervised methods, Association of Computational Linguistics (ACL), Cambridge, MA, Pp. 189-196.
- Yates, R.B. and Neto, B.R. (1999) Modern Information Retrieval. Addison-Wesley, New York.
- Yeom, G., Jo, T. and Yeom, Y. (2008) Table Based Single Pass Algorithm for Clustering News Articles in NewsPage.com, *Proceedings of the international conference on Computational Science and Its Applications, Part II (ICCSA '08)*, Pp. 1184-1193
- Yi, J., Zhang, L., Wang, J., Jin, R. and Jain, A.K. (2014) A single-pass algorithm for efficiently recovering sparse cluster centers of high-dimensional data, Proceedings of the 31st International Conference on Machine Learning, JMLR, Vol. 21, Pp. 1-9.
- Yin, X., and Han, J. (2003) CPAR: Classification based on predictive association rule, SDM 2003, San Francisco, CA.
- You, H. (2009) Robust Automatic Speech Recognition Algorithms for Dealing with Noise and Accent. Ph.D. Dissertation, University of California, Los Angeles, USA.
- Yu, W. and ZhengOu, W. (2007) A Fast KNN Algorithm for text categorization, Proceedings of the 6th International Conference on Machine Learning and Cybernetics, Hong Kong, Pp. 3436–3441.
- Zhai, C.Z. and Lafferty, J.D. (2004) A Study of Smoothing Methods for Language Models Applied to Information Retrieval, ACM Transactions on Information System (TOIS), Vol. 22, No.2, Pp.179–214.
- Zhao, J., Huang, J.X. and Ye, Z. (2014) Modelling Term Associations for Probabilistic Information Retrieval, ACM Transactions Information Systems, Vol.32, No.2, Article 7, 47.
- Zheng, Z. and Webb, G.I. (1998) stochastic attribute selection committees, Proceedings of the Australian Joint Conference on Artificial Intelligence, Berlin, Springer, Chapter 10, Pp. 321–332.
- Zhou, D., Truran, M., Brailsford, T., Wade, V. and Ashman, H. (2012) Translation techniques in cross-language information retrieval, ACM Computing Surveys, Vol. 45, No. 1, Article 1, Pp. 1-44.
- Zuccon, G., Azzopardi, L.A. and Rijsbergen, K.V. (2009) The quantum probability ranking principle for information retrieval, ICTIR, LNCS 5766, Springer-Verlag Berlin Heidelberg, Pp. 232-240.