References

[1] Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann publications, United States of America, 2006.

[2] Daniel t. Larose, *Data mining methods and models*, John Wiley & Sons, Inc, United States of America, 2006.

[3] Wesley Chu, Tsau Young Lin (Eds.), *Foundations and Advances in Data Mining*, vol 180, Springer-Verlag Berlin Heidelberg, 2005.

[4] David Taniar, *Research and Trends in Data Mining Technologies and Applications*, Idiea Group publication, 2007.

[5] David Kirk Evans, Judith L. Klavans, Kathleen R. McKeown, "Columbia Newsblaster: Multilingual News Summarization on the Web", Department of Computer Science, Columbia University, NY.

[6] Andreas Hotho, Alexander Maedche, Steffen Staab, "Ontology-based Text Document Clustering", DCAI 2001.

[7] Hannes Wettig, Jussi Lahtinen, Tuomas Lepola, Petri Myllym¨aki, Henry Tirri, "Bayesian Analysis of Online Newspaper Log Data", *Proceedings of the 2003 Symposium on Applications and the InternetWorkshops (SAINT 2003Workshops)*, IEEE Computer Society, Los Alamitos, California, 2003, Pp. 282–287.

[8] Sameh H. Ghwanmeh, "Applying Clustering of Hierarchical K-means-like Algorithm on Arabic Language", *International Journal Of Information Technology,* Volume 3 Number 1 2006 Issn 1305-2403.

[9] Noam Slonim, Naftali Tishby, "The Power of Word Clusters for Text Classification", *23rd European Colloquium on Information Retrieval Research*, 2001.

[10] Y.C. Fang, S. Parthasarathy, F. Schwartz, "Using Clustering to Boost Text Classification", *2001 IEEE International Conference on Data Mining*, Ohio State University, 2001.

[11] Peng Dai, Uri Iurgel, Gerhard Rigoll, "A Novel Feature Combination Approach for Spoken Document Classification with Support Vector Machines", University of Duisburg-Essen, Duisburg, Germany.

[12] Bei Yu, John Unsworth , "An Evaluation of Text Classification Methods for Literary Study", University of Illinois at Urbana-Champaign.

[13] L. Douglas Baker, Andrew Kachites McCallum, "Distributional Clustering of Words for Text Classification", *ACM SIGIR 98*, 1998.

[14] Marijus bernotas, Kazys Karklius, Remigijus Laurutis, Asta Slotkienė , "The Peculiarities Of The Text Document Representation, Using Ontology And Tagging-Based Clustering Technique ", *Information Technology And Control*, Vol.36, No.2, 2007.

[15] Gair, James W. Selected and Edited by Barbara C. Lust, *Studies in South Asian Linguistics: Sinhala and Other South Asian Languages*, New York / Oxford: Oxford University, 1998.

[16] Gair, James W., *Sinhala - The Indo-Aryan Languages*, George Cardona & Dhanesh Jain (eds.), London / New York: Routledge (Routledge Language Family Series), 2003.

[17] Aleksander Kotcz, Joshua Alspetor, "SVM-based filtering of e-mail spam with content specific misclassification costs", *2001 IEEE International Conference on Data Mining*, 2001.

[18] Bjornar Larsen, Chinatsu Aone, "Fast and effective text mining using linear-time document clustering", *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pg: 16 - 22 , ISBN:1-58113-143-7 , 1999.

[19] Catherine Blake, Wanda Pratt, "Better rules, fewer features: A semantic approach to selecting features from text", *2001 IEEE International Conference on Data Mining*.

[20] Cheng Xiang Zhai, Atulya Velivelli, Bei Yu, "A cross-collection mixture model for comparative text mining", *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pg: 743 – 748, 2004 , ISBN:1-58113-888-1 .

[21] Chid Apte, Fred Damerau, Sholom Weiss, "Text Mining with Decision Rules and Decision Trees", IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY 10598.

[22] Julia Itskevitch, "Automatic hierarchical e-mail classification using association rules ", Belorussian State Polytechnic Academy, 1997.

[23] Marko Grobelnik, Dunja Mladenic, Natasa Milic-Frayling, "Text Mining as Integration of Several Related Research Areas: Report on KDD'2000 Workshop on Text Mining", *ACM SIGKDD*, 2000.

[24] Waikato Environment for Knowledge Analysis (WEKA), version 3.4.11, University of Waikato, New Zealand.

[25] Prof. J.B. Dissanyake, "Some Salient Features of the Sinhala Alphabet".

[26] Gihan V. Dias, "Challenges of enabling IT in the Sinhala Language", *27th Internationalization and Unicode Conference*, Berlin, Germany, April 2005.

[27] Gihan Dias, Aruni Goonetilleke, "Development of Standards for Sinhala Computing", *1st Regional Conference on ICT and E-Paradigms*, Sri Lanka, 2004.

[28] Samaranayake V. K., Nandasara S. T., Dissanayake J. B., Weerasinghe A.R., Wijayawardhana H., *An Introduction to UNICODE for Sinhala Characters*, University of Colombo School of Computing, 2003.

[29] Gihan Dias, "Using IT in Local Languages".

[30] Gihan Dias, "Representation of Sinhala in Unicode".

[31] Muthu Nedumaran, "Sinhala Unicode developer workshop".

[32] Yuen-Hsien Tseng , *"FJU Test Collection for Evaluation of Chinese Text Categorization"*, 2004.

[33] Text mining tool, Magenta Technology text understanding approach, available at http://www.magenta-technology.com/en/technology/

[34] Heikki Hyötyniemi**, "**Text Document Classification with Self-Organizing Maps", *Proceedings of STeP'96*. Jarmo Alander, Timo Honkela and Matti Jakobsson (eds.), Publications of the Finnish Artificial Intelligence Society, pp. 64-72.

[35] Honkela T., Kaski S., Lagus K., Kohonen T., *Newsgroup Exploration with WEBSOM Method and Browsing Interface,* Helsinki University of Technology, Report A32, 1996.

[36] Krista Lagus, Timo Honkela, Samuel Kaski, Teuvo Kohonen, "WEBSOM - A Status Report", *Proceedings of STeP'96*. Jarmo Alander, Timo Honkela and Matti Jakobsson (eds.), Publications of the Finnish Artificial Intelligence Society, pp. 73-78.

[37]   Kanoksri Sarinnapakorn, Miroslav Kubat, "Combining Sub classifiers in Text Categorization: A DST-Based Solution and A Case Study", *IEEE transactions on knowledge and data engineering*, vol. 19, no. 12, December 2007.

[38]   Ji He, Ah-Hwee Tan, Chew-Lim Tan, "A Comparative Study on Chinese Text Categorization Methods", A-H. Tan, P.Yu (Eds), PRICAI 2000 workshop on text and web mining, Melbourne, pp. 24-35, August 2000.

[39]   Fotis Lazarinis, Jesus Vilares Ferro, John Tait, "Improving Non-English Web Searching" (iNEWS07), SIGIR 2007 workshop report, ACM SIGIR Forum, Vol. 41 No. 2, December 2007.

[40]   Andrea L. Houston, Kenneth R. Walsh, "Using an AI-Based Tool to Categorize Digitized Textual Forms of Organizational Memory", *Proceedings of the 29th Ann& Hawaii International Conference on System Sciences (HICSS-29)* – 1996, 1060-3425/96 IEEE.

[41]   Bled, Slovenia, "Machine Learning in Text Data Analysis", ICML-99 Workshop, June 30, 1999.

[42]   *"Natural Language Computing",* http://research.microsoft.com/research/pub/ China - Natural Language Computing - Home.mht.

[43]   Shan Chen, Damminda Alahakoon, Maria Indrawan, *Background knowledge drive ontology discovery*, Monash University.

[44]   Chung-Hong Lee, Hsin-Chang Yang, Sheng-Min Ma , "A Novel Multilingual Text Categorization System using Latent Semantic Indexing", *First International Conference on Innovative Computing, Information and Contro,* Volume II (ICICIC'06), pp. 503-506.

[45]   Guo D., Berry M.W., Thompson B.B., Bailin S., "Knowledge Enhanced Latent Semantic Indexing", *Information Retrieval*, Volume 6, Number 2, April 2003, pp.225-250(26).

[46]   Mark Girolami, A. Kaban, "On an Equivalence between PLSI and LDA", *Proceedings of SIGIR 2003*, 2003.

[47]   Dean Wright, *Integrating Language Identification with Text Classification,* UMBC, 2004.

[48]   Daniel Boley, Maria Gini, Kyle Hastings, Bamshad Mobasher, Jerry Moore, "A client-side Web agent for document categorization", *Internet Research: Electronic Networking Applications and Policy*, Volume 8 · Number 5 · 1998 · pp. 387–399, © MCB University Press · ISSN 1066-2243.

[49]   Toni Giorgino, "An Introduction to Text Classification", 2004.

[50]   "Sinhala hodiya", URL http://www.LANandWAN.com/Sinhala/hoodiya.htm

[51]   N. Jennings, K. Sycara, M. Wooldridge, "A Roadmap of Agent Research and Development", *Autonomous Agents and Multi-Agent Systems*, 1, 275–306, Kluwer Academic Publishers, Boston, 1998.

[52]   Michael Chau, Daniel Zeng, Hsinchun Chen, Michael Huang, David Hendriawan, "Design and evaluation of a multi-agent collaborative Web mining system".

[53]   Guilherme Bittencourt, Frederico L. G. Freitas, "An Ontology-based Architecture for Cooperative Information Agents".

[54]   D.A.Meedeniya, A.S.Perera, "A Comparative Study on Data Representation to Categorize Text Documents", *20th International Conference on Software Engineering and Knowledge Engineering (SEKE'08)*, 2008.

[55]   The UCI KDD Archive, Available at
        http://kdd.ics.uci.edu/databases/nsfabs/nsfabs.html   and
        http://kdd.ics.uci.edu/databases/reuters_transcribed/ reuters_transcribed.html.

[56]   Helena Ahonen-Myka, "Processing of Large Document Collections", University of Helsinki.

[57]   David L., Olson Dursun Delen, *Advanced Data Mining Techniques,* Springer-Verlag Berlin Heidelberg, USA, 2008. ISBN: 978-3-540-76916-3

[58]   Ian H.Witten, Eibe Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Second edition, Morgan Kaufmann Publishers, CA, 2005.

[59]   Pavel Praks, Jiri Dvorsky, Vaclav Snasel, "Latent Semantic Indexing for Image Retrieval Systems".

[60]   Landauer, T. K., Foltz, P. W., & Laham, D. ," Introduction to Latent Semantic Analysis, *Discourse Processes,* 25, 259-284, 1998.

[61]   Yogesh Raja - Shaogang Gong, **"**Gaussian Mixture Model (GMM)", Queen Mary and Westfield College, England.

[62]   Hichem Sahbi, "A Particular Gaussian Mixture Model for Clustering", Cambridge University, UK.

[63]   Erik Norvell, "Gaussian Mixture Model Based Audio Coding in a Perceptual Domain", 2005.

[64]   Amos Storkey," Learning from Data: Density Estimation - Gaussian Distribution", School of Informatics.

[65] Samy Bengio, "Statistical Machine Learning from Data : Gaussian Mixture Models", 2006.

[66] Kuei-Hsien, "K Means Clustering , Nearest Cluster and Gaussian Mixture", 2005.

[67] "A tutorial on Clustering Algorithms", available at http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/indix.html

[68] *Matlab: the language of technical computing* – Help

[69] Richard J.Roiger, Michael W.Geatz, *Data Mining: A Tutorial-based Primer*, Pearson Educations, 2005.

[70] D.A.Meedeniya, A.S.Perera, "An Adaptive Technique to Categorize Indic Language Documents", *International Conference on Advanced Computing Technologies (ICACT '08)*, India, 2008.

[71] Thomas K Landauer, Susan Dumais, "Latent semantic analysis".

[72] Sinhala text document corpus, publicly available at http://vijayaba.cse.mrt.ac.lk/~dulani/