References

[1] Bhardwaj, Aanshi, and Veenu Mangat, "A novel approach for content extraction from web pages", *Recent Advances in Engineering and Computational Sciences (RAECS),* 2014, pp. 1-4.

[2] Fayyad U M, "Data mining and knowledge discovery: Making sense out of data", *IEEE Intelligent Systems* , vol.11,(5), 2007, pp: 20-25

[3] "Machine learning: neural networks, genetic algorithms, and fuzzy systems", John Wiley & Sons Inc., 2007

[4] Blumberg, Robert and Shaku Atre, "The problem with unstructured data", DM REVIEW 13, 2008, pp: 42-49

[5] "The Theory and practice of online learning", Athabasca University Press, 2004

[6] Esra Saraç and Selma Ayşe Özel, "Web Page Classification Using Firefly Optimization", *IEEE International Symposium on Innovations in Intelligent Systems and Applications,* 2013, pp: 1-5.

[7] MD. Ezaz Ahmed and Preeti Bansal, "Clustering Technique on Search Engine Dataset using Data Mining Tool", *Third International Conference on Advanced Computing & Communication Technologies*, 2013, pp: 86 – 89.

[8] M´arius Sajgal´ık, Michal Barla, M´aria Bielikov´a, "From ambiguous words to key-concept extraction", *24th International Workshop on Database and Expert Systems Applications,* 2013, pp: 63-67.

[9] R Malhotra and A. Sharma, "A Neuro-Fuzzy Classifier for Website Quality Prediction", *International Conference on Advances in Computing, Communications and Informatics (ICACCI),* 2013, pp: 1274 – 1279.

[10] Tao Yue, Jianhua Sun and Hao Chen, "Fine Grained Mining and Classification of malicious web pages, *Fourth International Conference on Digital Manufacturing and Automation (ICDMA),* 2013, pp: 616 – 619.

[11] Sheau-Ling Hsieh, Wen-Yung Chang, Chi-Huang Chen and Yung-Ching Weng, "Semantic Similarity Measures in the Biomedical Domain by Leveraging a Web Search

Engine", *IEEE Journal of Biomedical and Health Informatics*, Volume: 17, Issue: 4, 2013, pp: 853 – 861.

[12] Qingjie Meng and Changqing Gong, "Research of Web Information Classifying Based on Neural Network", *6th International Conference on Information Management, Innovation Management and Industrial Engineering*, Volume: 1, 2013, pp: 625 – 628.

[13] Jingtian Jiang, Xinying Song, Nenghai Yu and Chin-Yew Lin, "FoCUS: Learning to Crawl Web Forums", *IEEE Transactions on Knowledge and Data Engineering*, Volume 25, 2013, pp : 1293-1306.

[14] Peng Wang, Mingqi Zhou, Yue You and Xiang Zhang, "A New Vision-Based Method for Extracting Academic Information from Conference Web Pages", *IEEE 24th International Conference on Tools with Artificial Intelligence*, Volume: 1, 2012, pp: 976 – 981.

[15] A Pakgohar and M Khalili, "A Probabilistic Relational Model for Keyword Extraction", *International Conference on Statistics in Science, Business, and Engineering (ICSSBE)*, 2012, pp: 1-5.

[16] J Krutil, M Kudelka and V Snasel, "Web Page Classification based on Schema.org Collection", *Fourth International Conference on Computational Aspects of Social Networks (CASoN)*, 2012, pp : 356 – 360.

[17] S Samarawickrama, L Jayaratne, "Effect of Named Entities in Web Page Classification", *Fourth International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM)*, 2012, pp: 38 – 42.

[18] Zhaohui Xu, Fuliang Yan, Jie Qin and Haifeng Zhu, "A Web Page Classification Algorithm Based On Link Information", *Tenth International Symposium on Distributed Computing and Applications to Business, Engineering and Science* , 2011, pp: 82 – 86.

[19] N Sato, K Komiya, K Fujimoto, Y Kotani, "Categorization of Product Pages Depending on Information on the Web", *Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE)*,2011, pp: 393 – 398.

[20] Wang Zhixing and Chen Shaohong, "Web Page Classification based on Semi-supervised Naive Bayesian EM Algorithm", *IEEE 3rd International Conference on Communication Software and Networks (ICCSN)*, 2011, pp: 242 – 245.

[21] D Taylan, M Poyraz, S Akyokus, M C Ganiz, "Intelligent Focused Crawler: Learning which Links to Crawl", *International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*,2011, pp: 504 – 508.

[22] S A Ozel, "A Genetic Algorithm Based Optimal Feature Selection for Web Page Classification", *International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, 2011, pp: 282 – 286.

[23] Van Lam Le, I Welch, Xiaoying Gao and P Komisarczuk, "Two-Stage Classification Model to Detect Malicious Web Pages", *International Conference on Advanced Information Networking and Applications*,2011, pp: 113 – 120.

[24] C Boden, T Hafele and A Loser, "Classification Algorithms for Relation Prediction", *IEEE 27th International Conference on Data Engineering Workshops (ICDEW)*, 2011, *pp:* 46 – 52.

[25] Qingjie Meng and Changqing Gong, "Web Information Classifying and Navigation based on Neural Network", *Second International Conference on Signal Processing Systems (ICSPS)*, Volume: 2, 2010, pp: 431-433.

[26] Boyi Xu, Jing Wang and Hongming Cai, "A Web Page Classification Algorithm and Its Application in E-government System", *Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010)*, Volume: 4, 2010, pp: 1767 – 1771.

[27] R J Oskouei,"Identifying Students' Behaviors Related to Internet UsagePatterns", *International Conference on Technology for Education (T4E)*, 2010, pp: 232-233.

[28] Daniela Xhemali, Christopher J. Hinde and Roger G Stone, "Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages", *IJCSI International Journal of Computer Science Issues*, Vol. 4, No. 1, 2009

[29] S Aliakbary, H Abolhassani, H Rahmani and B Nobakht, "Web Page Classification Using Social Tags", *International Conference on Computational Science and Engineering*, Volume: 4, 2009, pp: 588 – 593.

[30] Win Thanda Aung and Khin Hay Mar Saw Hla, "Random Forest Classifier for Multi-category Classification of Web Pages", *IEEE Asia-Pacific Services Computing Conference*, 2009, pp: 372 – 376.

[31] Dongwon Lee, Hung-sik Kim, Eun Kyung Kim, Su Yan, Johnny Chen and Jeongkyu Lee, "LeeDeo: Web-Crawled Academic Video Search Engine", *Tenth IEEE International Symposium on Multimedia*,2008, pp: 497 – 502.

[32] Jing Wang, Hongming Cai, Boyi Xu and Lihong Jiang, "CUCS: A Web Page Classification Algorithm for Large Training Set", *IFIP International Conference on Network and Parallel Computing*, 2008, pp: 440 – 445.

[33] Shiqun Yin, Fang Wang, Zhong Xie and Yuhui Qiu, "Study on Web-page Classification Algorithm Based on Rough Set Theory", *International Symposiums on Information Processing*, 2008, pp: 202 – 206.

[34] Hsinchun Chen, "IEDs in the Dark Web: Genre Classification of Improvised Explosive Device Web Pages", *IEEE International Conference on Intelligence and Security Informatics*, 2008, pp: 94 – 97.

[35] Rong Liu, ianzhong Zhou and Ming Liu, "A Graph-based Semi-supervised Learning Algorithm for Web Page Classification", *Sixth International Conference on Intelligent Systems Design and Applications*, Volume: 2, 2006, pp: 856 – 860.

[36] Mu-Hee Song, Soo-Yeon Lim, Dong-Jin Kang and Sang-Jo Lee, "Automatic Classification of Web Pages based on the Concept of Domain Ontology", *12th Asia-Pacific Software Engineering Conference*,2005

[37] Xiaogang Peng and B Choi, "Automated Web Page Classification in a Dynamic and Hierarchical Way", *IEEE International Conference on Data Mining*, 2002, pp: 386 – 393.

[38] "Introduction to Data Mining", Pearson Education, 2007

[39] "Data Mining – Concepts and Techniques", Elsevier, 2011

[40] "Massively Open: How Massive Open Online Courses Changed The World", CreateSpace Independent Publishing Platform, 2013

[41] "A Concise Introduction to Software Engineering", Springer – Verlog London Limited, 2008

[44] Jiawei Han, "Data mining techniques", *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, *ACM*, 2003, pp: 545

[45] "Applied logistic regression", John Wiley & Sons, 2004

[46] Panda, Mrutyunjaya, and Manas Ranjan Patra, "Network intrusion detection using naive bayes", *International Journal of Computer Science and Network Security*, 7.12, 2007, pp: 258-263

[47] Cunningham, Padraig, and Sarah Jane Delany, "k-Nearest neighbour classifiers", *Mult Classif Syst*, 2007

[48] Leo Breiman, "Random Forests Machine Learning", vol.45, 2001

[49] "Semantics and verification of data flow in UML 2.0 activities", Electronic Notes in Theoretical Computer Science – Elsevier, 2005

[50] "A corpus of Late Modern English texts" International Computer Archive of Modern and Medieval English – Elsevier, 2005

[51] "Class and subclass declarations", Software pioneers - Springer Berlin Heidelberg, 2002

[52] "C++: The Complete Reference", Tata McGraw-Hill Education, 2008

[53] "Object Oriented Programming with C++", Tata McGraw-Hill Education, 2008

[54] "A systematic analysis of performance measures for classification tasks", Information Processing & Management - Elsevier, 2009

[55] "An Introduction to ROC Analysis", Pattern recognition- Elsevier, 2006