

{tag} International Journal of Computer Applications  
Foundation of Computer Science (FCS), NY, USA

[Volume 179](#)

-  
[Number 45](#)

Year of Publication: 2018

Authors:

Pooja Vinod Nainwani, Purvi Prajapati

10.5120/ijca2018916994

{bibtex}2018916994.bib{/bibtex}

### **Abstract**

Classification of Web pages is one of the challenging and important task as there is an increase in web pages in day to day life provided by internet. There are many ways of classifying web pages based on different approach and features. This paper explains some of the approaches and algorithms used for the classification of webpages. Web pages are allocated to pre-determined categories which is done mainly according to their content in Web page classification. The important technique for web mining is web page classification because classifying the web pages of interesting class is the initial step of data mining. The agenda of this paper is first to introduce the concepts related to web mining and then to provide a comprehensive review of different classification techniques.

### **References**

1. Yu H, Han J, Chang KC. PEBL: Web page classification without negative examples. IEEE Transactions on Knowledge and Data Engineering. 2004 Jan;16(1):70-81.

2. Fiol-Roig G, Miró-Julià M, Herraiz E. Data mining techniques for web page classification. In Highlights in Practical Applications of Agents and Multiagent Systems 2011 (pp. 61-68). Springer, Berlin, Heidelberg.
3. Nayak MA. A Comparative Study of Web Page Classification Techniques.
4. Qi X and Davison B.D. (2009) Web Page Classification: Features and Algorithms. ACM Computing Surveys, Vol. 41, No. 2, Article 12.
5. S. Markkandeyan1 · M. Indra Devi, "Efficient Machine Learning Technique for Web Page Classification".
6. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2016 Oct 1.
7. Kwon OW, Lee JH. Web page classification based on k-nearest neighbour approach. In Proceedings of the fifth international workshop on Information retrieval with Asian languages 2000 Nov 1 (pp. 9-15). ACM.
8. Patil AS, Pawar BV. Automated classification of web sites using Naive Bayesian algorithm. In Proceedings of the international multiconference of engineers and computer scientists 2012 Mar 14 (Vol. 1, pp. 519-523).
9. Kan MY, Thi HO. Fast webpage classification using URL features. In Proceedings of the 14th ACM international conference on Information and knowledge management 2005 Oct 31 (pp. 325-326). ACM.
10. Herrouz A, Khentout C, Djoudi M. Overview of web content mining tools. arXiv preprint arXiv:1307.1024. 2013 Jul 2.
11. Malarvizhi R, Saraswathi K. Web Content Mining Techniques Tools & Algorithms—A Comprehensive Study. International Journal of Computer Trends and Technology (IJCTT). 2013 Aug; 4(8):2940-5.
12. Aggarwal CC, Zhai C. A survey of text classification algorithms. In Mining text data 2012 (pp. 163-222). Springer, Boston, MA.
13. Chen H, Fuller SS, Friedman C, Hersh W. Knowledge management, data mining, and text mining in medical informatics. In Medical Informatics 2005 (pp. 3-33). Springer US.
14. Kavitha S, Vijaya MS. Web Page Categorization using Multilayer Perceptron with Reduced Features. International Journal of Computer Applications. 2013 Jan 1;65(1).
15. Marath ST, Shepherd M, Milios E, Duffy J. Large-scale web page classification. In System Sciences (HICSS), 2014 47th Hawaii International Conference on 2014 Jan 6 (pp. 1813-1822). IEEE.
16. Yuchang Lu. "Application of SVM in web page categorization" , 2006 IEEE International Conference on Granular Computing, 2006
17. Rongfang Bie. "Automatic web pages categorization with ReliefF and Hidden Naive Bayes" , Proceedings of the 2007 ACM symposium on Applied computing - SAC 07 SAC 07, 2007
18. Shaohong, Chen, and Wang Zhixing. "Web page classification based on Semi-supervised Naïve Bayesian EM algorithm", 2011 IEEE 3rd International Conference on Communication Software and Networks, 2011.
19. AbdulHussien AA. Comparison of Machine Learning Algorithms to Classify Web Pages.
20. Gabriel Fiol-Roig. "Data Mining Techniques for Web Page Classification", Advances in Intelligent and Soft Computing, 2011.

## Index Terms

Computer Science

Information Sciences

### **Keywords**

Web page classification, Web Mining, Data Mining, uniform resource locator (URL), SVM (Support Vector Machine), KNN (K-Nearest Neighbor), Naïve Bayes, Artificial Neural Network.