



Semantic Web mining for Content-Based Online Shopping Recommender Systems

Ibukun Tolulope Afolabi (Covenant University, Ota, Nigeria), Opeyemi Samuel Makinde (Covenant University, Ota, Nigeria) and Olufunke Oyejoke Oladipupo (Covenant University, Ota, Nigeria)

Source Title: International Journal of Intelligent Information Technologies (JIIT) (/gateway/journal/1089) 15(4) (/gateway/issue/200352)

Copyright: © 2019

Volume: 15

Issue: 4

Article: 3

Pages: 16

ISSN: 1548-3657

EISSN: 1548-3665

EISBN13: 9781522564355

DOI: 10.4018/IJIT.2019100103

[Cite Article](#)  [Favorite](#)  [Full-Issue Download](#) 

[View Full Text HTML](#) >

(/gateway/article/full-text-html/237965)

[View Full Text PDF](#) >

(/gateway/article/full-text-pdf/237965)

Abstract

Currently, for content-based recommendations, semantic analysis of text from webpages seems to be a major problem. In this research, we present a semantic web content mining approach for recommender systems in online shopping. The methodology is based on two major phases. The first phase is the semantic preprocessing of textual data using the combination of a developed ontology and an existing ontology. The second phase uses the Naïve Bayes algorithm to make the recommendations. The output of the system is evaluated using precision, recall and f-measure. The results from the system showed that the semantic preprocessing improved the recommendation accuracy of the recommender system by 5.2% over the existing approach. Also, the developed system is able to provide a platform for content-based recommendation in online shopping. This system has an edge over the existing recommender approaches because it is able to analyze the textual contents of users feedback on a product in order to provide the necessary product recommendation.

Request access from your librarian to read this article's full text.

Full Text Preview

1. Introduction

It is obvious that the importance of data in the 21st century cannot be over emphasized, it is due to this fact that application areas such as medicine, transportation, ecommerce, banking and many other areas, seek to harness its power for various reasons. Data can be structured or unstructured, when it is structured, it has a degree of organization that is readily searchable and quickly consolidates into information. Structured data is well defined, predictable, and managed by an elaborate infrastructure (Lee, 2017). As a rule, most units of data in the structured environment can be located very quickly and easily. Typically, structured data is managed by a database management system (DBMS) and consists of records, attributes, keys, and indexes. Unstructured data on the other hand are not organized in a specific manner and does not have a specific data type or orientation. Without preprocessing, storing this data type in a table is impossible. Examples are; call centre data, email, social media data and so on (Johnson & Kumar, 2012). Data mining involves recognizing patterns from large data set using various mathematical and machine learning methods such as Support Vector Machines, Classification, Artificial Neural Network (ANN), etc. (Xu, Zhang, & Li, 2011). Unstructured data mining is the practice of looking at relatively unstructured data and trying to get more refined data sets out of it. It regularly comprises of removing data from sources not generally utilized for data mining activities. In unstructured data mining, advancements would separate the data, searching for identifiers and bits of data. Finally, semi-structured data mining refers to data that does not exist in a relational database but rather has some hierarchical properties that make it less demanding to analyze. Mining semi-structured data includes finding and separating helpful information from semi-structured data sets. The first step of unstructured mining is the data processing and the techniques used are dependent on the type of data to be mined.

Web mining lies in between and copes with semi-structured data and/or unstructured data. Web mining calls for creative use of data mining and/or text mining techniques and its distinctive approaches. Mining the web data is one of the most challenging tasks for the data mining and data management scholars because there are huge heterogeneous, less structured data available on the web that one can easily get overwhelmed with (Zhang & Segall, 2008). The contents of data from the web may be a collection of facts that webpages are meant to contain, and these may consist of text, structured data such as lists and tables, and even images, video and audio. Therefore, web mining extract valuable information or patterns from the Web hyperlink structure, page content, and data usage. In recent years with the rapid progress of the internet, one can easily access various kinds of information on the World Wide Web (WWW). However, it has become very inconvenient for users to obtain relevant interesting information on the WWW as information resources on the WWW is growing rapidly and there is a large amount of redundant and irrelevant information on web pages. Web mining can be categorized into web usage mining, web structured mining and web content mining (Gunasundari & Eswari, 2013; Mihai, 2009). Web usage mining deals with extricating patterns and data from server logs to understand client or user actions including where they are from, the number of clients or users who clicked an item on the web page and the sort of activities being done on the website (Mishra & Tripathi, 2015). Web structure mining deals with analyzing the nodes and connection structure of a website using graph theory. There are two things that can be obtained from this: the structure of a website in terms of how it is connected to other sites and the document structure of the website itself, as to how each page is connected (Anami, Wadawadagi, & Pagi, 2014). Web content mining is the process of mining useful information from the contents of web pages and web documents, which are mostly text, images and audio/video files. Techniques used in this discipline have been heavily drawn from natural language processing (NLP) and information retrieval (Asvial, Budiyanoto, & Gunawan, 2014; Gunasundari & Eswari, 2013).

Continue Reading (</gateway/article/full-text-html/237965>)

References

- Agarwal, P., Vaithyanathan, R., Sharma, S., & Shroff, G. (2012). Catching the Long-Tail: Extracting Local News Events from Twitter. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media* (pp. 379–382). AAAI. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4639%5Cnhttp://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4639/5011> (<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4639%5Cnhttp://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4639/5011>)
- Follow Reference Ajoudanian S. Jazi M. D. (2009). Deep Web Content Mining. *World Academy of Science, Engineering and Technology*, 49(1), 501–505.
- Follow Reference Anagnostopoulos I. Razis G. Mylonas P. Anagnostopoulos C. N. (2015). Semantic query suggestion using Twitter Entities. *Neurocomputing*, 163(March), 137–150. 10.1016/j.neucom.2014.12.090
- Follow Reference Anami B. S. Wadawadagi R. S. Pagi V. B. (2014). Machine Learning Techniques in Web Content Mining: A Comparative Analysis. *Journal of Information & Knowledge Management*, 13(01), 1450005. 10.1142/S0219649214500051
- Follow Reference Asvial, M., Budiyanoto, S., & Gunawan, D. (2014). An intelligent load balancing and offloading in 3G - WiFi offload network using hybrid and distance vector algorithm. In *Proceedings of the IEEE Symposium on Wireless Technology and Applications, ISWTA* (pp. 36–40). IEEE. 10.1109/ISWTA.2014.6981191
- Follow Reference Bhardwaj, A., & Mangat, V. (2014a). A novel approach for content extraction from web pages. In *Proceedings of the 2014 Recent Advances in Engineering and Computational Sciences (RAECS)* (pp. 1-4). doi:10.1109/RAECS.2014.6799616
- Follow Reference Bhardwaj A. Mangat V. (2014b). An improvised algorithm for relevant content extraction from web pages. *Journal of Emerging Technologies in Web Intelligence*, 6(2), 226–230. 10.4304/jetwi.6.2.226-230
- Follow Reference Boratto L. Carta S. Fenu G. Saia R. (2017). Semantics-aware content-based recommender systems: Design and architecture guidelines. *Neurocomputing*, 254, 1339–1351. 10.1016/j.neucom.2016.10.079
- Follow Reference Bucur C. (2015). Using opinion mining techniques in tourism. *Procedia Economics and Finance*, 23, 1666–1673.
- Follow Reference Cai, D., Yu, S., Wen, J., & Ma, W. (2003). Extracting Content Structure for Web Pages Based on Visual Representation. In *Proceedings of the 5th Asia-Pacific Web Conference - APWeb '03* (pp. 406–417). Academic Press. 10.1007/3-540-36901-5_42
- Follow Reference Capelle, M., Frasinca, F., Moerland, M., & Hogenboom, F. (2012). Semantics-based news recommendation. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics - WIMS '12*. Academic Press. 10.1145/2254129.2254163
- Follow Reference Capelle, M., Hogenboom, F., Hogenboom, A., & Frasinca, F. (2013). Semantic news recommendation using wordnet and Bing similarities. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing - SAC '13*. Academic Press. 10.1145/2480362.2480426
- Follow Reference Dubey P. Nair P. S. (2015). Recommendation System for Web Mining: A Review. *International Journal of Computers and Applications*, 109(11), 1–6. 10.5120/19229-0941

- Follow Reference Dublin T. C. Boyce S. Pahl C. (2007). Developing domain ontologies for course content the development of ontologies. *Journal of Educational Technology & Society*, 10(3), 275–288. Retrieved from www.ifets.info/journals/10_3/19.pdf (http://www.ifets.info/journals/10_3/19.pdf)
- Ghani, R., Jones, R., Mladenec, D., Nigam, K., & Slattery, S. (2000, August). Data mining on symbolic knowledge extracted from the web. In *Workshop on Text Mining at the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Vol. 56). Retrieved from <https://pdfs.semanticscholar.org/235d/e838bab8bf6bb624416f793ba12f3f941f43.pdf> (<https://pdfs.semanticscholar.org/235d/e838bab8bf6bb624416f793ba12f3f941f43.pdf>)
- Follow Reference Gunasundari C. Eswari N. (2013). A Study of Web Content Mining. *International Journal of Innovative Research in Science & Engineering*, 3(6), 253–259.
- Follow Reference Guo, Y., Tang, H., Song, L., Wang, Y., & Ding, G. (2010). ECON: An approach to extract content from web news page. In *Advances in Web Technologies and Applications - Proceedings of the 12th Asia-Pacific Web Conference, APWeb 2010* (pp. 314-320). doi:10.1109/APWeb.2010.11
- Follow Reference Hoang T. B. N. Mothe J. (2018). Location extraction from tweets. *Information Processing & Management*, 54(2), 129–144. 10.1016/j.ipm.2017.11.001
- Follow Reference Huang, X., Gao, Y., Huang, L., Zhang, Z., Li, Y., Wang, F., & Kang, L. (2017). Web content extraction using clustering with web structure. In *Proceedings of the International Symposium on Neural Networks* (pp. 95-103). Academic Press. 10.1007/978-3-319-59072-1_12
- Follow Reference Johnson F. Kumar S. (2012). Web content mining techniques: A survey. *International Journal of Computers and Applications*, 47(11), 44–50. 10.5120/7236-0266
- Follow Reference Kakol M. Nielek R. Wierzbicki A. (2017). Understanding and predicting web content credibility using the content credibility corpus. *Information Processing & Management*, 53(5), 1043–1061. 10.1016/j.ipm.2017.04.003
- Follow Reference Kapadia N. R. Patel K. (2012). Web content mining techniques – A comprehensive survey. *International Journal of Research in Engineering and Applied Sciences*, 2(2), 1869–1877.
- Follow Reference Költringer C. Dickinger A. (2015). Analyzing destination branding and image from online sources: A web content mining approach. *Journal of Business Research*, 68(9), 1836–1843. 10.1016/j.jbusres.2015.01.011
- Follow Reference Lee I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293–303. 10.1016/j.bushor.2017.01.004
- Li, J., & Ezeife, C. I. (2006). Cleaning Web Pages for Effective Web Content Mining. In *Dexa 2006* (pp. 560-571). Springer. Retrieved from https://link.springer.com/chapter/10.1007/11827405_55 (https://link.springer.com/chapter/10.1007/11827405_55)
- Follow Reference Lihui C. Chue W. L. (2005). Using Web structure and summarisation techniques for Web content mining. *Information Processing & Management*, 41(5), 1225–1242. 10.1016/j.ipm.2004.08.003
- Follow Reference Liu Q. Shao M. Wu L. Zhao G. Fan G. Li J. (2017). Main content extraction from web pages based on node characteristics. *Journal of Computing Science and Engineering*, 11(2), 39–48. 10.5626/JCSE.2017.11.2.39
- Follow Reference Mathiyalagan M. (2014). Mining the Web for Plagiarized Web Content using Similarity Measure Technique. *International Journal for Research in Science Engineering and Technology*, 1(1), 32–38. Retrieved from <http://ijrset.in/20141205.pdf> (<http://ijrset.in/20141205.pdf>)
- Follow Reference Menaka C. Nagadeepa N. (2016). Clustering and labeling of images under web content mining. *International Journal of Advanced Research in Computer Science*, 7(5), 75–79.
- Follow Reference Mishra S. Tripathi S. P. (2015). An artificial intelligence approach for web mining. *International Journal of Research and Development in Applied Science and Engineering*, 2(1), 1–5.
- Follow Reference Musto C. Lops P. de Gemmis M. Semeraro G. (2017). Semantics-aware recommender systems exploiting linked open data and graph-based features. *Knowledge-Based Systems*, 136, 1–14. 10.1016/j.knosys.2017.08.015
- Follow Reference Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. *Stanford Knowledge Systems Laboratory*, 25. doi:10.1016/j.artmed.2004.01.014
- Follow Reference Prakash K. B. Rajaraman A. (2016). Mining of bilingual Indian web documents. *Procedia Computer Science*, 89, 514–520. 10.1016/j.procs.2016.06.103
- Follow Reference Rohilla P. Sharma O. (2015). Web content mining: An Implementation on social websites. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(7), 108–111. doi:10.17148/IJARCCCE.2015.4724
- Follow Reference Sambanthan K. T. Dhenakaran S. S. (2017). Optimized Web Content Mining. *Life Science Journal*, 14(2), 41–47. doi:10.7537/marslsj140217.05
- Follow Reference Schedl M. Widmer G. Knees P. Pohle T. (2011). A music information system automatically generated via Web content mining techniques. *Information Processing & Management*, 47(3), 426–439. 10.1016/j.ipm.2010.09.002
- Follow Reference Sethi M. Gaur G. (2015). Web mining techniques for extraction of news. *International Journal of Advancement in Engineering Technology, Management & Applied Science*, 2(4), 34–47.
- Follow Reference Sivakumar J. Ravichandran K. S. (2013). A review on semantic-based web mining and its applications. *IACSIT International Journal of Engineering and Technology*, 5(1), 186–192.
- Follow Reference Steinberger J. Jezek K. (2012). Evaluation Measures for Text Summarization. *Computer Information*, 28(2), 251–275.

- Follow Reference Uzun E. Agun H. V. Yerlikaya T. (2013). A hybrid approach for extracting informative content from web pages. *Information Processing & Management*, 49(4), 928–944. 10.1016/j.ipm.2013.02.005
- Follow Reference Vijayarani S. Ilamathi J. Nithya M. (2015). Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16. Retrieved from <http://www.ijcsn.com/Documents/Volumes/vol5issue1/ijcsn2015050102.pdf> (<http://www.ijcsn.com/Documents/Volumes/vol5issue1/ijcsn2015050102.pdf>)
- Follow Reference Wise A. F. Cui Y. Jin W. Q. Vytasek J. (2017). Mining for gold: Identifying content-related MOOC discussion threads across domains through linguistic modeling. *Internet and Higher Education*, 32, 11–28. 10.1016/j.iheduc.2016.08.001
- Follow Reference Wu Y. C. (2016). Language independent web news extraction system based on text detection framework. *Information Sciences*, 342, 132–149. 10.1016/j.ins.2015.12.025
- Follow Reference Xu, G., Zhang, Y., & Li, L. (2011). *Web Mining and Social Networking*. doi:10.1007/978-1-4419-7735-9
- Follow Reference Yulian, F., Zongwei, L., Yun, X., & Winston, Z. (2009). A semantic DOM approach for webpage information extraction. In *Proceedings - International Conference on Management and Service Science, MASS 2009*. Academic Press. doi:10.1109/ICMSS.2009.5302541
- Follow Reference Zhang B. Wang X. F. (2012). Content extraction from Chinese web page based on title and content dependency tree. *Journal of China Universities of Posts and Telecommunications*, 19(Suppl. 2), 147–151, 189. 10.1016/S1005-8885(11)60429-9
- Follow Reference Zhang Q. Segall R. S. (2008). Web Mining: A Survey of Current Research, Techniques, and Software. *International Journal of Information Technology & Decision Making*, 7(4), 683–720. 10.1142/S0219622008003150
- Follow Reference Zhang Y. Jiao J. (2007). An associative classification-based recommendation system for personalization in B2C e-commerce applications. *Expert Systems with Applications*, 33(2), 357–367. 10.1016/j.eswa.2006.05.005

Request Access

You do not own this content. Please login to recommend this title to your institution's librarian or purchase it from the IGI Global bookstore (</article/semantic-web-mining-for-content-based-online-shopping-recommender-systems/237965>).

Username or email:

Password:

[Log In >](#)

[Forgot individual login password? \(/gateway/login/reset-password/\)](/gateway/login/reset-password/)

[Create individual account \(/gateway/login/create-account/\)](/gateway/login/create-account/)

Research Tools

[Database Search \(/gateway/\)](#) | [Help \(/gateway/help/\)](#) | [User Guide \(/gateway/user-guide/\)](#) | [Advisory Board \(/gateway/advisory-board/\)](#)

User Resources

[Librarians \(/gateway/librarians/\)](#) | [Researchers \(/gateway/researchers/\)](#) | [Authors \(/gateway/authors/\)](#)

Librarian Tools

[COUNTER Reports \(/gateway/librarian-tools/counter-reports/\)](#) | [Persistent URLs \(/gateway/librarian-tools/persistent-urls/\)](#) | [MARC Records \(/gateway/librarian-tools/marc-records/\)](#) | [Institution Holdings \(/gateway/librarian-tools/institution-holdings/\)](#) | [Institution Settings \(/gateway/librarian-tools/institution-settings/\)](#)

Librarian Resources

[Training \(/gateway/librarian-corner/training/\)](#) | [Title Lists \(/gateway/librarian-corner/title-lists/\)](#) | [Licensing and Consortium Information \(/gateway/librarian-corner/licensing-and-consortium-information/\)](#) | [Promotions \(/gateway/librarian-corner/promotions/\)](#) | [Online Symposium Series \(/gateway/librarian-corner/online-symposium-series/\)](#)

Policies

[Terms and Conditions \(/gateway/terms-and-conditions/\)](#)

[http://www.facebook.com/pages/IGI-](http://www.facebook.com/pages/IGI-Global/138206739534176?ref=sgm)

[Global/138206739534176?ref=sgm](http://www.facebook.com/pages/IGI-Global/138206739534176?ref=sgm)

<http://twitter.com/igiglobal>

<https://www.linkedin.com/company/igiglobal>



<http://www.world-forgotten-children.org>

<https://publicationethics.org/category/publisher/igi-global>

Copyright © 1988-2021, IGI Global - All Rights Reserved