

Semantic educational data extraction using structural domain relationships

Manchikatla Srikanth *

Associate Professor Dept of CSE, Vaageswari College Of Engineering Hyderabad -Karimnagar Highway, Beside LMD Police Station, Ramakrishna Colony, Karimnagar, Telangana 505481

*Corresponding author E-mail: srikanthmanchikatla24@gmail.com

Abstract

In the mining industry', some of the domains are most popular and it plays a vital role in the specific area. Educational Mining and Web-Data Extraction are the two important factors play a leading role in mining industry. The main objective of the proposed system is to extract the related contents from web using semantic (relating to meaning in language or logic) principles as well as to allow the providers to dynamically generate the web pages for educational content and allow the users to search and extract the data from server based on content. The main model of this system is to illustrate the adaptive learning system. For demonstration we consider the semantic principles for Educational content over dynamic environment. This site allows the providers to create web pages related to educational content dynamically and this will be getting approved by the Administrator to live in process. Once the site is live the users can search for the exact content present into the site based on semantic principles. The proposed model is designed for dynamic web data extraction and content analysis from the extracted data due to educational principles. In the proposed system Semantic Web Extraction (SWE) procedures are highly analyzed and utilized for content manipulations. Energetic data extraction scheme for users based on educational content rather than header, title, meta tags and descriptions.

Keywords: Educational Content Analysis; Web Mining; Data Extraction; Semantic Principles; Content Based Data Extraction.

1. Introduction

Energetic Information/data Processing is the major task in data mining industry. Dynamic data extraction and web based content retrieval are the major terms to deal with this domain. In web-based-application scenarios there are numerous issues occurring over processing levels, they are: data-overloading, server-overloading, data-security and many more [1]. All these are handled and processed under single server is a challenging task over mining industry. Specifically, in educational domains the content management process is crucial because the number of students are increasing means the server spacing is also more and processing strategy will also be getting increased. So, that some intelligent process mining techniques are required to process the student performance analysis scheme [2].

This kind of processing are usually handled by means of classical data mining algorithms such as SVM, ANN and so on, these kinds of algorithms produces acceptable yields as well. However, the condition behind those models are crucial while the processing data is more and this kind of algorithms taking much time to process the data compare to our proposed algorithms[3]. In this proposed system the following four methodologies are applied to extract the web-content as well as manipulate the student educational content analysis, they are:

- i). Resource linking Algorithm.
- ii). Latent Semantic Analysis (LSA) Algorithm.
- iii). Content Based Recommendation Algorithm and.
- iv). Data Searching and Ranking.

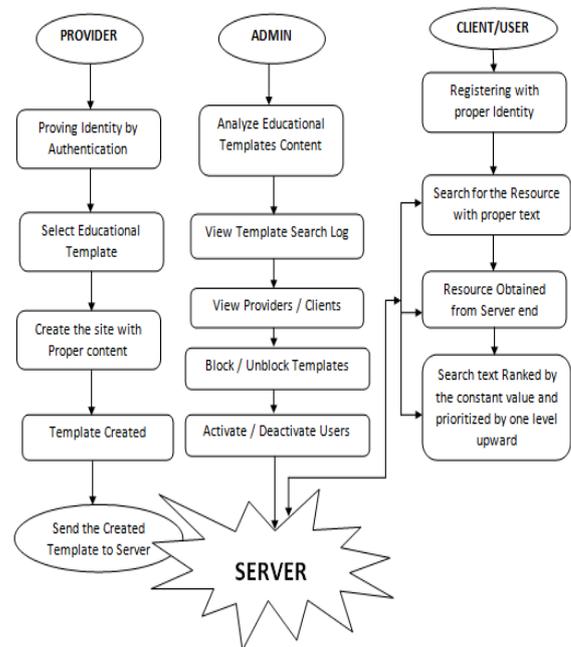


Fig. 1: Proposed System Model.

a) Resource Linking Algorithm

The resource linking is the process of analyzing the web-links, which is used to analyze the data/information presented into the WebPages [4]. This algorithm is used to manipulate the different-

kinds of associations with different servers. The kind of associations will be manipulated either by means of number of data presented into the respective location of the server as well as transactions performed into the server. For all the Resource linking Algorithm is especially used to identify the associations presented into multiple servers in multiple locations.

In the areas of knowledge-discovery and data mining domains, this Resource linking Algorithm plays a vital role to operate with high-level-of-accuracy, as well as which is efficient in the process of data processing, data-transformation, data-analysis and data-virtualization. The major analysis of links is estimated based on the following norms such as: trust, deadline, financial-aspects and machine learning processes [5].

b) Latent Semantic Analysis (LSA) Algorithm

Generally Semantic Analysis is used for the process of Natural Language manipulation tasks, which is associated with the syntactic and semantic feature extraction process of the given data/information.

The process of Latent Semantic Analysis algorithm mainly concentrates on the nature of language, sentence-formations, phrase-accumulations and cross-word checking [6]. This process also involving the unwanted feature-removal process that is also mainly focused as well as used for ethical data/information maintenance schemes in data mining industry. This process helps to analyzers to remove the unwanted sentences, phrases and stopwords, to result with the high-efficient proper sentences in final. For all the entire process of Latent Semantic Analysis algorithm is used to produce the best sentences and grammatical contents over designing the server based applications.

c) Content Based Recommendation Algorithm

The Content Based Recommendation algorithm is derived for Content Personalization over input document/data/information. This process introduces an approach to personalize educational content based on semantic web information. For this, two main mechanisms are developed: a content based template generator that allows providers to generate template profiles representing the user preferences, and a content-based recommendation algorithm estimates the user's interest in unknown content by matching their template to metadata descriptions of the content. These two features are integrated into a personalization system for providing more accuracy and clarified results to analyzers [7].

d) Data Searching and Ranking

The data user or clients can Search for the educational content provided by the resource provider. Before retrieval the ranking system creates the rank for the user searching templates, so that the user can get the more appropriate templates at another search without any confusions or mismatches [8].

2. Problems identified in earlier approach

In the past approaches profile based personalized educational content Search does not support runtime data checking based on content. A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminately. Such 'one-profile-fits-all' strategy certainly has drawbacks given the variety of queries. Evidence reported in is that content based personalization may not even help to improve the search quality for some disjoint queries, though exposing user profile to a server has put the user's privacy at risk.

2.1. Proposed system design

In the proposed approach we implement adaptive educational content authoring principles, which involve:

- i). Authoring the educational content itself.
- ii). Providing domain specific metadata structures (conceptual, semantic), which are used by an adaptive system and.
- iii). Interconnecting the educational content with these metadata.

The contribution of this proposed approach is based on twofold:

- i). A method for hierarchical relationship extraction.

- ii). Experimental analysis on the impact of Computer Aided Metadata creation on educational data authoring.

3. Literature survey

In the year of 2011, the authors "M. Bieliková, M. Barla, and M. Šimko" proposed a paper titled "Lightweight Semantics for the 'Wild Web'" [1, 9], in that they described such as: the present Web has numerous angles. It is never again just a place for content introduction. The Web is increasingly a place where we really invest energy performing different undertakings, a place where we search for fascinating data in light of talks, sentiments of others, and in addition a place where we spend some portion of our entertainment and relaxation time. Also, the Web gives a framework to applications that offer different administrations. In this paper we focus on portrayal and procurement of lightweight semantics for the "wild" Web, which is an absolute necessity on the off chance that we need to move to a "more quick witted" Web and web applications, which adapt to dynamic substance and consider client highlights to convey customized involvement. We as of now have instruments that construe suggestions for a specific client where the setting is known and the substance is depicted utilizing a specific type of semantics (which is an instance of just a couple of islands in the immense sea of the Web). In moving to the "wild Web" we don't have many hints about the substance itself. All the more frequently we have a photo of exercises performed inside specific substance, which can help at any rate and in addition the substance itself. We introduce our proposition of lightweight semantics models together with their social upgrades and talk about a few parts of lightweight semantics procurement on the "wild Web" as expansive and dynamic data space. We talk about cases of ways to deal with towards a change of satisfying our data needs in view of thinking on semantic portrayal of the web content. These cases are late outcomes started from the PeWe (Personalized Web, pewe.fiit.stuba.sk) examine bunch at the Institute of Informatics and Software Engineering at the Slovak University of Technology in Bratislava.

In the year of 2014, the authors "M. Bieliková, M. Šimko, M. Barla, J. Tvarožek, M. Labaj, R. Móra, I. Srba, and J. Ševcech" proposed a paper titled "ALEF: from Application to Platform for Adaptive Collaborative Learning" [2,10], in that they described such as: Web 2.0 has tremendously affected instruction. It encourages access and accessibility of learning content in assortment of new configurations, content creation, learning custom-made to understudies' individual inclinations, and coordinated effort. The scope of Web 2.0 apparatuses and highlights is continually developing, with concentrate on clients and ways that empower clients to mingle, offer and cooperate on (client created) content. In this section we show ALEF—Adaptive Learning Framework that reacts to the difficulties postured on instructive frameworks in Web 2.0 period. Other than its base usefulness—to convey instructive substance—ALEF especially concentrates on influencing the figuring out how to process more proficient by conveying customized learning knowledge by means of customized proposal, and empowering students to work together and effectively partake in learning by means of intuitive instructive segments. Our current and effectively used arrangement fills in as the medium for introducing key ideas that empower acknowledging Web 2.0 standards in training, in particular lightweight models, and three segments of structure foundation vital for steady development and incorporation of understudies specifically into the instructive procedure—explanation system, input framework and gadgets. These influence conceivable to devise to and execute different instruments for proposal and cooperation—we additionally show those strategies for customized suggestion and joint effort with their assessment in ALEF [13], [14].

4. Conclusion

In the proposed approach, the typical educational mining process is considered into the data mining domain and the summary proves the nature and accuracy of the proposed system in detail. This approach produces result with high-accuracy and good-performance with the help of four different algorithms such as: Resource linking Algorithm, Latent Semantic Analysis (LSA) Algorithm, Content Based Recommendation Algorithm and Data Searching and Ranking. All these algorithms are clearly stated in the above descriptions. For all the entire system is clearly stated with proper definitions and the nature of work guarantees the accuracy and performance in results.

References

- [1] M. Bieliková, M. Barla, and M. Šimko, "Lightweight Semantics for the 'Wild Web'," Proc. IADIS Int. Conf. WWW/Internet, IADIS Press, pp. xxv-xxxii, 2011.
- [2] M. Bieliková, M. Šimko, M. Barla, J. Tvarožek, M. Labaj, R. Móro, I. Srba, and J. Ševcech, "ALEF: from Application to Platform for Adaptive Collaborative Learning". In Recommender Systems for Tech. Enhanced Learning, Springer, pp. 195-225, 2014. https://doi.org/10.1007/978-1-4939-0530-0_10.
- [3] P. Brusilovsky, "Adaptive Hypermedia," User Modeling and User-Adapted Interaction, vol. 11, no. 1-2, pp. 87-110, 2001. <https://doi.org/10.1023/A:1011143116306>.
- [4] P. Brusilovsky, "KnowledgeTree: A Distributed Architecture for Adaptive E-learning," Proc. 13th Int. World Wide Web Conf. Alternate Track Papers & Posters, pp. 104-113, 2004. <https://doi.org/10.1145/1013367.1013386>.
- [5] P. Brusilovsky, J. Knapp, and J. Gamper, "Supporting Teachers as Content Authors in Intelligent Educational Systems," Int. J. Knowledge and Learning, vol. 2, no. 3/4, pp. 191-215, 2006. <https://doi.org/10.1504/IJKL.2006.010992>.
- [6] P. Brusilovsky and E. Millán, "User Models for Adaptive Hypermedia and Adaptive Educational Systems," The Adaptive Web, LNCS 4321, Springer, pp. 3-53, 2007. https://doi.org/10.1007/978-3-540-72079-9_1.
- [7] P. Cimiano, Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. Springer US, 2006.
- [8] P. Cimiano, A. Hotho, and S. Staab, "Comparing Conceptual, Divisive and Agglomerative Clustering for Learning Taxonomies from Text," Proc. 16th European Conf. Artificial Intelligence (ECAI 2004), pp. 435-439, 2004.
- [9] A.I. Cristea and A. de Mooij, "Designer Adaptation in Adaptive Hypermedia Authoring," Proc. 2003 Int. Symp. Information Technology (ITCC 2003), IEEE, pp. 444-448, 2003. <https://doi.org/10.1109/ITCC.2003.1197570>.
- [10] P. de Bra, D. Smits, and N. Stash, "The Design of AHA!," Proc. 17th ACM Conf. Hypertext and Hypermedia (HYPERTEXT '06), ACM, pp. 133-134, 2006.
- [11] J. de Knijff, F. Frasincar, and F. Hogenboom, "Domain Taxonomy Learning from Text: The Subsumption Method versus Hierarchical Clustering," Data & Knowledge Engineering, vol. 83, pp. 54-69, 2013. <https://doi.org/10.1016/j.datak.2012.10.002>.
- [12] O.L. de Lacalle and M. Lapata, "Unsupervised Relation Extraction with General Domain Knowledge," Proc. 2013 Conf. Empirical Methods in Natural Language Processing, pp. 415-425, 2013.
- [13] D. Dicheva and C. Dichev, "Authoring educational topic maps: can we make it easier?" 5th IEEE Int. Conf. Advanced Learning Technologies, IEEE, pp. 216-218, 2005. <https://doi.org/10.1109/ICALT.2005.73>.
- [14] D. Dicheva and C. Dichev, "Helping Courseware Authors to Build Ontologies: The Case of TM4L," Frontiers in Artificial Intelligence and Applications, vol. 158, IOS Press, pp. 77-84, 2007.
- [15] H.N. Fotzo and P. Gallinari, "Learning 'Generalization/Specialization' Relations between Concepts – Application for Automatically Building Thematic Document Hierarchies," Proc. 7th Int. Conf. Computer-Assisted Inf. Retr., pp. 143-155, 2004.
- [16] M. Grigoriadou and K. Papanikolaou, "Authoring Personalised Interactive Content," First International Workshop on Semantic Media Adaption and Personalization (SMAP 2006), pp. 80-85, 2006. <https://doi.org/10.1109/SMAP.2006.12>.
- [17] Z.S. Harris, Mathematical Structures of Language, Wiley, 1968.
- [18] J. Harinek and M. Šimko, "Improving Term Extraction by Utilizing User Annotations," Proc. ACM Symp. Document Engineering 2013 (DocEng '13), ACM, pp. 185-188, 2013. <https://doi.org/10.1145/2494266.2494306>.
- [19] S. Harispe, S. Ranwez, S. Janaqi, S. and J. Montmain. "Semantic Measures for the Comparison of Units of Language, Concepts or Instances from Text and Knowledge Base Analysis". arXiv preprint arXiv:1310.1285, 2013.
- [20] M. Hatala, D. Gasevic, J. Jovanovic, M. Siadaty, and C. Torniai, "Ontology Extraction Tools: An Empirical Study with Educators," IEEE Trans. Learning Technologies, vol. 5, no. 3, pp. 275- 289, 2012. <https://doi.org/10.1109/TLT.2012.9>.