Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 151 (2019) 1225–1230

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Web Search and Data Mining (WSDM)
April 29 - May 2, 2019, Leuven, Belgium

An intelligent approach for the design and development of a personalized system of knowledge representation

Amelec Viloria[a]*, Omar Bonerge Pineda Lezama[b]

Universidad de la Costa (CUC), Barranquilla 080003, Colombia [a]
Universidad Tecnológica Centroamericana (UNITEC), Tegucigalpa 11101, Honduras [b]

Abstract

This article proposes a generic presentation system for hypermedia systems of adaptive teaching that is highly independent from the representation of domain knowledge and the application state maintenance. Generality is achieved by providing an application framework for the definition of ontologies that best fit a domain or a specific author. The presentation of the pages to be generated is described in terms of classes and relationships of the ontology. For this purpose, a web page ranking algorithm based on automatic learning is used, specifically, the algorithm for Advanced Cluster Vector Page Ranking (ACVPR). This algorithm provides the user a powerful meta-search tool that presents a ranking order of the web page to quickly meet custom needs, especially when the search is erroneous or incomplete.

Keywords: adaptive hypermedia; ontologies; knowledge representation; user modeling; interface design tools; teaching on the web; algorithm for advanced cluster vector page ranking.

1. Introduction

The fast development of web-based training systems is allowing the increase of the student´s autonomy, while educational applications are reaching a more diverse and heterogeneous audience. This scenario raises a renewed interest in the development of hypermedia systems that permit the automatic adaptation to different types of users, platforms and situations, which considers the evolution of each user over time (Alam, M. and Sadaf, K., 2015)[1]. A

* Corresponding author. Tel.: +573046238313
E-mail address: aviloria7@cuc.edu.co

10.1016/j.procs.2019.04.176

priority concern, explicitly or implicitly present in this research framed in the field of adaptive teaching is to find an adequate representation of the pedagogical knowledge (Chen, C. and Zhang, C., 2014)[2]. Each tutoring system building tool establishes its own way of structuring the domain, so that the designer describes the subject in terms established by the tool, which takes charge of the selection, presentation, dynamic sequencing of the teaching material, as well as the interaction with the user. In the existing tools, the hypermedia documents are generated according to pre-set page designs that the authors of the course can not configure (Zhu, H., et al.,2017)[3], for example, or they are required to provide a description of all the pages to be generated as in (Ferretti, S. et al., 2016)[4].

The purpose of this work is to create a generic hypermedia presentation system that provides the course authors a simple specification paradigm to define non-trivial elements of page design and an adaptive presentation, independent from the contents (Malhotra, D. and Rishi, O., 2018)[5], (Malthankar, S. and Kolte, S., 2016)[6], (Zhang, G. et al., 2012)[7].

The automatic learning model based on logistic regression can be developed using the Statistics R tool to predetermine the suitability of a web page to meet the user's custom search needs (Xiang, B.et al., 2010)[8]. The learning model will determine the most appropriate page ranking order for a specific web user in relation to several parameters, such as the page load speed, response time, page navigation security, and custom relevance. The scientific evaluation related to the calculation of the matrix of confusion, specificity, sensitivity, etc. easily checks the model fit to the proposed system.

This research pretends to merge automatic learning with big data analysis for the implementation of the search algorithm and the proposed custom page ranking, that is, the ACVPR algorithm in the form of an intelligent metasearch system tool to help the end user list the web with the most relevant knowledge and the custom knowledge representation.

With respect to the use of explicit presentation models for the dynamic generation of web documents, (Malhotra, D. and Rishi, O., 2016)[9] proposes a system based on adaptive models of page design to allow the HTML dynamic display of the information stored in databases. The most significant restriction of this system in relation to the objectives proposed in this work lies in its limited expressive capacity for the representation of knowledge, which in the mentioned system is limited to a relational data model.

2. System Design

The system design for the proposed metasearch tool has three sub-phases. The detailed description of each phase and a simplified block diagram are shown in Fig 1.

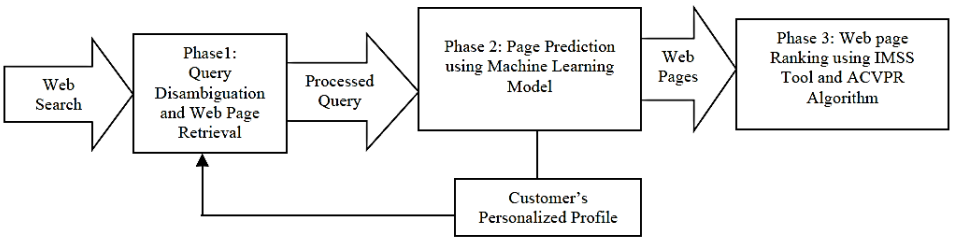

Fig. 1. System design of proposed IMSS-P tool and ACVPR (Malhotra, D. and Rishi, O., 2017)[10]

The recommendation engine module can be used to build a user's profile using web technology semantics. The expanded custom query is also passed to the number of search engines in the background of the metasearch tool. To predict a user's preference for a specific web page, an automatic learning model based on logistic regression was developed. The response variable to be predicted is the feedback with respect to the relevance of the ranked web link in the output of the metasearch tool. The data must be in .csv format as required by the Statistics R tool. The .csv format file will consist of the data about the following five variables (Malhotra, D et al., 2017)[11]:

- Comments represent the user relevance response for the previous web link in his browsing history and can take either of two values, Yes or No.
- The load represents the load experience of the user's web page and can take one of the two values, Good or Bad.
- The response represents the user's response time experience and can take one of two values, Good or Bad.
- Security represents the security protocol function provided by the candidate website and can take either of two values, Yes or No.
- Custom represents the use of the function, that is, the custom expansion of the function.

The summary of the generated model is shown in Fig 2.

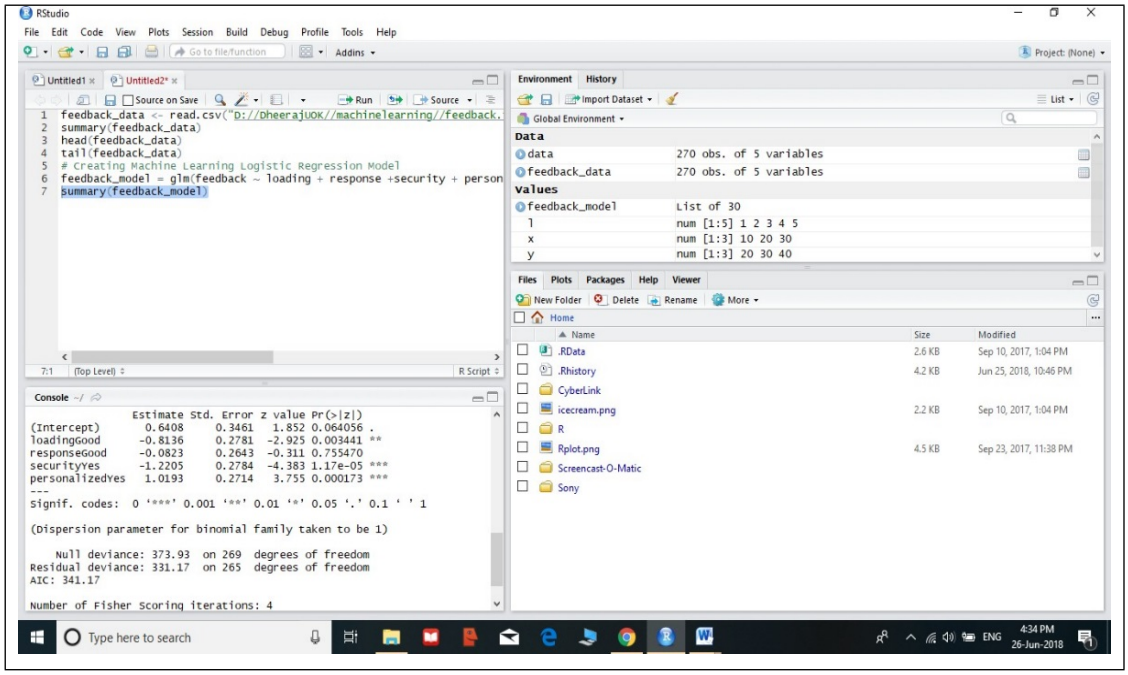

Fig. 2. Generalized linear model generation for IMSS-P

3. Discussion and Results

The proposed research aims to design the ACVPR algorithm and its implementation in the form of IMSS-P tool to determine the effectiveness and efficiency of the proposed approach. The ACVPR algorithm and IMSS-P tool are discussed below.

The IMSS-P tool is deployed on the Google cloud platform and is coded using Python programming and Statistics R tool to verify the efficiency and effectiveness of the proposed ACVPR algorithm. The IMSS-P is an advanced version of IMSS-AE tool proposed by Malhotra, D. and Rishi, O (2018)[5], in the sense that IMSS-P employs machine learning logistic regression to predict web user's personalized preferences and hence improves the search precision. The IMSS-P is compatible with mobile phone interface and supports biometric sign in, i.e. fingerprint scanner. Moreover, IMSS-P is a generic metasearch tool and is not limited to e-Commerce search queries like IMSS-AE. The IMSS-P also provides advanced information on its interface such as the search engine information from which a specific URL is fetched, and search accuracy statistics (Zhou, D. et al., 2018)[12]. The IMSS-P supports two

search modes (i) intelligent metasearch and (ii) advanced metasearch. The advanced mode is for technical users and lets them select various tabs as shown on the interface of the tool (Liu, Y. et al., 2017)[13]. These tabs include options to choose background search engines to be used by meta search tools, i.e. a user is free to select one or three search engines out of Google, Bing, and Kartoo. Furthermore, the user has options to select search predictors to determine the ranking order in advanced mode, i.e. website response time, webpage loading speed and browsing security. The user can choose any of these parameters to figure out the custom page ranking order in the output of the search tool (Yang, Y. et al, 2012)[14], (Torres-Samuel M. et al; 2018)[15].

3.1 Knowledge representation templates

The templates are defined by an HTML extension based on JavaServer Pages (JSP) (Ahmad, M. et al., 2017)[16] (Kamatkar S. et al; 2018)[17], which allows interleaving control statements (between <% and %>) and Java expressions (between <% = and %>) in the HTML code. In these templates, the designer can use all the elements of presentation of the HTML language (lists, tables, frames, links, forms, etc.) like entering, using very simple Java expressions, and the domain elements to be presented. For example, a simple template for the Algorithm class could be the following:

```
<h2> <%= title %> </h2>
<h3> Previous concepts </h3>
<%= prerequisites %>
<h3> Procedure </h3>
<%= procedure %>
<h3> Examples </h3>
<%= examples %>
<h3> Proof of Correction </h3>
<%= correction %>
```

In these templates, the author of the presentation just must reference attributes and relationships of the presented class (in the example, in bold) (Bouadjenek, M. et al., 2016)[18]. The presentation system internally deals with aspects such as automatic list management (multi-valued relationships, i.e. algorithm examples), or the recursive application of templates to objects referenced according to their class (i.e. the algorithm's correction theorems). The HTML elements that surround the presentation of the algorithm (frame structure with a contextual index on the left and Previous/Next buttons at the bottom) correspond to the template for the DomainObject root class. For example, the resulting page for the Dijkstra algorithm with this presentation template can be seen in Fig 3.

While the construction of templates is within reach from any familiar web designer with HTML and JSP technologies, the definition of ontologies is a delicate task that requires participation of an advanced designer, trained in the use of the tool (Aoki, Y. et al., 2015)[19], (Sanchez L. et al; 2018)[20]. Once defined the ontology and the associated presentation models, the construction of the domain model is affordable for the average author (Adamopoulos, P., 2014)[21], (Viloria A. et al; 2018)[22]. The introduction of modifications on the templates and rules of presentation can be an easy step for this class of author towards a more advanced level of use.

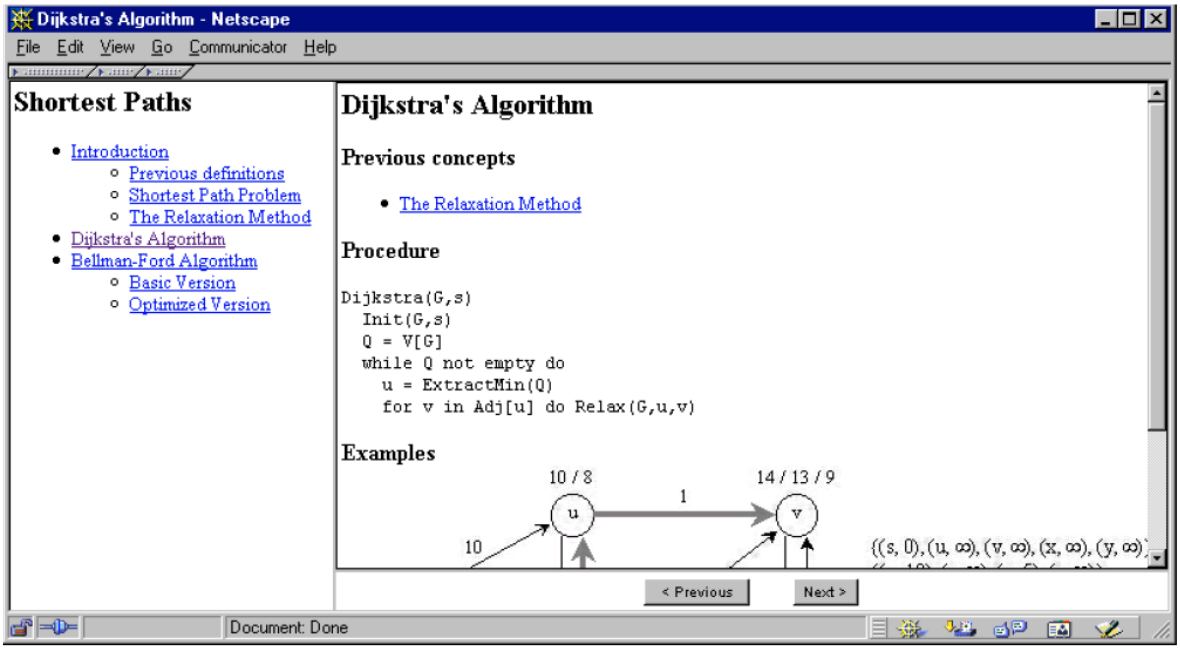

Fig. 3. Web page generated for a topic of Algorithm type

4. Conclusions

This research paper proposes the ACVPR algorithm and the architecture of a metasearch system. This search tool is backed by intelligent and advanced technologies such as the semantic web and big data analysis based on Hadoop2 to predict meta keywords to expand the user's search query and to manage a massive number of web links returned by engines of background search.

The proposed knowledge representation system allows to reproduce the domain models used in a wide range of hypermedia systems. The dynamic generation of the independent presentation update of the application's status makes the tool compatible with other support applications for adaptive courses. The approach allows the specification of the presentation separate from the construction of contents, favoring the reuse and consistency of the presentation, thus reducing the cost of development.

References

[1] Alam, M. and Sadaf, K., 2015. Labeling of Web Search Result Clusters using Heuristic Search and Frequent Itemset. Procedia Computer Science, Elsevier,216-222.

[2] Chen, C. P., & Zhang, C. Y., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, Elsevier, 275, 314-347

[3] Zhu, H., Ou, C. X., Van den Heuvel, W. J. A. M., & Liu, H.,2017. Privacy calculus and its utility for personalization services in e- commerce: An analysis of consumer decision-making. Information & Management, Elsevier, 54(4), 427-437.

[4] Ferretti, S., Mirri, S., Prandi, C., & Salomoni, P., 2016. Automatic web content personalization through reinforcement learning. Journal of Systems and Software, Elsevier, 121, 157-169.

[5] Malhotra, D., & Rishi, O. P.,, 2018. An intelligent approach to design of E-Commerce metasearch and ranking system using next- generation big data analytics. Journal of King Saud University-Computer and Information Sciences, Elsevier

[6] Malthankar, S. V., & Kolte, S., 2016. Client Side Privacy Protection Using Personalized Web Search. Procedia Computer Science, Elsevier, 79, 1029-1035.

[7] Zhang, G., Li, C. and Xing, C., 2012. A Semantic++ Social Search Engine Framework in the Cloud. In Semantics, Knowledge and Grids (SKG), 2012 Eighth International Conference, IEEE, 270-278

[8] Xiang, B., Jiang, D., Pei, J., Sun, X., Chen, E., & Li, H., 2010. Context-aware ranking in web search. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, ACM, 451-458.

[9] Malhotra, D. and Rishi, O.P., 2016. IMSS-E: An Intelligent Approach to Design of Adaptive Meta Search System for E-Commerce Website Ranking. Proceedings of the International Conference on Advances in Information Communication Technology & Computing, ACM, doi>10.1145/2979779.2979782.

[10] Malhotra, D. and Rishi, O.P., 2017. IMSS: A Novel Approach to Design of Adaptive Search System Using Second Generation Big data Analytics. Proceedings of International Conference on Communication and Networks, Springer, 189-196.

[11] Malhotra, D., Malhotra, M. and Rishi, O.P., 2017.An Innovative Approach of Web Page Ranking Using Hadoop- and Map Reduce- Based Cloud Framework. Proceedings of Advances in Intelligent Systems and Computing, Vol.654, CSI, Springer, 421-427.

[12] Zhou, D., Zhao, W., Wu, X., Lawless, S., & Liu, J., 2018. An iterative method for personalized results adaptation in cross-language search. Information Sciences, Elsevier, 430, 200-215.

[13] Liu, Y., Bi, J.W. and Fan, Z.P., 2017. Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. Information Fusion, 36, 149-161.

[14] Yang, Y. F., Hwang, S. L., & Schenkman, B.,2012. An improved Web search engine for visually impaired users. Universal Access in the Information Society, 11(2), 113-124.

[15] Torres-Samuel M., Vásquez C.L., Viloria A., Varela N., Hernández-Fernandez L., Portillo-Medina R. Analysis of Patterns in the University World Rankings Webometrics, Shanghai, QS and SIR-SCimago: Case Latin America. In: Tan Y., Shi Y., Tang Q. (eds) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol 10943. Springer, Cham. 2018

[16] Ahmad, M. W., Doja, M. N., & Ahmad, T., 2017. Enumerative feature subset based ranking system for learning to rank in presence of implicit user feedback. Journal of King Saud University-Computer and Information Sciences. Elsevier

[17] Kamatkar S.J., Tayade A., Viloria A., Hernández-Chacín A. (2018) Application of Classification Technique of Data Mining for Employee Management System. In: Tan Y., Shi Y., Tang Q. (eds) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol 10943. Springer, Cham

[18] Bouadjenek, M. R., Hacid, H., Bouzeghoub, M., & Vakali, A., 2016. Persador: personalized social document representation for improving web search. Information Sciences, Elsevier, 369, 614-633.

[19] Aoki, Y., Koshijima, R. and Toyama, M., 2015. Automatic Determination of Hyperlink Destination in Web Index. In Proceedings of the 19th International Database Engineering & Applications Symposium, ACM, 206-207.

[20] Sanchez L., Vásquez C., Viloria A., Cmeza-estrada (2018) Conglomerates of Latin American Countries and Public Policies for the Sustainable Development of the Electric Power Generation Sector. In: Tan Y., Shi Y., Tang Q. (eds) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol 10943. Springer, Cham

[21] Adamopoulos, P., 2014. On discovering non-obvious recommendations: Using unexpectedness and neighborhood selection methods in collaborative filtering systems. Proceedings of the 7th ACM international conference on Web search and data mining, ACM, 655- 660.

[22] Viloria A., Lis-Gutiérrez JP., Gaitán-Angulo M., Godoy A.R.M., Moreno G.C., Kamatkar S.J. (2018) Methodology for the Design of a Student Pattern Recognition Tool to Facilitate the Teaching - Learning Process Through Knowledge Data Discovery (Big Data). In: Tan Y., Shi Y., Tang Q. (eds) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol 10943. Springer, Cham