Journal of
Creative Practices in Language Learning and Teaching (CPLT) Special Issue: *Generating New Knowledge through Best Practices in Computing and Mathematical Sciences*
Volume 7, Number 1, 2019

Towards Interface Design for Virtual Database

Zanariah Idrus, Noor Hasnita Abdul Talib, Siti Zaleha Zainal Abidin, Nasiroh Omar and Zainura Idrus

Received: 22 May 2018. Accepted: 15 Feb 2019/Published online: 28 Feb 2019

ABSTRACT
Today, big data has become as one of the important contribution in database management. It led to innovative ways of storing and organizing data which include structured and unstructured data. The unstructured data such as in news, reports, chats and surveys are basically loaded with heavy text data and numerous format. Thus, these data become challenging to be used for diverse purpose and are not appropriate to be stored in database. However, virtual database method has the capability to organize the unstructured data, and reconstruct into firm and concrete data. This approach carry out two major processes in databases which are mining and managing the data. However, the main problem is the insufficient support between people using databases and the heap of data collection. This is due to unawareness of clustered data organization as information is stored implicitly. Thus, this paper presents the conception of clustered data using the interface design model. Alignment of features and connections between the interface and knowledge composition allow users to access knowledge proficiently.

Keywords: Interface model, Data extraction, Data clustering

✉ Z.Idrus
Faculty of Computer and Mathematical Sciences,
Universiti Technologi MARA,Malaysia.
E-mail: zanaidrus@uitm.edu.my

N.H.Abdul Talib
E-mail: nhasnita@uitm.edu.my

S.Z.Z.Abidin N.Omar
E-mail: sitizaleha533@uitm.edu.my

N.Omar
E-mail: nasiroh@uitm.edu.my

Z.Idrus
E-mail: zainura@uitm.edu.my

1 INTRODUCTION

Nowadays, information has been considered as one of a major asset of an organization. There are many types of information sources and database has been one of it. Sources can further being categorized as structured or unstructured. Two categories of unstructured information are textual and non-textual data [1]. Textual data typically come from sources such as text documents, web pages, newspaper article, personal blogs, discussion forums and other areas while non-textual come from images, sounds and video files. Information extraction (IE) is the automatic extraction of structured information from unstructured sources [2].

The structured information can be in the form of the entities, the relationships between entities, or the attributes. The differences of the way data sources are organized, the vocabulary used and the data-access or query mechanisms used had made the data combination from multiple source are difficult to be performed. The transformation from the unstructured or semi-structured information into structured knowledge has also becoming one of the main challenges of today's knowledge society [3]. Hence, Virtual database (VDB) technology has enabled the external data sources to become an extension of an organization's relational database system [4]. Thus it is crucial to have an interface model that has the capability to provide the end-user an environment to access such structured virtual database. Systems commonly have many types of presentation views and human-computer interface management which can be from a computer science viewpoint that focuses on the developing quality human-computer interfaces process [5]. Interfacing of human with computer (database engine) is a way of extracting the required information along with its meanings [6].

Currently, there is lack of focus in interface design model for virtual database environment. This becomes vital as users are dealing with unstructured and structured data in managing information. Hence, in our study, the primary objective is to develop an interface model that has the capability to provide the end-user an environment to access such structured virtual database. Systems were basically had various types of presentation views and various types of interactions [7]. Even though data display has advanced, the modern version still relying on visual data representation [6]. Human-computer interface management focuses on the process of developing quality human-computer interfaces, including their representation, design, implementation, execution, evaluation, and maintenance [5].

2 RELATED WORKS

Researchers are searching for a variety of information management strategies to establish order in the text wilderness due to information missed and opportunities lost. Two main strategies are information retrieval (IR) and information filtering [8]. Information extraction (IE) is about finding particular pieces of data in natural-language documents, thus structured information is extracts from unstructured text [9]. IE works as a front end for high precision information retrieval or text routing [10]. Two techniques used in retrieving information from database are basically data extraction and clustering techniques. Data extraction is a process which starts from information finding in the documents for example from particular paragraphs, table entries, or

graphical elements from figures which then being transformed into a canonical form [11]. In contrast, data clustering is use to partition a set of data objects into clusters based on their characteristics similarity [12]. This tool is widely used with applications such as astronomy, bioinformatics, machine learning, and pattern recognition [13].
A virtual database (or VDB) is used to hold components which incorporate data from various data sources [14]. Users can access VDB in an integrated manner through a single, uniform application program interface (API). Four main components in the VDB structures are basically the Mapper, Publisher, Executor and Wrapper. VDB technology enables applications rapid deployment. The characteristics can be either one of the following: enormous data sources, independent data sources or the centralized control data sources can't have a combination of structured and unstructured data [15].

An interface comprises of visualization and interaction. Visualization is the output of graphical information while interaction is the input of information [16]. According to the information suggested by authors in [7], the use of three type models are required. Firstly, task-based models are a tasks-oriented human computer system. Second, object-oriented models are models of a system using objects. Finally, data-centered models are models of a system consist of data. Task models are about representing the domain based on task due to the importance of HCI in human-computer interaction [17] [18]. Next is the object-oriented paradigm in which the domain is represented by the existing objects including the objects relationships and messages. In data-centered models approach, information of the entities, data flows and processes are kept in a data dictionary. Mobi-D1 (Model Based Interface Designer) supports user centered design by using model-based interface development [19]. This model uses a series of declarative models where consisting of interrelated components of user-task, dialog and presentation. This is different with model-based systems that using either one or two models in isolation and have no explicit notion as to how the various model elements are organized into an interface design.

This study discusses the modelling concept of interface design for virtual database. Virtual database has successfully performs two main operations on database which are data extraction and data clustering management in database management system. The basic principle of interface design should be simple, fulfill task requirement and well-communicated. Thus, this proposed model shows the front-end of virtual database in the form of user interface design.

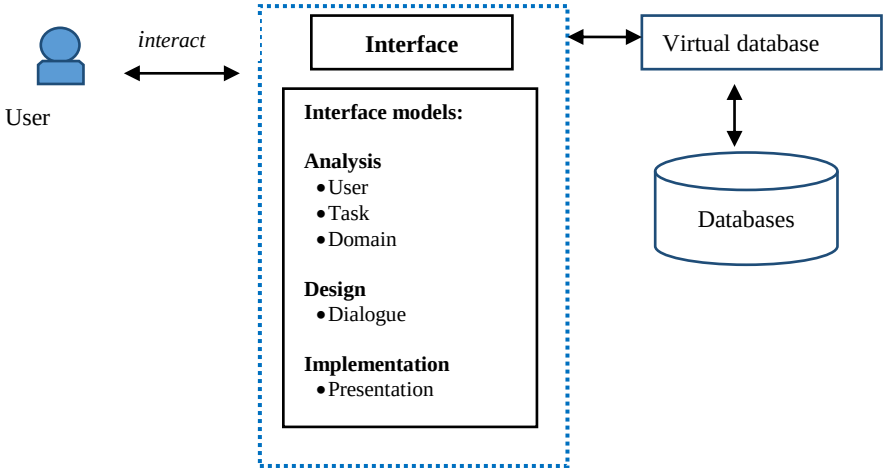

Fig. 1: Overview of interface model

As in Figure 1 above, the overview of interface model involves user, interface models, virtual database, and database. Users interact with the database system using the interface mechanism. The interface model is divided into three phase which are analysis, design and implementation. The analysis phase consists of user model, task model, and domain model. The design phase include dialog model and presentation model is in the implementation phase. Data are stored in the virtual database which has been structured and integrated with disparate databases.

User model describes the characteristics of the diversity type of users in order to create the personalized of user interface. It represents the roles of each users and can be categorized as application independent and application dependent. Users are given a set of tasks to accomplish. Task model described the abstract of user interface and concentrate on what user need to do. It is used as guidance in generating different modalities and identifies the set of activities and actions accordingly. Thus users have to perform it in achieving a concrete goal or purpose. In other words it identifies users' scenarios. Next is to modeled it and represent to the domain model. Domain model provide the special features for creating use interface. It represents the attributes, methods and relationships in a system. It determines the objects that users can view, access and manipulate. Both task and domain model has a strong bonding with each other in order to ensure all actions and data elements are embedded in the user interface. Dialog model established information and interaction based on the characteristics of each users. It is a sequence of interactions using user interface elements of tasks and actions such as transitions of navigation from a screen to another screen. User, task and dialog are the most significant models as they cover all the relevant aspects of an interface design. Presentation model identify the presence of objects or information in different dialog situation. It include the hierarchical structure of screen to be displayed and objects in the interaction. It also identify devices users may access a service and align conceptual elements in dialog model into the platform elements.

Virtual database has the advantages of merging multiple databases, extracting, clustering data and creating universal and unified information. User interface is needed to be integrated to the virtual databases to be able to expose to the information. It may be in the form of metadata, statistical, graphical data and visual analysis. In view of that, this study map interface model with the virtual database.

$$VD\_UI=A\ (D+T+U) +D\ (Dg) + I\ (P) \qquad (1)$$

As shown in equation Eq.(1), user interface in virtual database (VD_UI) should include the analysis phase (A) which consists of domain (D), task (T), and user (U). The design phase (D) comprise of dialog elements in the interface. Finally, in the implementation phase (I) presentation have the responsibility to give high values to the information displayed.

4 CLUSTER AND INTERFACE MODEL

There are rich labels for various online document publications. The hierarchical clustering capable to group the metadata into clusters virtually of the similar characteristics in the form of hierarchy. The clusters include the characteristics retrieved from title, authors, affiliation, addresses, email, phone number, date, keywords, abstract, uniform resource locator (URL) addresses, and publication. The interface model corresponds with the user interface interaction and display results in more meaningful manners and with a better user acceptance. This new approach provides options for developers in this field which can be used as an alternative method for interface design.

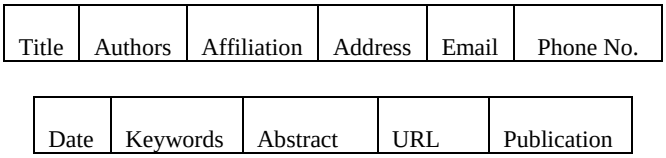

Fig. 2: Characteristics for online document publication

5 CONCLUSION

Virtual database is able to store, cluster unstructured data and organize structured data in a universal way. However, there is lacking of support between users and the clustered data in virtual database which may lead to information misrepresent. Data are virtually stored, users are unaware of the cluster structure and consistently fail to understand the data and their meaning. Thus, this paper presents an interface model for virtual database to visualize the cluster and represent in a more meaningful information. The model becomes the intermediate component between users and virtual database.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknologi MARA and Ministry of Higher Education Malaysia for the financial support under the national grant 600- RMI/RAGS 5/3 (20/2012).

REFERENCES

[1] Yafooz, W. M. S., Abidin, S. Z. Z., Omar, N., & Idrus, Z.: Managing unstructured data in relational databases. Proceedings - 2013 IEEE Conference on Systems, Process and Control, ICSPC 2013, (December), 198–203. http://doi.org/10.1109/SPC.2013.6735131 (2013)

[2] Sarawagi, S.: Information Extraction. Communications of the ACM, 1(3), 261–377. http://doi.org/10.1561/1500000003 (2007)

[3] Singh, S. P., Darbari, H., Kumar, A., Mehta, S., Jain, N., Simran, P., & #2, K.: Bilingual Data Extraction and Auto Summarisation, 4350–4354. (2016)

[4] Rajaraman, A., & Norvig, P.: Virtual database technology: transforming the Internet into a database. IEEE Internet Computing, 2(4), 55–58. http://doi.org/10.1109/4236.707691 (1998)

[5] Hartson, H. R., & Hix, D.: Human-computer Interface Development: Concepts and Systems for Its Management. {ACM} Comput. Surv., 21(1), 5–92. http://doi.org/10.1145/62029.62031(1989)

[6] Patel, N., & Hughes, D.: Revolutionizing human-computer interfaces: the auditory perspective. Interactions. Retrieved from http://dl.acm.org/citation.cfm?id=2065336 (2012)

[7] Benyon, D., & Keynes, M. Domain Models for User Interface Design, 44(0), 1–16. (1993)

[8] Cowie, J., & Lehnert, W.: Information Extraction. Communications of the ACM, 39(1), 80–91. http://doi.org/10.1145/234173.234209 (1996)

[9] Mooney, R. J., & Bunescu, R.: Mining knowledge from text using infomation extraction. ACM SIGKDD Explorations Newsletter, 7(1), 3–10. http://doi.org/10.1145/1089815.1089817 (2005)

[10] Soderland, S.: Learning Information Extraction Rules for Semi-Structured and Free Text. Machine Learning, 34(1), 233–272. http://doi.org/10.1023/A:1007562322031 (1999)

[11] Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K.: The cost structure of sensemaking. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93, 269–276. http://doi.org/10.1145/169059.169209 (1993)

[12] Tkaczyk, D.: New Methods for Metadata Extraction from Scientific Literature. Dissertation, University of Warsaw (2015)

[13] Sheng, W., Chen, S., Member, S., & Sheng, M.: Adaptive Multisubpopulation Competition and Multiniche Crowding-Based Memetic Algorithm for Automatic Data Clustering, 20(6), 838–858 (2016)

[14] Sathya, S., & Victor Jose, M.: Application of Hadoop MapReduce technique to Virtual Database system design. 2011 International Conference on Emerging Trends in Electrical

and Computer Technology, ICETECT 2011, 892–896. http://doi.org/10.1109/ICETECT.2011.5760245 (2011)

[15] Gupta, A., Harinarayan, V., & Rajaraman, A.: Virtual Database Technology, 26(4), 57–61 (1997)

[16] Mitchell, K., Kennedy, J., & Barclay, P.: A framework for user-interfaces to databases. Proceedings of the Workshop on Advanced Visual Interfaces, 90. http://doi.org/10.1145/948449.948462 (1996)

[17] F. Paterno and V. V. Alfieri: Task Models in Interactive Software Systems. http://doi.org/10.1142/9789812389718 (2001)

[18] V. Tran, J. Vanderdonckt, M. Kolp, and S. Faulkner: Generating User Interface from Task, User and Domain Models. Second International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services (2009)

[19] Puerta, A. R.: Introduction to Model-Based User Interface. IEEE Software. https://pdfs.semanticscholar.org/9a9b/86488ccb64276fdbe47d5322ee27144ac835.pdf (1997)