

Zuzana Juchova Juroskova - Boris Bucko*

USER ORIENTED APPROACH IN INFORMATION COMMUNICATION TECHNOLOGY RESEARCH

User-oriented approach is changing the way of information systems development and affects the processes, methodologies and paradigms proposed by software architects and developers. In line with that trend we carried out two separate studies where the users were involved in the whole process of the development. The first study used experiment, data mining and text mining and was focused on identifying opportunities for obtaining values of bibliographic elements to digitize objects in order to maximize the quality and quantity of delivering meta-data value elements of physical objects Slovak library database from a user's perspective. The second study used experiment and was focused on the design of the architectural framework for information system solutions for knowledge management in the organization by using the principles of model-driven architecture (MDA) and ontology modeling language within the CIM (Computer Independent Model) and PIM (Platform Independent Model) levels.

Keywords: Database, text mining, datamining, ontology, Computer Independent Model, Platform Independent Model.

1. Introduction

Information society has changed significantly recently. While a few years ago it was difficult to gain access to the information, today we are overwhelmed with information and we are trying to find the best ways of how to organize, search, process, store, represent and use Big Data. We develop new technologies and procedures for data mining, we improve predictive analytics, we create new jobs (external and internal algorithm analytics) and we also set up new types of organizations (i.e. data mediators). The only thing that has not changed significantly is the way how users search information. We can assume that users search information using keyword(s). Interpretation of search results contents and type of its representation influence the popularity and reusability of the specific tool or service.

Database systems and information systems can be considered as the one of the most frequently used information environments. The development of database and information systems has a long history. Software architects and developers tend to use various approaches, methodologies and paradigms in order to implement innovative solutions. The creation of web sites (containing databases and information systems) based solely on the opinions and proposals of the application developers proved to be insufficient. Jacob Nielsen [1] states that estimates of user behavior are wrong up to 75%, which

results into significant changes towards traditional approaches and puts emphasis on user requirements. The quality of the proposals and selected procedures is thus subject of user testing. Nowadays user testing has become usual part of development of the databases and information systems.

For this reason, it appears to be necessary to develop flexible and easy to maintain databases and information systems that can meet the requirements of users in the shortest possible time. In line with this trend - focusing on user requirements and meeting their needs we have carried out two separate studies.

2. User oriented approach

These studies used various methods to acquire input data:

- Analysis,
- Synthesis,
- Questionnaire method,
- Group discussions.

and also methods for data processing:

- Experiment,
- Data mining,
- Text mining.

* Zuzana Juchova Juroskova, Boris Bucko
University Science Park, University of Zilina, Slovakia
E-mail: zuzana.juchova@uwp.uniza.sk

2.1 Study 1

The key source of information for final theses (bachelor, master or PhD) and also for degree theses (rigorous and habilitation) are still libraries which offer databases of printed and electronic objects. These objects can be obtained through online catalogs OPAC (online public access catalog). These databases contain a large number of elements of different metadata schemas describing physical objects. The number of these objects is increasing in the process of mass digitization of database objects which include elements of digitalized objects description. In 2012 the Slovak National Library began mass digitization of object. Within this digitization each digital object contained at least two elements:

- Persistent identifier,
- Metadata Dublin Core.

Project "Digital library" calculates with the usage of qualified Dublin Core format and also with complete MARC21 record which is transformed for digital repository using MODS with all descriptive metadata [2]. It should be noted that metadata elements forming part of digital libraries are primarily intended for the user but are also used to describe web pages, documents, etc. to make the search much easier.

There was used a traditional approach to transform repository in the Slovak National Library – proposal and implementation was done mostly by internal employees of the ICT (information communication technologies) department. During the 12th Annual Conference for Librarians called "The function of libraries under the changed economic conditions and the specific user requirements during the economic crisis" [3] there was discussed disappointment and user unpopularity over the current online library catalog. According to this opinion we tried to investigate the cause of the mentioned problem. The research was conducted by one author as a part of her dissertation thesis [4] *the aim of which was to identify possibilities for obtaining bibliographic elements values during the objects digitization with an aim to maximize the quality and the quantity of delivered metadata elements values of physical objects within Slovak database library. And this should be done from a user perspective.*

Given that the use of online catalog was standardized and established in accordance with applicable standards and recommendations we focused on a content analysis of search results and used the user approach to solve this problem. That is why we focused on acquisition of the most desirable bibliographic elements. We also focused on identifying accessible bibliographic elements provided within specific object records in Slovak library which are actually used by the user during the object decision-making process. Afterwards we determined multiplicity of the users most desirable filled metadata elements of the objects from the Slovak library online database. From the original size of the statistical series

4 203 474 records we were able to generate 119 858 records of text documents from ORACLE database system. These text documents also had filled in the most desirable element - element containing keywords.

The results of the initial analysis showed that the real number of relevant values of the most desired metadata elements provided by a database is only 1.36. According to the total number of records in the databases of the Slovak library we have judged this multiplicity as unsatisfactory, because the user is not able to make decision based on available metadata elements. It is not clear which object is subjectively relevant for the user, which is the major issue in the process of meeting information needs of the user. The size of text documents and the number of pages were the main criteria to identify the most suitable method to supplement existing Slovak library database on a generated metadata element values. These values were necessary to create a new digital library database with text mining method and as criteria of deliberate choice of the text documents were used:

- Condition 1 - filling field 650 - Topical term
- Condition 2 - the necessity of the existence and availability of digital copies of these objects. the need to obtain.

Afterward these generated objects were:

1. Converted to plain text - to get rid of formatting and to prevent language encoding errors within Slovak language.
2. Each object was processed separately in the Rapid Miner software. By assigning the weights to each term we were able to get list of terms represented by chosen vector. Using TF-IDF option we were able to create vector based on the frequency and the reversed frequency of the occurrence of terms (words) in the document [5]. The removal of stop words or edit words to the basic form were not used because these options are not available for texts written in the Slovak language.
3. After the generation of terms, we ignored the conjunctions, prepositions, etc.

The generated keywords were compared with actual values from metadata database elements from Slovak library. We identified a percentage of their conformity. The final match was 77.14% which means that with a probability of 77.14% the generated keyword also has a tag value 650 – Topical term in Slovakia library database.

2.2 Study 2

The complexity of information systems (IS) is constantly increasing. This trend is apparent not only from the technological possibilities but also from the performance requirements that are placed on information system functionality. Functional requirements are based on the

new requirements of the users. Because the user is the one who is the information system developed for. However, it is only in those cases where the requirements of the users are correctly processed by business or system analysts and then translated into the functional specifications of IS. If user requirements are not processed properly, it may result into information system which does not meet the needs of organizations and further system modification and maintenance is difficult and costly. Nowadays it is not possible to focus only on functionality of information system without any deeper knowledge of organization's structure and needs. Linking business perspective and the perspective of a specific information system is therefore still an interesting field of research.

Applying the principles of the ISO / IEC / IEEE 42010: 2011 Standard, the developers are trying to get rid of problems that come with object oriented IS development by establishing connection between business perspective and software perspective in the system architecture [6]. These views form two partial architectures - the business architecture and the software architecture. Each of the partial architectures has different language to describe architecture. Languages for perspective/view description are today the mostly represented as graphical models and model-driven development becomes a fundamental principle or a paradigm of IS development.

Model Driven Architecture (MDA) is based on creation of models and transformations between these models. MDA specifies four levels of abstraction:

1. Computation Independent Model - CIM
2. Platform Independent Model - PIM
3. Platform Specific Model - PSM
4. Implementation Model - IM

Levels of abstraction mentioned above are fundamental paradigm of MDA. The first three levels are graphical models and the last level one is made of source code.

Models of different levels can be gradually transformed to the source code and vice versa – the changes in source code can be applied back in the models using inverse MDA. This simplifies not only the development but also the maintenance of such created information systems. Transformational links between models and source code enable to keep the models, source code, files and documentation in a consistent state. Transformations between levels are an important aspect of MDA. There is a continuous effort to semi-automate these transformations, or if it is possible to completely automate this process. With regards to the type of the MDA it is necessary to create transformation rules between source and target model. In the terms of the application of Object-Oriented Systems Analysis and Design – OOSAD during the IS proposal, the greatest emphasis is put on the PIM - PSM transformation in both directions. Nowadays, there are various CASE tools and the MDA tools that enable transformations. Top level (CIM)

to a lower level (PIM) transformations are still not described in the OMG documents. Only partial results of solving the problem are available [7 - 10].

Within our proposed solutions, we decided to use ontology. Ontology as a tool for developing IS is currently used in different contexts and stages which can be seen in various scientific articles [11 - 14].

Most systems do not fail for technical reasons but because they do not address the real needs of the customers. The lack of the focus in the analytical stages represented by the CIM level (within MDA) only postpones real IS development problems. These problems can later occur during the implementation and during the maintenance of information systems. Organizations need to use information technology to support specific processes. Create support for these processes is not an easy task which cannot be done by configuring ready-made software package. That is why it is usually necessary to create tailor made system which can meet user requirements. Specific processes need to be logically linked to data, information and knowledge through the implementation of knowledge management.

According to the information mentioned above we set the objective to: *Design an architectural framework for information system solutions for knowledge management in the organization by using the principles of model-driven architecture (MDA) and ontology as a modeling language in the levels of CIM (Computer Independent Model) and PIM (Platform Independent Model).* Research was conducted by one author as a part of his dissertation thesis [15]. To achieve that objective, we divided it to the 4 partial objectives:

1. The use of ontologies in the IS development leads to new discipline - ontological engineering. Ontological engineering is mostly associated with knowledge engineering. If we consider MDA as the IS development principle - the first partial objective of the solution is to integrate the principles of MDA into ontology and knowledge engineering.
2. The general development of the IS according to principles of the ISO / IEC / IEEE 42010: 2011 standard - Systems and Software Engineering leads to the creation of an architectural framework that forms the methodology of the software-intensive systems development. Creation of an architectural framework for the development of knowledge-based systems using ontologies with the emphasis on the modeling of CIM and PIM level is the second partial objective.
3. It is necessary to confirm the theoretical solution. Therefore, application of the created architectural framework within specific domain with experimental verification of the analysis and design of IS is the third partial objective.
4. The use of ontology as a modeling tool in MDA can remove the transformations between CIM and PIM levels. Validate and confirm the use of ontologies over other modeling

languages in model driven architecture is the fourth partial objective.

3. Results and discussion

In the terms of study 1 which was based on user-oriented digitization process and with the usage of the ultimate algorithm we were able to generate keywords with 77.17% success rate. These keywords are requested by the users. In the case of real usage of the proposed approach within Slovak library we recommend to semi-automate the process - employee of the library should select only those keywords which are in accordance with the methodology to value creation of metadata elements. To fully automate this process, it would be possible to use algorithm made by specialists in the field of library and information science. This algorithm could use controlled vocabularies, thesaurus, word weighting, etc. Further research could examine various options of search results representation which could simplify the decision making process within choosing the right object from searched records set.

According to study 2 we can assume that ontology can be mapped to the model independent on the computer processing (CIM) and also on a Platform Independent Model (PIM). Consequently, it is possible to view these models in graphical form - as a business processes in BPMN notation or use cases in UML notation. Using this concept, it is possible to speed up the modeling process and also eliminate the need for transformation between CIM and PIM levels. In order to fully eliminate the need for transformation the mapping has to be extended to the whole scope of the CIM and PIM levels modeling in the respect of notation, restrictions, conventions, standards and file types. Nevertheless, the proposed solution is able to consolidate, clarify and accelerate the design, creation and maintenance of business processes and use case diagrams in terms of CIM / PIM transformations and provides a platform for further development in this research area. The advantage of the proposed solution is that the manual creation of the transformations between CIM and PIM levels is no longer necessary since all information from these levels are united under a single ontology. CIM and PIM levels can be generated directly from the ontology. Any changes made during the design phase of the IS are edited directly in the ontology and will be then reflected in the CIM/PIM models which are generated from this ontology.

4. Conclusion

The approach to generate values of the most desirable metadata elements for adding new digital library database is considered by us as a ready-to-implement solution even without implementing the recommendations listed in the Results and discussion, as we agree with the Big Data specialist opinion: "Rather to have a lot of worse data than a small amount of better data" [16].

According to our studies - we have confirmed that experts without satisfying user needs are not able to create or modify high quality information system. Users enter the development process more often than we could think. Users test even the partial results and their requirements shape every aspect of IS development. Due to this fact it is important that the user with his/her requirements has to be perceived as the basic prerequisites for the creation of IS.

Model-driven development which includes model driven architecture allows to divide analysis, design and implementation of IS to individual models and levels which abstract the complexity of the IS into the smaller, transparent units. These units are easy to maintain and allow to clearly define the actual stage of the IS development.

To ensure that the models will share information it is necessary to create interoperability between MDA levels. This interoperability is provided by model transformations which are usually made manually. The principle of using ontology as a modeling language in the model driven architecture enables to automate model creation and also eliminates the need to manually transform models within the various levels of abstraction.

Acknowledgements

This paper is supported by the following project: University Science Park of the University of Zilina (ITMS: 26220220184) supported by the Research&Development Operational Program funded by the European Regional Development Fund.



References

- [1] NIELSEN, et al.: *Evidence-Based User Experience Research, Training, and Consulting*. [Online]. Available: <https://www.nngroup.com/topic/research-methods/?page=2>. [Accessed 19 10 2016].
- [2] JUCHOVA JUROSKOVA, Z., ZAVODSKA, A.: *Using of Text Mining in Online Public Access Catalog*. Uden, L., et al. (Eds.): KMO 2015, LNBIP 224, pp. 451-462. Springer, Heidelberg (2015).
- [3] Statna vedecka kniznica v Banskej Bystrici.: 12th annual Conference for librarians, [Online]. Available: <http://www.svkbb.eu/aktualne-informacie/716-fungovanie-kniznic-v-zmenenych-ekonomickych-podmienkach-a-pecificke-poia>. [Accessed 01 10 2016].
- [4] JUCHOVA JUROSKOVA, Z.: *The Measurement of Information about Cultural Heritage*. [PhD. thesis], University of Zilina, p. 102 (2012).
- [5] ROBERTSON, S.: *Understanding Inverse Document Frequency: On Theoretical Arguments for IDF*. J. Documentation 60(5), 503-520 (2004). [Online]. Available: <http://www.emeraldinsight.com/doi/full/10.1108/00220410410560582>. [Accessed 06 10 2016].
- [6] ISO/IEC/IEEE 42010:2011. 2011, Systems and software engineering - Architecture description.
- [7] SURYN, W., KHERRAF, S., LEFEBVRE, E.: Transformation from CIM to PIM Using Patterns and Archetypes. *Software Engineering Conference, Australian*, 2008, 338-346.
- [8] RODRIGUEZ, A., FERNANDEZ-MEDINA, E., PIATTINI, M.: *CIM to PIM Transformation: A Reality*. IFIP, vol. 255, Research and Practical Issues of Enterprise Information Systems II, vol. 2, 1239-1249. Springer : Boston, 2008.
- [9] ZHANG, W., MEI, H., ZHAO, H., YANG, J.: Transformation from CIM to PIM: A Feature Oriented Component-Based Approach. *Model Driven Engineering Languages and Systems*, vol. 3713, 248-263. Springer (2005).
- [10] CAO, X.-X., MIAO, H.-K., XU, Q.-G.: *Modeling and Refining the Service-Oriented Requirement*. 6th intern. symposium on Theoretical Aspects of Software Engineering, 2008, 159-165. TASE'08. 2nd IFIP/IEEE International Symposium on. IEEE (2008).
- [11] KANG, D., et al.: An Ontology-Based Enterprise Architecture. *Expert Systems with Applications*, 2010, 37.2: 1456-1464.
- [12] HICKEY, A. M., DAVIS, A. M.: An Ontological Approach to Requirements Elicitation Technique Selection. *Ontologies*, 2007, 403-431, Springer.
- [13] VONGDOIWANG, W., BATANOV, D. N.: An Ontology-Based Procedure for Generating Object Model from Text Description. *Knowledge and Information Systems*, 2006, 10.1: 93-108.
- [14] VARGAS-VERA, M., et al.: MnM: Semi-Automatic Ontology Population from Text. *Ontologies*, 373-402, 2007, Springer.
- [15] BUCKO, B.: *Model Driven Architecture and Ontologies* [PhD. thesis]. University of Zilina : Zilina, SR, 2016, p. 107.
- [16] MAYER-SCHONBERGER, V., CUKIER, K.: *Big Data: A Revolution that will Transform How we Live, Work and Think*, Hodder&Stoughton, 2013, p. 256.