Conference on ENTERprise Information Systems / International Conference on Project MANagement / Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2016, October 5-7, 2016

Maintaining Requirements using Web Usage Data

Jorge Esparteiro Garcia[a,b,*], Ana C. R. Paiva[b,c]

[a]*Polytechnic Institute of Viana do Castelo, Praca General Barbosa, Viana do Castelo, Portugal*
[b]*INESC TEC, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal*
[c]*Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal*

Abstract

The use of the World Wide Web has had a huge growth and there is a greater variety of web applications with an increasing importance in society and in supporting the development to all kinds of business. Often, most of websites are providing support services that must be maintained and improved over time. This maintenance and upgrade can be difficult because frequently the requirements are no longer actual and/or often not even exist documented. Websites are increasingly monitoring usage data, and this type of information is increasingly abundant. Analyzing the usage of the websites can help identify improvements and help to maintain the website and its software requirements. This paper presents REQAnalytics, a recommender system that collects the information about the usage of a website, processes it and generates recommendations to the requirements specification of the website. This research work also presents an experimental evaluation of a case study based on an online newspaper website. The results showed that REQAnalytics can produce reports in a language closer to the business, identify most and less used functionalities, requirements that can be split in two or more requirements and give support to the maintenance of requirements of the website being analyzed.

* Corresponding author. Tel.: +351 258 809 610;
E-mail address: jorgegarcia@esce.ipvc.pt

1. Introduction

Software quality has become a topic of major concern in the field of Software Engineering. Particularly, websites are being used not only for displaying static information but increasingly as core business tools, particularly through web services, intranets and web applications that run as support applications to business development[1,2].

The websites and web applications are a type of software that must be available 24 hours a day. Furthermore, they should please the customer and must be maintained and improved to adjust to needed changes through the website lifecycle[3]. The main reasons of software project failures are the incorrect requirements elicitation, requirements changes and their uncontrolled evolution during the software project lifetime[4]. Uncontrolled requirement changes cause negative impacts in software development, such as, excessive costs and a system unable to answer to stakeholders' needs.

The requirements management allows to maintain stability and agreement among stakeholders' requirements, by means of the analysis of change effect and their monitoring during software lifetime[5]. Software organizations are improving the methods they use to collect, analyze, document, and maintain their requirements in a structured software requirements specification written in natural language. However, a Software Requirements Specification is difficult to keep current, specially when the software project is a website that evolves during its lifetime[6]. Through time, this causes the requirements become outdated and do not reflect the current state of websites. Furthermore, during software lifetime it is difficult to determine what requirements change requests should be answered first and the information of traceability between the requirements and the implementation is frequently lost.

Nowadays, there is little support for websites and web applications evolution, despite the evolution of website accounts for major development costs. The evolution of websites involves the development of new requirements, requirements' updates and implementation changes. These maintenance activities may come from different stakeholders, such as developers, systems engineers, users and service integrators.

There are methods for measurement, data collection and data analysis of websites and web applications throughout their lifetime. These methods are called Web Analytics. The main objective of Web Analytics is to provide the right direction to online users. This can be done by doing required and impactful changes in the web site[7].

Nowadays, existing Web Analytics tools are able to gather diverse data about the usage of a website. The use of websites generates large amounts of information that may be used for different purposes like assessment of quality of web-products[8], pattern recognition or to statistical analysis[7].

It is our belief that information about the usage of websites may help requirements maintenance which can be a contribution to the overall quality of the service provided. Existing approaches and tools do not take advantage of this data. In addition, Web Analytics tools have some limitations relating to services maintenance. Web Analytics tools have focused on analysis and reporting of business metrics, like number visits and traffic sources, which interest is mainly to marketers[9] but analysis directed to the improvement analysis is not currently done. Usage data is disregarded for the improvement of the quality of a web application. So, the potential of web usage data analysis is yet to be explored[10].

Recommendation Systems for Software Engineering (RSSE) can help developers to find alternative decisions in a wide range of software engineering tasks from reusing code to writing effective bug reports. The overall goal is to provide the right information, at the right time, to the right person. This would allow requirements engineers to spend their limited time on more important aspects of the project[17].

This paper presents REQAnalytics, a recommender system that collects the web data usage from a website and suggests requirements' changes. Recommender systems have been widely used in e-commerce websites to provide user personalization[11] like product, content or service recommendations. Recommender Systems for Software Engineering (RSSE) is a novel approach to support developers in decision making.

The remainder of this paper is organized as follows: related work is discussed in Section 2, Section 3 presents an overview of REQAnalytics, the recommender system developed to assist the requirements maintenance. Section 4 presents an experimental evaluation of a case study in an online newspaper and the results achieved. Conclusion and future directions are in Section 5.

2. Related Work

There are Recommendation Systems that have been widely used in e-commerce websites to provide user personalization[11] like product, content or service recommendations. Recommendation Systems provide information items estimated to be also valuable for a software engineering task in a given context[17]. Recommender Systems for Software Engineering (RSSE) is a novel approach to support developers in decision making.

SRRS[18] proposed a Security Requirements Recommendation System that uses prior knowledge about security requirements approaches, combined with user preferences, to recommend the most appropriate approach for specific project characteristics, and presents a full-order ranking of all approaches. Castro-Herrera[19] presented a hybrid Recommendation System to identify potential users who could reply to unanswered posts in online forums. A content-based recommender technique analyses the text of unanswered posts, and compares them to previous posts. Cleland-Huang[20] proposed a feasible approach that utilizes data mining and recommender systems to scale-up the fundamental processes of requirements elicitation and prioritization.

In addition, nowadays, the Requirements Management of a website is only done to solve detected problems, satisfy or correct the failures of the software requirements specification[21] not focused in its improvement, and therefore in the improvement of the overall quality of the website. Several studies[22,23] have been developed related to Recommender Systems of Software Engineering , however none of them produce recommendations to the Software Requirements Specification based on the web usage data of the website.

In other related work, Danylenko an Lowe[24] suggest recommendation systems based on context-aware composition to enable a system designer to postpone and automate decisions regarding efficiency non-functional requirements, such as performance. Though, this approach focus on handling efficiency non-functional requirements during the software development process and is carried out before deploy of the system.

A recommendation-based approach to requirements reuse is presented by Dumitru[25]. The proposed recommendation approach is content-based filtering, where a vector of keywords (derived from the description of the new software project) is matched with the keywords extracted from requirements artifacts from the repository of already completed software projects.

Despite the fact of several studies[17,23] have been carried related to RSSEs, none of them produce recommendations to the Software Requirements Specification based on techniques of web mining or based on web usage data.

Using the data gathered from a web analytics tool, REQAnalytics generates recommendations to the software requirements specification and, ultimately, to the website itself. These recommendations also allow to analyze typical paths taken by users (which may be useful to define or improve workflows) and suggest changes to the functional requirements and to the website in a language closer to the business.

3. The REQAnalytics System

This section describes REQAnalytics, an innovative approach through a web based Recommender System that supports the task of Requirements Management. REQAnalytics is a recommender system that, using the web usage data of a website and the information of the mapping of the functional requirements with the web pages and elements, suggests recommendations to the software requirements specification. Despite having some features commonly available in a Requirements Management tool, it cannot be categorized as so, due to the fact that its goal is to complement and support the task of requirements maintenance through several recommendations to the requirements.

The main features of REQAnalytics include:

- **Mapping Tool.** This tool is integrated within REQAnalytics and relates functional requirements with the web pages and elements of the website as described in another case study of REQAnalytics[12]. Also, a similar methodology is presented in a different research work[13].
- **Most used Requirements Navigation Paths Report.** A report that shows the most frequent paths taken by users along the web service. However, instead of showing the paths as sequences of web pages visited, as it is common in other tools, it shows the functional requirements executed along those paths.
- **Traceability Matrix Report.** Correlate the links between the functional requirements and the web pages and elements of the website

- **Requirements Analytics (Statistics and Main Metrics) Report.** This report lists the most accessed Functionalities (functional requirements), and other metrics like Entry Functionalities, Exit Functionalities and Requirements Bounce Rate.
- **Report of Recommendations to the Software Requirements Specification based on web usage data.** This report shows the main recommendations to the Software Requirements Specification. Some results have already been achieved in two different case studies of REQAnalytics[14,15]:
    - o Create New Requirement
    - o Requirements Prioritization Change
    - o Delete existing Requirement
    - o Split exiting Requirement

This system is designed to be used through a web browser. Developed in PHP and using a MySQL database as support, REQAnalytics analyses the web usage and navigation of a specific website to generate recommendations to improve the quality of the software requirements specification.

To generate these recommendations, we developed a web based mapping tool included in REQAnalytics that allows to map the functional requirements of the website with its functionalities. To collect the web data usage from the website, REQAnalytics uses a web analytics tool, OWA - Open Web Analytics that allows to gather this kind of data and save it to a database.

As presented in other research works[14,15], the REQAnalytics system is divided in four different phases:

- Requirements mapping - map the functional requirements with the functionalities (pages and HTML elements) of the website.
- Collect Web usage data - use of the web analytics tool OWA for collecting web usage data (pages viewed, clicked web elements, traversed paths along web pages, session duration, entry pages, exit pages).
- Analysis of the data collected - the data provided by OWA is analyzed and intersected with the mapping information defined during requirements mapping.
- Generation of recommendation report – generation of a high level recommendations report with possible improvements of the requirements specification and, ultimately, of the website itself.

3.1. Mapping of requirements

The functional requirements of the website under analysis are mapped with the web pages and HTML elements present in the website. The main idea is to trace functional requirements with the implementation artifacts like web pages and elements of the website. Identifiers for each traceability element such as requirement identifiers, web page URL and HTML element identifiers are stored within the database and correlated to the functional requirements.

This mapping is a very important task in the project, since it will allow to establish correlations between the functional requirements and the implementation in order to allow REQAnalytics analyze and then generate recommendations to the software requirements specification baseline. This mapping is established using a high-level mapping web tool included in the system. This web tool can work with any web application or website since it was developed inside a bookmarklet that works in any web browser.

3.2. Collect Web usage data

The purpose of this phase is to collect the web usage data available using a web analytics tool. The analytics tool used in this system is Open Web Analytics[16]. It will collect several data like pages visited, DOM elements clicked, traversed paths along web pages, entry pages, exit pages and duration of session. These data will be stored in a support database for further analysis.

3.3. Analysis of the data collected

The REQAnalytics lists and analyses the web usage data collected by the web analytics tool, with the mapping information stored in the first phase. In this analysis, we intend to identify possible improvements of the website. For

this purpose, we analyze the paths traversed by users while visiting the website to be able to identify, among other improvements, possible shortest paths, workflow changes, which are the most and least used functionalities (functional requirements), create new functionalities not included in the previous requirements specification, change the priority of the requirements and detect patterns on the navigation paths.

3.4. Generation of recommendation report

After the analysis made in the previous phase, a detailed report is generated with several improvement recommendations for the website under review. This report is built in a language closer to the business than the language used by existing web analytics tools because it is based on the requirements associated with the web pages and elements. Examples of recommendations are:

- Show the most and least used website functionalities, mapped with the respective requirement
- Creation of new requirements
- Eliminate requirements whose functionality is not used
- Change the priority of the functional requirements
- Detect the most used navigation paths (i.e., the most used sequence of performed functionalities along the website).

4. Case Study

This section presents the results of the application of REQAnalytics in a news website: Cidade Tomar, a regional newspaper of Portugal. The aim is to assess the real ability of the system to suggest recommendations to the software requirements specification and to the navigation paths. As described in Section 3, the first phase of the REQAnalytics is to identify the functional requirements of the website that will be used by the analysis of the recommender system engine.

The functional requirements of the software requirements specification are firstly exported to an XML document in the format accepted by REQAnalytics. Then the mapping between each requirement and the web page and element that implements it is established. This mapping is made using the web based mapping tool include in the REQAnalytics system.

The web usage data for this case study is collected using OWA. The time period used in this study was 2 weeks (14 days) before the recommender system analysis has started.

The goals defined for this specific case study were:

- Display the Top Entry Functionalities of the website
- Display the Top Exit Functionalities of the website
- Suggest Recommendations to split requirements

4.1. Results

REQAnalytics relates the information gathered by the web analytics tool with the mapping information in order to perform the analysis. After the analysis, the REQAnalytics generates reports with recommendations to the software requirements specification and website.

4.1.1. Entry Functionalities

Displays the top entry functionalities of the requirements specification. The table shows the list requirements related to web pages that were visited as "first pages". The list is sorted downward by the number of clicks.

Fig. 1. Report generated by REQAnalytics with the top Entry Functionalities

4.1.2. Exit Requirements

Shows the top exit functionalities of the requirements specification. The table shows the list requirements related to web pages that were visited as "last pages". The list is sorted downward by the number of clicks.

Fig. 2. Report generated by REQAnalytics with the top Exit Functionalities

4.1.3. Split a Requirement in two or more Requirements

The REQAnalytics system detects when a web page (mapped to a requirement) has a considerable number of visits without user leaving to other functionalities of the website. Currently, REQAnalytics assumes that, when this number

is above ten, it is a likely case that there is need to detail the requirements specification (e.g., dividing a requirement in two). Figure 3 exemplifies this situation.
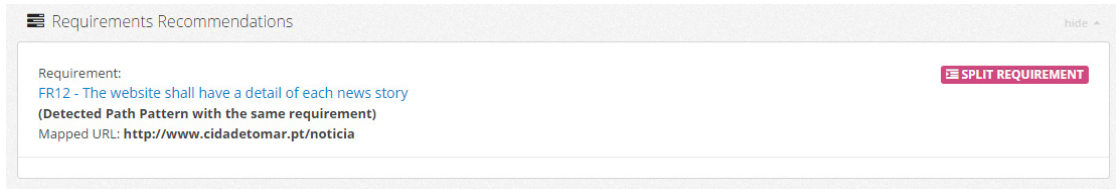

Fig. 3. Recommendation to Split a Functional Requirement in two or more New Requirements

It is assumed by the system that when this situation occurs it may be due to three types of possible events on the website or web application under analysis:

- **Functionalities not previously mapped.** There are functionalities related with the same web page and web elements that were not described by any requirement.
- **Functionalities not initially documented in the Software Requirements Specification.** There are some functionalities that were implemented on the website that were not included in the initial specification baseline and therefore the user of REQAnalytics could not have mapped them and include them in the analysis.
- **New functionalities not documented in the Software Requirements Specification.** When there are new functionalities that were not yet inserted in the requirements specification.

These results suggest that recommender systems for software engineering can be used in a meaningful way to help the requirements maintenance during the lifecycle of a website. The recommendations provided by the REQAnalytics helps to maintain the software requirements specification updated. Documenting software requirements reduce the project risk by reducing uncertainty in implementation of the software. In addition, software requirements documented and updated may ensure that they are addressed during software design and testing.

5. Conclusions

This paper presents a recommender system, REQAnalytics. It maps functionalities/requirements with web pages and elements of a website and relates this information with the web usage data in order to generate recommendations to the website under analysis.

Commonly, traceability information is kept among requirements and test cases and the majority of the existing tools are incapable of automatically generate and maintain traceability relationships. The presented methodology for requirements traceability and software requirements management can be applied in an evolutionary context that supports and manages functional requirements during software lifetime.

Regarding reports produced, existing web analytics tools generate mainly navigation statistics using the codes of the web elements which is difficult to read and analyze. REQAnalytics presents reports in a language closer to the business because it is based on functionalities. With this king of recommendations, it is easier to meet the expectations of the customers and users because a common language is used.

REQAnalytics also helps in the task of requirements management, which contributes to the quality of the web service itself. However additional work is needed to compare these results to those obtainable using other type of recommendations.

As future work we intend to use other Web Mining methods and techniques to analyze the web usage data. Since there are emerging new methods and techniques applied to web usage mining, an opportunity for future work would be apply other Web Mining methods to achieve new type of recommendations to the requirements specification. REQAnalytics can only be used in websites or web applications. Applying this same approach in mobile applications would be also a promising research work.

References

1. Singh B, Singh HK. Web Data Mining research: A survey. In: *2010 IEEE International Conference on Computational Intelligence and Computing Research*. IEEE; 2010:1-10. doi:10.1109/ICCIC.2010.5705856.

2. Taghipour N, Kardan A, Ghidary SS. Usage-based web recommendations. In: *Proceedings of the 2007 ACM Conference on Recommender Systems - RecSys '07*. New York, New York, USA: ACM Press; 2007:113. doi:10.1145/1297231.1297250.

3. Malviya BK, Agrawal J. A Study on Web Usage Mining Theory and Applications. In: *2015 Fifth International Conference on Communication Systems and Network Technologies*. IEEE; 2015:935-939. doi:10.1109/CSNT.2015.247.

4. Standish Group. CHAOS Report. *Group* 1995:11-13. doi:0674.

5. Ibrahim N, Wan Kadir WMN, Deris S. Propagating requirement change into software high level designs towards resilient software evolution. In: *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*.; 2009:347-354. doi:10.1109/APSEC.2009.55.

6. Wiegers KE. *Automating Requirements Management*.; 1999.

7. Kumar L, Singh H, Kaur R. Web analytics and metrics. In: *Proceedings of the International Conference on Advances in Computing, Communications and Informatics - ICACCI '12*. New York, New York, USA: ACM Press; 2012:966. doi:10.1145/2345396.2345552.

8. Singal H, Kohli S, Sharma AK. Web analytics: State-of-art & literature assessment. In: *2014 5th International Conference - Confluence The Next Generation Information Technology Summit (Confluence)*. IEEE; 2014:24-29. doi:10.1109/CONFLUENCE.2014.6949041.

9. Pai D, Ravindran B, Rajagopalan S, Srinivasaraghavan R. Automated faceted reporting for web analytics. In: *Proceedings of the 4th International Workshop on Web-Scale Knowledge Representation Retrieval and Reasoning - Web-KR '13*. New York, New York, USA: ACM Press; 2013:9-16. doi:10.1145/2512405.2512406.

10. Akerkar R, Badica C, Burdescu DD. Desiderata for research in web intelligence, mining and semantics. In: *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics - WIMS '12*. New York, USA: ACM Press.

11. Pazzani MJ, Billsus D. Content-based recommendation systems. 2007:325-341.

12. Garcia JE, Paiva ACR. A Requirements-to-Implementation Mapping Tool for Requirements Traceability. *J. Softw.* 2016;11(2):193-200. doi:10.17706/jsw.11.2.193-200.

13. Paiva ACR, Faria JCP, Tillmann N, Vidal RAM. A model-to-implementation mapping tool for automated model-based GUI testing. In: Lau K-K, Banach R, eds. *ICFEM'05 Proceedings of the 7th International Conference on Formal Methods and Software Engineering*.Vol 3785. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005:450-464. doi:10.1007/11576280.

14. Garcia JE, Paiva ACR. REQAnalytics: A Recommender System for Requirements Maintenance. *Int. J. Softw. Eng. Its Appl.* 2016;10(1):129-140. doi:10.14257/ijseia.2016.10.1.13.

15. Garcia JE, Paiva ACR. An Automated Approach for Requirements Specification Maintenance. In: *WorldCist'16 - 4th World Conference on Information Systems and Technologies*.; 2016. doi: 10.1007/978-3-319-31232-3_78.

16. Open Web Analytics. Available from http://www.openwebanalytics.com/. 2015. Available at: http://www.openwebanalytics.com/.

17. Robillard M, Walker R, Zimmermann T. Recommendation Systems for Software Engineering. *IEEE Softw.* 2010;27(4):80-86. doi:10.1109/MS.2009.161.

18. Romero-Mariona J, Ziv H, Richardson DJ. SRRS: a recommendation system for security requirements. In: *Proceedings of the 2008 International Workshop on Recommendation Systems for Software Engineering - RSSE '08*. New York, New York, USA: ACM Press; 2008:50.

19. Castro-Herrera C. A hybrid recommender system for finding relevant users in open source forums. In: *2010 Third International Workshop on Managing Requirements Knowledge*. IEEE; 2010:41-50. doi:10.1109/MARK.2010.5623811.

20. Cleland-Huang J, Mobasher B. Using data mining and recommender systems to scale up the requirements process. In: *Proceedings of the 2nd International Workshop on Ultra-Large-Scale Software-Intensive Systems - ULSSIS '08*. New York, New York, USA: ACM Press; 2008:3-6. doi:10.1145/1370700.1370702.

21. Khatoon A, Motla YH, Azeem M, Naz H, Nazir S. Requirement change management for global software development using ontology. In: *2013 IEEE 9th International Conference on Emerging Technologies (ICET)*. IEEE; 2013:1-6. doi:10.1109/ICET.2013.6743519.

22. Maalej W, Thurimella AK, eds. *Managing Requirements Knowledge*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. doi:10.1007/978-3-642-34419-0.

23. Robillard MP, Maalej W, Walker RJ, Zimmermann T, eds. *Recommendation Systems in Software Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. doi:10.1007/978-3-642-45135-5.

24. Danylenko A, Löwe W. Context-aware recommender systems for non-functional requirements. 2012:80-84.

25. Dumitru H, Gibiec M, Hariri N, et al. On-demand feature recommendations derived from mining public product descriptions. In: *Proceeding of the 33rd International Conference on Software Engineering - ICSE '11*. New York, New York, USA: ACM Press; 2011:181. doi:10.1145/1985793.1985819.