Analytics in Continuous Requirements Engineering

Marite Kirikova and Kaspars Purmalietis

Riga Technical University, Latvia
marite.kirikova@cs.rtu.lv, kaspars.purmalietis@rtu.lv

Abstract. In continuous requirements engineering, timely knowledge about required and possible changes is essential. However, even one of the most important sources of knowledge – users of the systems, not always are aware of existing problems and all relevant future trends that can cause the need for changes. Therefore, to be more aware of possible changes, different analytics approaches are becoming embraced by requirements engineering. This paper examines what are the potential sources of data to be investigated and what are potential methods for defining requirements for analytics tasks in continuous requirements engineering. The discussion is based on the FREEDOM framework that has previously been designed for the purposes of continuous requirements engineering.

Keywords: Requirements engineering · continuous engineering · requirements analytics.

1 Introduction

Continuous requirements engineering implies ability to rapidly react to any changes that might be required from a system, its organizational context, or its external environment [1]. So, here the basic task is to identify new requirements as soon as possible to ensure timely development reactions with respect to the decision making and implementation. Often the elicitation of requirements is hindered by missing knowledge about the requirements, e.g. as in cases when "users do not know what they want". Developers try to overcome this difficulty by using agile approaches such as user stories, to find the problems to be solved as soon as possible, and heavily relaying upon the socialization among the users and developers. User stories oriented approaches work well for the set of cases, where users are available and can, at least to some extent, define the problems to be solved. However, not all systems development contexts are so comfortable, as in many situations the need for changes might be not known, wanted, or properly externalized. Therefore requirements engineers start to embrace new approaches such as analytics. There are several ways how the analytics might be used in requirements engineering, e.g. analytics of requirements engineering process, analytics of organizational processes (e.g. KPIs of organizational processes), analytics of systems usage (e.g., analytics of usage of WebPages), etc.

In this paper we will analyze possible applications of analytics in requirements engineering based on FREEDOM framework [1], which has been developed for the purposes of continuous requirements engineering. The goal of the paper is to amalgamate and structure the types of analytics tasks and approaches applicable in continuous requirements engineering.

The paper is organized as follows. In Section 2 the related work on analytics in requirements engineering and requirements for analytics is discussed including the methods used for analytics purposes. In Section 3, some recommendations regarding usage of analytics in continuous requirements engineering are provided. The conclusions are presented in Section 4.

2 Related Work on Analytics in the Context of Requirements Engineering

Here the related work is considered from three aspects. First, in Section 2.1 the state of the art of the usage of analytics in requirements engineering is discussed. Next, in Section 2.2, the typology of analytics approaches/techniques is addressed, and in Section 2.3 the methods for defining requirements for analytics tasks are considered. Section 2.3 is included here, because, if we wish to apply analytics in requirements engineering – the requirements for these analytics tasks of requirements engineering also have to be defined. Thus, it is a requirements engineering for requirements engineering; and can be considered as a meta-requirements engineering activity.

2.1. Analytics in Requirements Engineering

The analytics in requirements engineering can help to identify the requirements which otherwise might be overlooked or un-recognized. For instance, in large software systems (systems size, functionality breadth, component maturity, supplier heterogeneity) it is advisable to apply software repositories mining for understanding, evaluating, and predicting the development, management, and economics of such systems [2]. There are different metrics that can be used in defining and analyzing requirements. The simplest ones are direct requirements metrics that, for instance, give evidence on the number of different types of requirements (e.g. “shall” vs. “should”). Other metrics may concern software defects, and their origin that shows, which requirements caused the software problems and what are the problems to be addressed in future. For instance, the author of [2] illustrates that, in one of the large scale systems, data analysis has revealed that approximately 50% of faults are injected and detected by the requirements phase; developers detect 95% of in-phase faults and 47% of out-of phase faults, so showing that there is a room for improvement in the requirements engineering process. However, analysis of requirements quality per se, does not addressed all issues relevant for the continuous requirements engineering.

In [3] Web usage data analytics is applied for requirements monitoring. The authors propose REQAnalytics, an innovative approach through a web based Recommender system that supports the task of requirements management. EQAnalytics 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 their elements, suggests recommendations to the software requirements specification.

The REQAnalytics system is divided in four different phases: (1) Requirements mapping - mapping the functional requirements with the functionalities (pages and HTML elements) of the website; (2) Collecting Web usage data - using of the web analytics tool for collecting web usage data (pages viewed, clicked web elements, traversed paths along web pages, session duration, entry pages, exit pages); (3) Analyzing the data collected - the data provided by the tool is analyzed and intersected with the mapping information defined during requirements mapping; (4) Generation of recommendation report – generation of a high level recommendations report with possible improvements of the requirements specification and, ultimately, of the website itself. The tool can provide such recommendations as to create new requirement, to change the priority of the requirement, delete existing requirement, and split existing requirement.

In general, the requirements monitoring is a well researched area. In [4] an overview of 37 monitoring frameworks revealed from 330 publications is presented. However, the authors conclude that most of existing approaches are restricted to certain kinds of checks, particular types of events and data, and often are limited to particular architectural styles and technologies. Thus it is not so simple to apply the approaches in complex situations, where many different systems shall interact, e.g. in the context of systems of systems.

One more stream in application of analytics in requirements engineering is the use of visual analytics. Visual analytics for requirements tracing is discussed in [5]. Authors of [6] present the visual requirements analytics framework that explicitly models the user, emphasizing that machine computations only augment, but cannot replace human capabilities to perceive, relate, and conclude in the knowledge discovery and decision making process. The paper also provides a survey of eight visual analytics approaches used in requirements engineering applying such visualization approaches as multiple views, inter-view navigation, browsing, searching, query-drilling, filtering, and annotation. These approaches can help in anomaly detection, revealing project plan based social networks, risk assessment, revealing crosscutting concerns, understanding scope changes, mitigating user difficulty in analyzing particular models, and facilitating visual exploration.

We can see that the spectrum of analytics approaches used in requirements engineering is large, and many of these approaches might be useful in continuous requirements engineering.

In the next (sub)section, some analytics approaches/techniques will be discussed and the examples of their possible usage in continuous requirements engineering will be given.

2.2. Classification of Analytics Approaches

The analytics usually is based on a particular data mining techniques. Below the list of data mining approaches/techniques presented in [7] will be used to exemplify potential use of these techniques in continuous requirements engineering. Thus for each technique in the list, the corresponding example of its usage in requirements engineering will be given.

The following approaches/techniques [7] and examples are considered:

- Classification and class probability estimation – showing “yes” or “no” an item belongs to a class. The target is categorical (often binary, but not always), For instance, is it a mandatory requirement or a wish.
- Regression – predicts how much something will happen – predicts the extent of something (numerical target). For instance, how many hours the implementation of a particular requirement may take.
- Similarity matching – attempts to identify similar individuals based on data known about them. For instance, finding requirements which are similar to wrongly defined requirements.
- Clustering – attempts to group individuals together based on their similarities, but it is not purpose driven. For instance, similar faults that could point to a systemic error in the requirements.
- Co-occurrence grouping (frequent item set mining or association rule discovery or market-basket analysis) – attempts to find associations between entities based on transactions involving them. For instance – which application features are usually used one after another?
- Profiling (or behaviour description) – what is the typical behaviour of an individual, group, or population. For instance, what are web browsing habits of a particular user group?
- Link prediction – predicts connection between data items. For instance, connecting an implementation problem to a quality characteristic of a requirement.
- Data reduction – replacing a larger data set with a smaller, but with more insightful one. For instance, changing specific user preferences with role preferences. Usually involves data loss, but gains insight.
- Causal modeling attempts to show, which actions influence others. This includes randomized controlled experiments. The approach is based on the assumptions and drawing the conclusions always has to present also the assumptions, to validate them. For instance, causal modeling can be applied for analyzing what are the causes of wrongly defined requirements, what are the group behaviours signalizing the need for extra communication, etc.

These approaches/techniques can be divided on supervised and non supervised methods. Classification, regression and causal modelling usually are supervised methods; similarity matching, link prediction, and data reduction can be both; and clustering, co-occurrence grouping, and profiling usually are applied as unsupervised methods.

The above-mentioned approaches/techniques are mainly used for structured data. However, also analysis of semi-structured information and unstructured information (e.g. text mining and sentiment analysis) is relevant in continuous requirements engineering.

As was suggested in [6], the analytics has to involve also human actors. Therefore we introduce here one more classification of the methods, namely automated, semi-automated, and manual. This allows for even wider differentiation of the analytics tasks, higher variability in the granularity of the analytics tasks, and a possibility to relate them to the feature models that can be used as a backbone for analytics requirements identification [8].

2.3. Methods for Defining Requirements for Analytics

In [9] a number of approaches for defining data analytics in warehouses have been amalgamated. The following methods are proposed and discussed (see also Fig. 1):

- Goal driven approaches: emphasis here is on the need to comply with organizational goals and strategies. Authors claim that these approaches work well only if business processes are designed and are combined with the goals.
- User driven approaches (or Demand Driven) approach – concentrates on participative requirements elicitation, data sources usually are not discussed there (they are considered later). This also may refer to so called sense making approaches [10].
- Data driven approaches (Supply driven) – the emphasis is on data sources and availability of data; their usefulness assessment is not the first issue discussed here.
- Process driven approaches – usually relate data to business process models. These are the most used approaches in practice. They imply definition of business process model, revealing main decisions, and then the actual data relevant in the decision making [11].
- Ontology driven approaches. These approaches can be based on self-developed ontologies or/and utilize standard domain ontologies.
- Technology driven approaches – aiming at identifying the current technology state that is reasonably applicable for the data gathering and storing.

Besides above mentioned ones, we can find also the following approaches:

- Feature driven approaches. These approaches usually utilize the feature model. In most of cases they are applied in software product line management, including the definition of requirements for analytics [12].
- Decision driven approaches [13] – treating decisions as "the first class citizens" and using decision models as the starting point of requirements identification; thus differring from the business process model based approaches which use process models as the starting point of requirements identification.

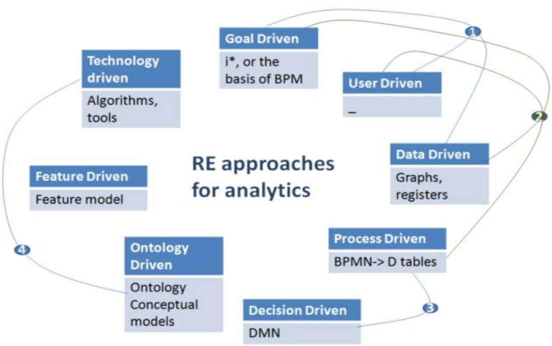

Fig. 1. Requirements engineering approaches for analytics

Above-mentioned requirements engineering approaches are represented in Fig. 1. Most of approaches relay upon particular modeling techniques. Some of these techniques are named below the titles of the approaches in Fig.1. Usually, for defining the requirements for analytics, several requirements engineering approaches are combined. The following combinations are illustrated in Fig. 1: (1) combining goal driven, user driven, and data driven approaches; (2) combining the approaches mentioned in (1) with the process driven approaches; (3) combining process driven and decision driven approaches; and combining (4) ontology driven and technology driven approaches. In the next section single approaches and their combinations will be suggested for particular data targets in continuous requirements engineering.

3 Application of Analytics in Continuous Requirements Engineering

We will use the FREEDOM framework [1] to illustrate the potential application of analytics in continuous requirements engineering (Fig. 2).

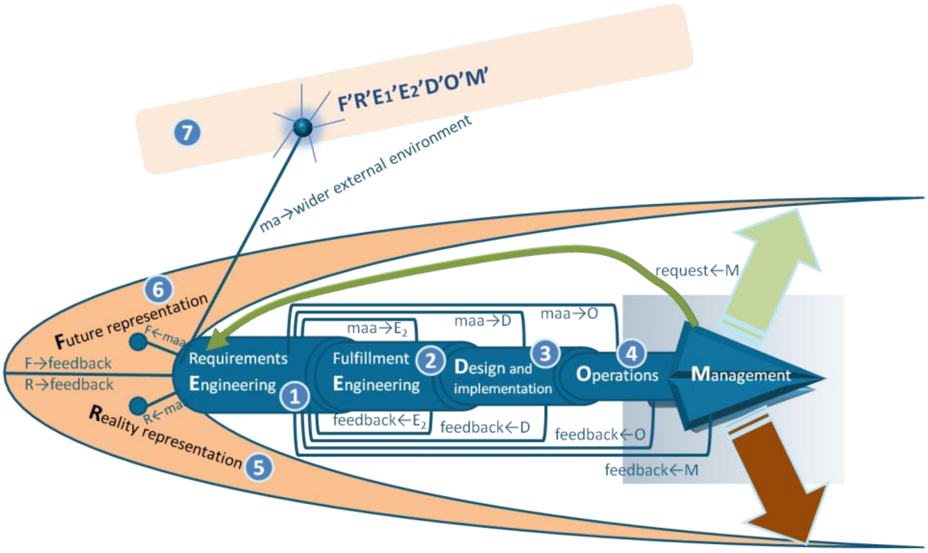

Fig. 2. Continuous requirements engineering in FREEDOM framework (ma - monitoring and analytics, maa - monitoring, analytics, and audit). For description of the framework see [1].

The framework represents several functions relevant in continuous requirements engineering. These functions are used here as analytics subdomains denoted by numbers 1-7 in the small circles in Fig. 1. We will focus here on the following issues:

- Application subdomain that illustrates what exactly in the FREEDOM framework is addressed by the analytics task. We have identified 7 subdomains (reflected with the numbers in Fig. 2, which represents the FREEDOM framework; the same numbers are used also in the first column of Table 1, namely: (1) Requirements Engineering function, (2) Fulfillment Engineering function, (3) Design and Implementation function (4) Operations function, (5) Reality representation function, (6) Future representation function, and (7) The external environment, where all phenomena that concern 6 above mentioned functions are relevant (e.g., new requirements methods, new ontologies, new software platforms, target group opinions, etc).

- Main data sources to be utilized in each subdomain.
- The requirements engineering approaches applicable in the subdomains. This is sort of meta-requirements engineering activity for the continuous requirements engineering.

The above-mentioned issues are amalgamated in Table 1 that has three columns: the first column represents a subdomain (by its number which corresponds to the number represented if Fig. 2), the second column represents the main data sources/data to be targeted in a particular subdomain, and the third column represents the requirements identification approach suitable for a particular subdomain.

Table 1. Analytics in continuous requirements engineering

Subdomain	*Data source/data to be targeted*	*Examples of possible requirements engineering approaches to be used*
1	Requirements specification, requirements elicitation process	User driven, feature driven
2	Project portfolios, Backlogs, project planning process	Decision driven, goal driven
3	Design artifacts, code, test cases, design process, implementation process	Technology driven
4	Business process, KPIs	Goal driven, process driven, decision driven
5	Artifacts representing the reality	Data driven, ontology driven
6	Artifacts representing the future	Data driven, ontology driven
7	Structures, semi-structured and unstructured data/information relevant to FREEDOM functions	Ontology driven, user driven, might be also data driven

Table 1 concerns only requirements engineering approaches; and it does not prescribe specific analytics methods or techniques, as many different techniques can be used in one and the same subdomain and vice versa. Nevertheless, as discussed in Section 2, when choosing the methods, it is necessary to take into consideration whether structured, semi-structured, or unstructured data/information is to be treated; and whether the fully automated, semi-automated, or even manual analytics methods are to be applied. To incorporate manual methods, the feature driven requirements engineering approach can be applied to define analytics requirements for the continuous requirements engineering. For this purpose a special, analytics services oriented, type of a feature model can be used that allows representing automatic, semi-automatic, and manual analytics services [8].

In Table 1 the requirements engineering approaches are assigned to subdomains based on the specifics of data/information that is the target for analytics in a particular subdomain. Only some examples of possible approaches are given on the basis of experiences reported in related works that were discussed in Section 2.

There are two levels at which the requirements engineering approaches for analytics can be applied in continuous requirements engineering. First, the approaches may be a part of continuous requirements engineering activities that are performed with respect to particular subdomain. The second level is requirements engineering for requirements and requirements engineering process itself (e.g. as in [2]). In both cases the requirements are to be stated for the activities of requirements engineering instead of requirements for the systems to be developed. Therefore the methods reflected in Table 1 can be considered as a part of meta-requirements engineering (requirements for requirements engineering) activities in continuous requirements engineering.

4 Conclusions

In this paper the possible usage of analytics in continuous requirements engineering was discussed. The discussion concerned the related work on usage of analytics in requirements engineering; also the methods of analytics were briefly overviewed in the context of requirements engineering; and the approaches for analytics requirements engineering where discussed. From the above discussion the following conclusions can be drawn:

1. The use of analytics becomes more and more common in requirements engineering.
2. In continuous requirements engineering the analytics can bring in knowledge that otherwise might be overlooked.
3. There are several subdomains – targets for analytics in continuous requirements engineering.
4. In each subdomain different sets (or single) requirements engineering approaches can be used and vice versa.
5. Different analytics approaches/techniques can be applied in a single subdomain and vice versa.
6. Requirements engineering for analytics in continuous requirements engineering can be regarded as a specific meta-requirements engineering activity when the analytics tasks are performed for the purposes of requirements engineering, i.e. the results of analytics are directly used by requirements engineers.
7. This meta-requirements engineering activity globally (for all subdomains as an integrated system) might be organized using feature driven approach.

Future research concerns (1) analysis of costs of introducing analytics tasks in the practice of continuous requirements engineering; and (2) development of methods for estimating feasibility of different analytics activities during continuous requirements engineering.

Acknowledgment: The research presented in this paper is sponsored in part by the Latvian National research program SOPHIS under the grant agreement No.10-4/VPP-4/11, and in part by the funding from the research project "Competence Centre of Information and Communication Technologies" of EU Structural funds, contract No. 1.2.1.1/16/A/007 signed between IT Competence Centre and Central Finance and Contracting Agency, project No. 1.14.

References

1. Kirikova, M.: Continuous Requirements Engineering in FREEDOM Framework: A Position Paper. In: Joint Proceedings of REFSQ-2016 Workshops, Doctoral Symposium, Research Method Track, and Poster Track co-located with the 22nd International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2016), Sweden, Gothenburg, 14–17 March, 2016, pp.1–6. RWTH, CEUR Press, Aachen (2016)
2. Selby, R.W.: Analytics-Driven Dashboards Enable Leading Indicators for Requirements and Designs of Large Scale Systems, IEEE Software Jan./Feb. 2009, pp. 41–49 (2009)
3. Garcia, J.E. and Paiva, A.C.R.: Maintaining Requirements Using Web Usage Data, Procedia Computer Science 100, pp. 625–633 (2016)
4. Vierhauser, M., Rabiser, R., Grunbacher, P.: Requirements Monitoring Frameworks: A Systematic Review, Information and Software Technology, 80, pp. 89–109 (2016)
5. Cleland-Huang, J., Habrat, R.: Visual Support in Automated Tracing, REV, 2007, no. 4. (2007)
6. Niu, N., Reddivari, S., Chen, Zh.: Keeping Requirements on Track via Visual Analytics. In: Proceedings of RE 2013, pp. 205–214. IEEE (2013)
7. Provost, F., Fawcett, T.: Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking, 1st ed., Loukides, M., Blanchette, M., Hearse, C. (eds.) United States of America: O'Reilly Media (2013)
8. Rudzājs, P., Kirikova, M.: Variability Handling in Multi-Mode Service Composition. In: Proceedings of 2nd International Conference on the Human Side of Service Engineering, Poland, Krakow, 19–23 July, 2014, pp.1–10 (2014)
9. Pestana, G., Catelas, P., Rosa, I.: A Multi-Driven Approach to Improve Data Analytics for Multi-Value Dimensions. In: Rocha, A., Correia, A.M., Costanzo, S., Reis, L.P. (eds.) New Contributions in Information Systems and Technologies, vol. 2, pp. 115–128. Springer International Publishing (2015)
10. Lau, L., Yang-Turner, F., Karacapilidis, N.: Requirements for Big Data Analytics Supporting Decision Making: A Sensemaking Perspective. In: Mastering Data-Intensive Collaboration and Decision Making: Research and Practical Applications in the Dicode Project, pp. 49–70. Springer International Publishing (2014)
11. Pape, T.: Prioritising Data Items for Business Analytics: Framework and Application to Human Resources. In: European Journal of Operational Research, vol. 252, no. 2, pp. 687–698 (2016)
12. Afzal, U., Mahmood, T., Zubair, Sh.: Intelligent Software Line Configuration: A Literature Review, Computer standards and interfaces, 48, pp. 30–48 (2016)
13. Taylor, J., Kitching, M.: Customer Retention in the Airline Industry: The Value of Predictive Analytics and How Using Decision Modeling Helps You Succeed (2015) http://www.slideshare.net/jamet123/the-value-of-predictive-analytics-and-decision-modeling