Automatic generation of software interfaces for supporting decision-making processes. An application of domain engineering and machine learning.

Andrea Vázquez-Ingelmo
GRIAL Research Group, Department of Computer Science, Research Institute for Educational Sciences. University of Salamanca
Paseo de Canalejas 169, 37008 Salamanca, Spain
andreavazquez@usal.es

Francisco J. García-Peñalvo
GRIAL Research Group, Department of Computer Science, Research Institute for Educational Sciences. University of Salamanca
Paseo de Canalejas 169, 37008 Salamanca, Spain
fgarcia@usal.es

Roberto Therón
GRIAL Research Group, Department of Computer Science, Research Institute for Educational Sciences. University of Salamanca
Paseo de Canalejas 169, 37008 Salamanca, Spain
theron@usal.es

ABSTRACT

Information dashboards are sophisticated tools. Although they enable users to reach useful insights and support their decision-making challenges, a good design process is essential to obtain powerful tools. Users need to be part of these design processes, as they will be the consumers of the information displayed. But users are very diverse and can have different goals, beliefs, preferences, etc., and creating a new dashboard for each potential user is not viable. There exist several tools that allow users to configure their displays without requiring programming skills. However, users might not exactly know what they want to visualize or explore, also becoming the configuration process a tedious task. This research project aims to explore the automatic generation of user interfaces for supporting these decision-making processes. To tackle these challenges, a domain engineering, and machine learning approach is taken. The main goal is to automatize the design process of dashboards by learning from the context, including the end-users and the target data to be displayed.

KEYWORDS

Automatic generation; Domain engineering; Meta-modeling; Information Dashboards; High-level requirements.

1 Introduction

The introduction of information systems to support and automatize the great diversity of activities has led an increment regarding the volume of generated data. However, the possession of significant quantities of data is not valuable if they are not analyzed and leveraged to create knowledge [1].

Data analysis has gained relevance over the years in different sectors [2] with the vision to accomplish robust information bases to support strategic decision-making processes [3].

Different people and profiles can participate during the process of making decisions, especially in interdisciplinary contexts. Given this situation, communication gaps could raise among the existing profiles, as they have different knowledge levels regarding different domains [4].

Each person involved must understand which implications the collected and analyzed data have to get the most out of decision-making processes. However, the gathered variables that might be relevant for supporting decisions could be complex, and users could need assistance to explore patterns and relationships among them.

One of the most powerful software tools for generating knowledge and exploiting datasets are information dashboards [5, 6]. Information dashboards are displays composed of graphical resources [7] and metrics to present information understandably [8], allowing pattern recognition or relevant indicators for decision-making processes.

However, designing a dashboard is not a trivial task [6]; it is necessary to take into account the users' necessities to assist them in reaching their goals. But this search process of needs process is usually complex, given the impossibility of knowing right from the beginning the exact kind of visualizations or metrics that would be beneficial for each individual user.

Current studies have identified different challenges regarding the design process of these tools, like functional flexibility, the influence of a user's knowledge or literacy about certain domains, the social impact, etc. [5]. Indeed, it is necessary to bear in mind that knowledge generation processes can vary depending on the person. Also, each person's goals regarding the same data can be completely different. However, building individual dashboards from scratch for every potentially involved profile would be unfeasible, requiring great quantities of time and resources.

The main focus of this research project is to tackle the automatic generation of personalized dashboards to raise the effectiveness of these tools as well as the productivity regarding their development, establishing relations among the users' concrete goals, preferences, abilities, etc. with the set of potential features and elements that would finally compose the personalized dashboard.

As it will be detailed, two paradigms are considered to address this challenge: the software product line (SPL) paradigm [9] and the model-driven engineering approach [10]. These two approaches are a good strategy to generate flexible and maintainable dashboards with different features.

On the other hand, current artificial intelligence (AI) methods can be leveraged and applied to the dashboards generation process. An AI model can be fed with users' characteristics [11] to offer the potentially best and most beneficial dashboard configurations for their contexts.

The rest of the paper is organized through the following sections: the second section introduces the hypotheses and objectives of this research project, followed by section 3, where the used methodology is detailed. Section 4 presents the current status of the research, finalizing with section 5, where the conclusions derived from this work are presented.

2 Hypotheses and objectives

The main hypothesis of this work is the following:

H1. The tailoring of user interfaces for supporting decision-making processes increment the efficiency and efficacy when extracting information and generating knowledge from the displayed data.

The goal of the research is to obtain a generative framework for the automatic and systematic development of information dashboards, where the tailoring task involves not only variability at the layout and visual design levels, but also at data and interaction mechanisms levels to foster individualization, usability, and flexibility. This goal also involves the application

of AI technology to provide beneficial dashboard configurations automatically.

A series of sub-objectives are posed to reach the mentioned main goal.

- Identify common characteristics of information dashboards at a meta-level (i.e., abstract level)
- Identify connection mechanisms to enable a model-driven approach to build concrete products of the SPL
- Implement mechanisms that foster interoperability to allow the connection of different data sources
- Define and implement reusable and configurable core assets to generate concrete products of the SPL
- Evaluate the SPL at a generative and functional level
- Evaluate the generated dashboards in terms of usability and tailoring capabilities
- Study the automatic adaptation of the dashboards depending on the users' characteristics an behavior through AI mechanisms
- Study the integration of the dashboards SPL within different technological ecosystems and case studies

The outcomes of the different phases of the research will support the test of posed hypothesis.

3 Methodology

3.1 Action-research methodology

This research project is approached as an iterative process where the knowledge gained through past experiences and the outcomes of the different cycles is crucial for the following stages. The Action-Research methodological framework [12] will be followed to accomplish this process.

Kemmis posed Action-Research [13] as an inquiry method carried out by the participants in social situations with the aim of improving and understanding their own social practices and their contexts.

Later, McTaggart & Kemmis described the characteristics of this methodology. The Action-Research methodology is based on a cyclic spiral of research and actions composed of a series of phases and sequences [14].

Therefore, Action-Research is an iterative process where each cycle provides an output that will be the input for the next cycle.

The iterative nature of the methodology enables the researcher to address previously identified problems, thus obtaining more refined solutions.

However, to be able to start the Action-Research cycles, it is necessary to formalize the problem to be addressed. Similar problems and previously developed solutions have been studied to understand the context and the current state of the field. The

methodology used for this step (a Systematic Literature Review) is detailed in the next section.

Once the problem is formalized, two Action-Research cycles are proposed to develop a proposal for generating dashboards and evaluate them in real contexts. Evaluation is necessary to obtain feedback to improve the proposal.

The chosen framework for software development is an agile approach based on SCRUM [15]. This framework provides the necessary processes, rules, practices, roles, and artifacts to increase the productivity of development teams through an iterative and incremental software development cycle [16].

A mixed methods research approach will be employed to evaluate the dashboards. The research will be conducted using both quantitative and qualitative methods [17], leveraging the two perspectives to obtain a wider view of the results to face the next Action-Research cycles.

3.2 Systematic literature review

As introduced above, a systematic literature review (SLR) is a powerful method to gain knowledge about previous solutions and similar problems. The SLR helps in the contextualization of the problem to be solved and provides new research lines by identifying weaknesses and strengths in previous solutions.

The SLR is conducted under the guidelines proposed by Kitchenham [18]. Following the [18, 19] guidelines, the SLR is composed of three main phases: planning, conducting, and reporting the study.

However, before planning the review, a preliminary search was performed to verify that no recent reviews about the target topic. If any recent SLR were found, there would not be any necessity to conduct a new one.

This preliminary search was performed using different electronic databases (Scopus, Web of Science (WoS), IEEE Xplorer and Springer) and using terms related to literature reviews ("SLR", "systematic literature review", etc.), as well as terms related to the target of the review ("dashboards").

The result of the previous search confirmed that, at the time of performing the queries, there were not any previous SLR about tailored dashboards, so the necessity of performing a literature review was justified.

3.3 Meta-modeling and domain engineering

Two methods are selected to tackle the design of tailored dashboards with automation possibilities.

Given the complexity of the dashboards' design processes, it is necessary to understand their domain deeply. Dashboards can present different features, different visual designs, different purposes, etc. However, dashboards also share common features that are always present.

These common features can be abstracted to obtain generic schemas or models that can help with the domain understanding and systematic reuse of software components. The technique for identifying shared properties and variabilities within a specific domain is called domain engineering [20].

Domain engineering is based on knowledge reuse regarding some specific domain. This approach is an essential phase of the software product line (SPL) paradigm [21, 22]. This methodology allows the reuse of software components and their configuration to match certain requirements; that is why identifying common features and variabilities is an essential step.

Once the domain has been studied, it is possible to develop a generic model (a meta-model) that captures every abstract property of dashboards, as well as the relationships among the identified entities.

Meta-models are crucial artifacts in model-driven paradigms [10, 23, 24], as they allow mapping entities from high-abstraction levels to more concrete entities and even code through transformations.

These two related methodologies increase not only productivity regarding software development, but also knowledge reuse, and are suitable methods to address several requirements from different profiles and contexts.

4 Dissertation status

The presented research project is currently in a conceptualization stage.

As previously introduced, the dashboards domain is a complex field of study, as several elements and disciplines are involved. That is why the contextualization and research of previous solutions are crucial for the development of the thesis.

A systematic literature review (SLR) has been performed to gain knowledge about this domain, as detailed before. The SLR is focused on how existing approaches and solutions have tackled tailoring capabilities of dashboards.

During the planning phase, the scope of the review was defined: research questions, inclusion and exclusion criteria, search strategy, query strings, and quality criteria. The research questions that the systematic review aims to answer are the following:

- **RQ1.** How have existing dashboard solutions tackled the necessity of tailoring capabilities?
- **RQ2.** Which methods have been applied to support tailoring capabilities within the dashboards' domain?
- **RQ3.** How the proposed solutions manage the dashboard's requirements?
- **RQ4.** Can the proposed solutions be transferred to different domains?
- **RQ5.** Has any artificial intelligence approach been applied to the dashboards' tailoring processes and, if applicable, how these approaches have been involved in the dashboards' tailoring processes?
- **RQ6.** How mature are tailored dashboards regarding their evaluation?

The data extraction process to conduct the present SLR has been divided into different phases in which various activities are performed. The PRISMA statement [25] has been used to detail the performed tasks during the whole review process, and it can be consulted in Figure 1.

In the end, 23 papers about tailoring capabilities within the dashboards domain were selected, and the different research questions were answered through them, obtaining a wide-view of how this tailoring challenge has been addressed before.

The SLR [26] was complemented with a systematic mapping of the literature that can be consulted at [27]. The gained knowledge about technical features (RQ1-RQ4), AI applications (RQ5), and evaluation (RQ6) will be applied to improve tailoring capabilities. Especially, the answer to RQ5 provided clarity on the lack of AI applications on this domain, which could be potentially useful to automate the configuration processes of dashboards to match concrete user requirements and contexts.

A dashboard meta-model has been developed using the gained knowledge and employing the methodologies described in section 3.3. The meta-model is composed of three main sections: user, layout, and components.
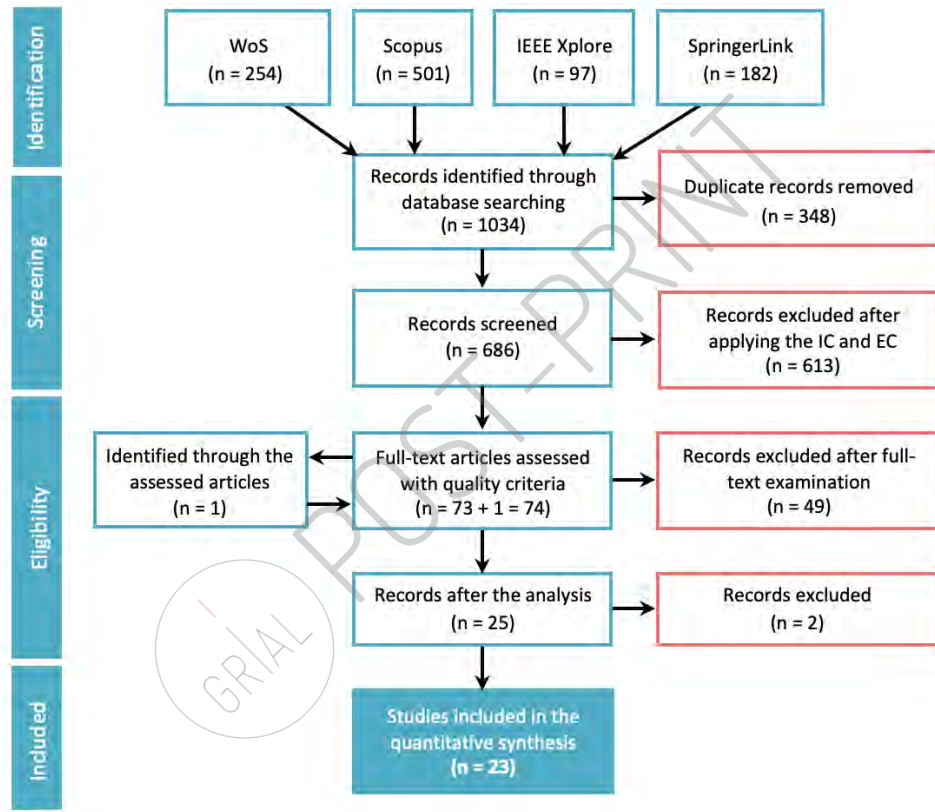

Figure 1: PRISMA flow. Adapted from [25]. Elaborated by the authors [27].

Each section is related to each other, and the whole meta-model describes a high-level view of dashboards with their common features and properties.

The user and layout meta-models can be consulted at [28, 29]. The next steps will involve the refinement (addition of constraints, rules, etc.) and instantiation of this meta-model to obtain concrete models and mapping them to real code through AI paradigms.

5 Conclusions

The proposed research project is focused on developing tailored solutions to support decision-making processes. This work outlines the methodologies and current status of the dissertation.

A systematic literature review has been performed to gain knowledge about the domain, and a dashboard meta-model has been developed to capture high-level features and properties.

The next steps will employ the meta-model to instantiate concrete dashboard models to generate their code subsequently.

The automatic generation of dashboards could improve the effectiveness of these tools by adapting them to specific users' needs without consuming significant quantities of resources and time.

ACKNOWLEDGMENTS

REFERENCES

[1] S Christian Albright, Wayne Winston, and Christopher Zappe. 2010. Data analysis and decision making. Cengage Learning, Mason, OH, USA.

[2] Dj Patil and Hilary Mason. 2015. Data Driven. " O'Reilly Media, Inc.".

[3] Ramesh Sharda, Dursun Delen, and Efraim Turban. 2013. Business intelligence: a managerial perspective on analytics. Prentice Hall Press.

[4] Martin J Eppler. 2007. Knowledge communication problems between experts and decision makers: An overview and classification. Electronic Journal of Knowledge Management 5, 3.

[5] Alper Sarikaya, Michael Correll, Lyn Bartram, Melanie Tory, and Danyel Fisher. 2018. What Do We Talk About When We Talk About Dashboards? IEEE Transactions on Visualization Computer Graphics 25, 1, 682 - 692.

[6] Stephen Few. 2006. Information dashboard design.

[7] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. 2008. Visual analytics: Definition, process, and challenges. In Information Visualization Springer, 154-175.

[8] Edward Tufte and P Graves-Morris. 2014. The visual display of quantitative information.; 1983 Graphics Press, Cheshire, CT, USA.

[9] Klaus Pohl, Gunter Böckle, and Frank J. Van Der Linden. 2005. Software Product Line Engineering: Foundations, Principles and Techniques. Springer-Verlag New York, Inc., New York, NY, USA.

[10] Anneke G Kleppe, Jos Warmer, and Wim Bast. 2003. MDA Explained. The Model Driven Architecture: Practice and Promise Addison-Wesley Longman Publishing Co., Inc., Boston, MA.

[11] J. Cruz-Benito, A. Vázquez-Ingelmo, J. C. Sánchez-Prieto, R. Therón, F. J. García-Peñalvo, and M. Martín-González. 2018. Enabling adaptability in web forms based on user characteristics detection through A/B testing and machine learning. IEEE Access 6, 2251-2265. DOI:http://dx.doi.org/10.1109/ACCESS.2017.2782678.

[12] Kurt Lewin. 1946. Action research and minority problems. Journal of social issues 2, 4, 34-46.

[13] Stephen Kemmis. 1984. Point-by-point guide to action research. Victoria: Deakin University.

[14] Robin Mctaggart and Stephen Kemmis. 1988. The action research planner. Deakin university.

[15] Ken Schwaber. 1997. Scrum development process. In Business object design and implementation Springer, 117-134.

[16] Ken Schwaber and Mike Beedle. 2002. Agile software development with Scrum. Prentice Hall Upper Saddle River.

[17] John W Creswell and J David Creswell. 2017. Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

[18] B. Kitchenham and S. Charters. 2007. Guidelines for performing Systematic Literature Reviews in Software Engineering. Version 2.3. Technical Report Report. School of Computer Science and Mathematics, Keele University.

[19] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. Keele, UK, Keele University 33, 2004, 1-26.

[20] Maarit Harsu. 2002. A survey on domain engineering. Citeseer.

[21] Markus Voelter and Eelco Visser. 2011. Product line engineering using domain-specific languages. In Software Product Line Conference (SPLC), 2011 15th International IEEE, 70-79.

[22] Paul Clements and Linda Northrop. 2002. Software product lines. Addison-Wesley, Boston, MA, USA.

[23] Nicolas Anquetil, Birgit Grammel, Ismênia Galvão, Joost Noppen, Safoora Shakil Khan, Hugo Arboleda, Awais Rashid, and Alessandro Garcia. 2008. Traceability for model driven, software product line engineering. In ECMDA Traceability Workshop Proceedings SINTEF, 77-86.

[24] Salvador Trujillo, Don Batory, and Oscar Diaz. 2007. Feature oriented model driven development: A case study for portlets. In Proceedings of the 29th international conference on Software Engineering IEEE Computer Society, 44-53.

[25] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and The Prisma Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine 6, 7, e1000097.

[26] A. Vázquez-Ingelmo, F. J. Garcia-Peñalvo, and R. Therón. 2019. Information Dashboards and Tailoring Capabilities - A Systematic Literature Review. IEEE Access 7, 109673-109688. DOI:http://dx.doi.org/10.1109/ACCESS.2019.2933472.

[27] Andrea Vázquez Ingelmo, Francisco J. García-Peñalvo, and Roberto Therón. 2019. Tailored information dashboards: A systematic mapping of the literature. In Proceedings of the Interacción2019 (Donostia, Spain2019). ACM.

[28] Andrea Vázquez Ingelmo, Francisco J. García-Peñalvo, Roberto Therón, and Miguel Ángel Conde González. 2019. Extending a dashboard meta-model to account for users' characteristics and goals for enhancing personalization. In Proceedings of the Learning Analytics Summer Institute (LASI) Spain 2019 (Vigo, Spain2019).

[29] Andrea Vázquez-Ingelmo, Francisco J. García-Peñalvo, and Roberto Therón. 2019. Taking advantage of the software product line paradigm to generate customized user interfaces for decision-making processes: a case study on university employability. PeerJ Computer Science 5(2019/07/01), e203. DOI:http://dx.doi.org/10.7717/peerj-cs.203.