A User Profiling Approach to Optimize the Production of Architectural Documentation

Matias Nicoletti

ISISTAN Research Institute, CONICET-UNICEN. Tandil, Argentina.

The Software Architecture Document (SAD) is a key artifact in the early stages of software development, as it serves to share the architectural knowledge among the key project stakeholders and to keep record of the main design decisions for satisfying the stakeholders' concerns. Even the best architectural solution could fail its purpose if it is not properly documented, distributed and understood by the stakeholders. However, documenting an architecture is a non-trivial and time-consuming activity, usually performed with limited resources. Thus, documenting a SAD should be planned to ensure that the invested efforts are worth the cost in terms of SAD quality.

An approach to plan and generate useful architectural documentation with limited resources is to produce *reader-oriented documentation.* We claim that documentation should be written from readers' perspective, rather than from writers' perspective. Stakeholders' interests should be assessed before producing the documentation to ensure it addresses stakeholders' information needs. However, identifying stakeholders' preferences is challenging, since explicit information about their interests is not usually available. In such cases, stakeholders' interests can be discovered automatically from their working contexts by means of User Profiling techniques.

Very few of the existing approaches to architectural documentation follow a stakeholder-centric documentation strategy. A relevant work is the View & Beyond approach [1], which proposes a role-based personalization of the SAD contents in the form of a matrix that links stakeholders' roles with SAD sections in terms of level of interest. A common drawback of current approaches is that they provide few or no guidelines to the documenter on how to generate the right SAD contents. Another issue is that stakeholders' interests are considered as static (regarding time) and are only derived from their roles. In practice, interests vary during a software project life-time and might depend on other factors in addition to the role, such us the stakeholder's reading history. This sort of simplified models of interests might be inaccurate and fail its purpose.

We propose an approach to assist the documenter in generating useful and low-cost architectural documentation delivered incrementally. To this end, we use two kinds of techniques: User Profiling and Optimization. As regard the personalization aspect, we base our approach on the construction of user (stakeholder) profiles, using Natural Language Processing (NLP) and User Profiling (UP) techniques [4]. In this work, a user profile comprises two parts: (i) *a priori* information about the user (e.g., the stakeholder's role and priority, predefined views for each role) and (ii) information extracted from the user's work environment (e.g., frequent terms mentioned or SAD sections frequently accessed). We claim that user profiling techniques can produce accurate models for stakeholders' interests.

On a second stage, the user profiles serve as the input of an optimization algorithm to maximize benefits of the SAD generation. In other words, we seek for the mini-

mal amount of documentation (or *documentation plan*) that maximizes the quality of the resulting SAD version with the minimum effort. In our context, the quality of the architectural documentation is directly related to the level of satisfaction of the key (high-priority) stakeholders. The final goal is to suggest the documenter a documentation plan to produce cost-effective SAD versions. We refer to this problem as the *Next SAD Version Problem* (NSVP) and tackle it with heuristic algorithms due to its time complexity (NP-hard).

The approach is supported by two software tools. On one hand, we have a Wiki-based environment with monitoring capabilities for stakeholders' actions and additional collaborative features, such as, chat and forums. On the other hand, there is an administrative control panel that provides several functionalities to support the documenter's tasks.

This thesis is currently in an advanced stage of completeness. As regards the user profiling aspect, we have already implemented a semi-automated algorithm to infer stakeholders' interests and also validated it in two experiments. The results obtained from these experiences were encouraging in the sense that we were able to predict stakeholders' interests with an acceptable precision [3]. Regarding the optimization aspect, we have formalized the NSVP as a mono-objective optimization problem. Currently, we are exploring different techniques to solve the NSVP, including exact techniques (e.g., Pseudo-Boolean SAT4J) and heuristic ones (e.g., NSGA-II). Also, preliminary evaluations have demonstrated the potential of using this sort of techniques to solve the NSVP [2]. The remaining work of this thesis is focused on the evaluation of the whole approach in a development scenario (as real as possible) in which a group of stakeholders with competing interests should accomplish a set of tasks using a SAD, which is released in several versions. The goal of this evaluation is to check whether an smaller (optimized) SAD can replace a standard SAD without losing its level of quality. The satisfaction level will be measured using subjects feedback.

References

1. P. Clements, F. Bachmann, L. Bass, D. Garlan, J. Ivers, R. Little, R. Nord, and J. Stafford. A practical method for documenting software architectures. In *Proc. of the 25th ICSE*, 2003.
2. J. A. Diaz-Pace, M. Nicoletti, S. Schiaffino, C. Villavicencio, and E. Sánchez. A stakeholder-centric optimization strategy for architectural documentation. In *Proc. of 3th MEDI*, 2013.
3. M. Nicoletti, J. A. Diaz-Pace, and S. Schiaffino. Discovering stakeholders' interests in wiki-based architectural documentation. In *Proc. of 16th CIbSE (IDEAS)*, 2013.
4. M. Nicoletti, S. Schiaffino, and D. Godoy. Mining interests for user profiling in electronic conversations. *Expert Systems with Applications*, 40:638 – 645, 2013.