Reducing Architectural Knowledge Vaporization by Applying the Repertory Grid Technique

Dan Tofan, Matthias Galster, and Paris Avgeriou

Department of Mathematics and Computing Science, University of Groningen
The Netherlands
{d.c.tofan,m.r.galster}@rug.nl, paris@cs.rug.nl

Abstract. The architecture of a software-intensive system is the composition of architectural design decisions. These decisions are an important part of Architectural Knowledge (AK). Failure to document architectural design decisions can lead to AK vaporization and higher maintenance costs. To reduce AK vaporization, we propose to apply the Repertory Grid Technique (RGT) to make tacit knowledge about architecture decisions explicit. An architect can use the RGT to elicit decision alternatives and concerns, and to rank each alternative against concerns. To validate our approach, we conducted a survey with graduate students. In the survey, participants documented decisions using the RGT. We compared these decisions with decisions documented using a basic decision template. Our results suggest that RGT leads to less AK vaporization, compared to conventional ways of documenting decisions.

Keywords: architectural knowledge, repertory grid, AK vaporization, survey.

1 Introduction

Bosch [2] considers the software architecture of a system as the composition of a set of architectural decisions. Kruchten et al. [9] define Architectural Knowledge (AK) about a system as design decisions and design. De Boer and Farenhorst [1] investigated various definitions of AK in literature. They obtained evidence for the importance of decisions in AK and argued that decisions represent a link from the problem domain to the solution domain.

Bosch [2] considers AK vaporization as a key problem in software architecture. AK vaporization contributes to expensive system evolution, lack of stakeholders' communication and reduced reusability [7]. Given the importance of AK vaporization, the architecting community has proposed various approaches for reducing it. For example, Bosch [2] argued for the representation of design decisions as first class entities in software architecture. Kruchten et al. [9] discussed an AK repository, and its underlying ontology. Jansen [7] studied decisions recovery and their dependencies. Kruchten et al. [8] described a decision view that incorporates design decisions in the '4+1' view model. Furthermore, the ISO/IEC 42010 standard provides recommendations for incorporating design decisions in architecture descriptions.

I. Crnkovic, V. Gruhn, and M. Book (Eds.): ECSA 2011, LNCS 6903, pp. 244–251, 2011.

In this paper we propose to use the Repertory Grid Technique (RGT) for capturing architectural decisions. In general, the RGT follows four steps: choose a decision topic, obtain alternatives, get concerns, and rate alternatives against each concern. The resulting grids can be used for hierarchical cluster analysis of alternatives and concerns. For details on the RGT please refer to [4]. In a previous study [10], we already identified some advantages and disadvantages of RGT. In this paper, we investigate how RGT may reduce AK vaporization.

2 Applying the RG Technique

We conducted a descriptive survey with the goal of investigating the reduction of AK vaporization by applying RGT. As sample, we invited graduate students enrolled in a Software Architecture course at the University of Groningen in 2010. Participation was voluntary and had no influence on grades. The study was scheduled to take place as part of an optional 2-hour seminar.

In the session, we first trained the students to use the RGT and then asked them to apply it to some architectural decisions. Participants used the RGT to describe decisions that occurred in the context of the course project. The course project required students to act as architects, and design a complex home automation system that interacts with the Smart Grid in order to sell and buy electric power. Throughout the semester, students worked in groups of five persons to architect the system. As our study took place halfway through the project, students had already made the important architectural decisions and thus possessed AK about the home automation system. Therefore, we captured the AK of students about some of the decision topics that occurred in their course project. Even though students worked as teams, we decided to apply the RGT at an individual level, for simplicity reasons.

To structure our study, we used Basili's Goal-Question-Metric (GQM) method. Our *goal* is to reduce AK vaporization from architects' viewpoint. This is important to both practitioners and researchers, because knowledge vaporization leads to increased maintenance costs. Our *question* is: Does the RG technique reduce AK vaporization, compared to a basic approach to document architectural decisions?

Selecting *metrics* for AK vaporization was more difficult. In our context, the decisions documented in architectural documents in the course project were based on a predefined template. Given their lack of experience, students were asked to describe the decision topic, alternatives, their pros and cons, decision outcome, and rationale. Moreover, they did not use any established approach for documenting decisions. We speculate that architectural documentations created by students are similar to documentations produced by practitioners who do not use systematic approaches for decision documentation. From our experience, we consider that decision documentation in industry is rarely systematic, even compared to the templates filled by students. On the other hand, the RG output is the result of a systematic, but unproven approach.

Existing work on AK vaporization has yet to offer techniques for measuring vaporization and to facilitate the evaluation of new approaches that claim to reduce AK vaporization. Thus, we defined the following three *metrics* to compare how RG-based decisions differ from decisions based on a basic decision template. First, we considered the number of explicit decision alternatives, followed by the number of concerns. The third metric was the ratio of the number of expressed rankings compared to the maximum of possible rankings. These metrics could be obtained from the RGT output as well as from the basic architectural documentations created by students. A higher number of explicit alternatives, concerns or rankings may suggest a reduction in AK vaporization. To increase reliability, two researchers conducted this analysis. For the ratio of rankings, we divided the number of explicit rankings by the maximum number of possible rankings. For example, if a decision has 4 explicit rankings, 2 alternatives and 3 concerns, then the ratio is 0.66.

Next, we describe the preparations for our data collection efforts. Due to time constraints for the study, we needed a RG tool that participants could use for the AK acquisition task. After piloting a few tools, we selected Idiogrid [6], because it supports self-administering the RGT. Next, we describe the four steps of RGT as performed in the study.

Choose Decision Topic. We prepared some decision topics for students to use in the survey. To apply RGT, participants needed to be experts on the decision topic, or be at least knowledgeable about the topic. To satisfy this condition, we analyzed the course project reports, delivered by students. For each of the six groups, we compiled a list of decision topics and alternatives. Next, we analyzed which decision topics appeared across all groups, to see which topics are more common. We identified four such topics: choice of user interface, programming language, communication technology, and operating system. To satisfy the expertise prerequisite, we asked each student to choose two out of the four decision topics, based on his/her familiarity.

Get Alternatives. According to Edwards et al. [3], the alternatives (or elements) used in the RGT can either be supplied to the participant or elicited from him/her. The former is suitable for investigations on a specific set of elements [3]. As we aimed to elicit AK from participants, we decided to ask participants to specify the alternatives for each selected topic.

Get Concerns. Similar to the previous step, we decided to elicit characteristics (or constructs) from participants, rather than to supply them. However, in our previous study [10], we used the triadic elicitation approach in individual interviews: we asked the expert in what way two of three alternatives (elements) are alike, but different from the third. For this study, we doubted that most students could successfully use the triadic approach in a self-administered session, through a dialog with a tool, instead of an interviewer. According to Grice et al. [5], grids based on sentence completion are suitable for any domain of experience, and are easy to complete. Following our pilot sessions, we concluded that it may be more intuitive for students to generate constructs through sentence completions, compared to the triadic elicitations. Therefore, we decided to use the sentence completion approach.

Rate Alternatives. For this step, we configured Idiogrid to use a 5-point rating scale, ranging from -2 to 2. When rating an alternative against a concern, lower values indicated agreement with the concern (i.e., affordable), while higher values indicated agreement with its opposite (i.e., expensive). The middle value indicated neutrality, uncertainty or lack of applicability.

When executing the survey, we first introduced the participants to the RG technique. Afterwards, we asked the participants to do an example grid session, for training purposes, with the topic of choosing between bars in town. Next, participants applied RGT, on decision topics of their choice. We asked the participants not to use the internet, or talk with each other during the study. We did not impose a time limit for the RG sessions. Due to the limited duration of the session, we skipped grids analysis and refinement. Instead, we decided to send each student an email with an analysis of his/her grids. Furthermore, two researchers were available to answer questions from participants during the study.

At the end, each participant filled in a questionnaire. We were interested in the profile of the participants. Moreover, we added questions on the study itself, e.g., to check whether the participants understood the instructions and questions, and the perceived difficulty of performing the RGT for the decision topics.

4 Analysis of Survey Results

Each of the 20 participants delivered two grids, except for two persons who had to leave earlier and produced only one grid. One participant delivered no grid at all. Overall, we obtained 36 grids, as well as paper-based post questionnaires from attendants. Additionally, students delivered 6 architectural documents, at the end of the course. As described in the survey design, we collected measurements for both, the grids and architectural documents.

To measure the number of explicit concerns, two authors of this paper conducted a content analysis on each decision description. Each researcher individually assigned a concern to every sentence of a decision's description. Next, we reviewed every sentence, and compared the assigned concerns. We considered an agreement if both researchers meant the same thing, but used different words. For example, if one assigned to a fragment the concern *cost* and the other one assigned *affordability,* then it is an agreement. Upon disagreements, we either agreed to use an existing concern from one of us, or we negotiated the assignment of a new concern. For a few sentences, we asked a third researcher to mediate. Initial inter-rater agreement was 51.8%. After negotiations, we achieved full agreement.

For the third metric, we needed to determine the ratio of explicit rankings, against the maximum number of possible rankings. We multiplied the number of explicit alternatives and the number of distinct concerns to obtain the maximum number of possible rankings.

4.2 Analyzing Metrics for All Decisions

Some students from the same course project group used the RGT on identical decision topics. We obtained 12 grids for which only one student from a course project group addressed that decision topic (single grids). Additionally, we obtained 7 double grids: two participants from the same group produced individually two grids on the same decision. Also, we got 2 triple grids, from three members of the same group capturing the same decision. Similarly, we obtained 1 quadruple grid, from four members of the same group.

To ensure a suitable comparison of the metrics obtained from the grids and the project reports, the grid must have been produced by a student who also co-authored the architectural document. Our raw data consisted of 12 data points of single grids, 7 data points of double grids, 2 data points of triple grids, and 1 data point of quadruple grids.

To analyze the data, we needed to filter outliers. We noticed that one data point of double grids and the one with quadruple grids had poor decisions descriptions in the architectural document, i.e., no alternatives considered. Therefore, we eliminated the two data points. In one of the triple grids, we eliminated a grid due to poorly phrased concerns and obtained a new data point with double grids. Similarly, we converted the other data point with triple grids into a new one with a single grid, by removing two poor quality grids. After filtering, we obtained 13 single-grid data points, and 7 double-grid data points. The numbers of alternatives and concerns in a double-grid data point were calculated by counting the distinct alternatives and concerns from the two grids.

The boxplots in figure 1 summarize the collected metrics. The median of the numbers of alternatives obtained with RGT is 4 for single-grid data points, and 6 for double-grid data points. Half of the data points for numbers of concerns in student reports (architectural documents), for single-grid data points, were equal to 5, 6, or 7. All ratios of rankings for grids are equal to 1.

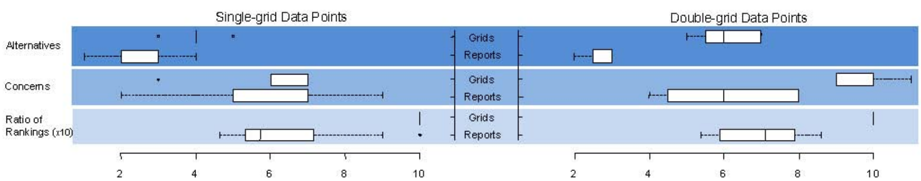

Fig. 1. Boxplots for each metric, for the two types of data points

To answer our research question, we define the following null hypotheses.

H_{0a}: The RGT does not influence the number of explicit alternatives.

H_{0b}: The RGT does not influence the number of explicit concerns.

H_{0c}: The RGT does not influence the ratio of explicit rankings.

The alternative hypotheses are the following.

H_{1a}: The RGT influences the number of explicit alternatives.

H_{1b}: The RGT influences the number of explicit concerns.

H_{1c}: The RGT influences the ratio of explicit rankings.

Next, we compare the means of each metric, for the decisions in grids and architectural documents, by using the Wilcoxon signed ranks test. We test each hypothesis on the single-grid data points and the double-grid ones. Table 1 summarizes the results, for the numbers (#) of alternatives, concerns, and ration, of grids (G) and reports (R).

Table 1. Hypotheses, metrics, means, standard deviations, and p-values for the two samples

H	Metric	13 Single-grid Data Points			7 Double–grid Data Points		
		Mean	Std. Dev.	*p*-value	Mean	Std. Dev.	*p*-value
H_a	# Alternatives G	4.00	0.41	0.002	6.14	0.90	0.016
	# Alternatives R	2.62	0.77		2.71	0.49	
H_b	# Concerns G	6.00	1.41	0.720	9.57	0.79	0.017
	# Concerns R	6.23	1.88		6.14	1.86	
H_c	Ratio G	1.00	0.00	0.003	1.00	0.00	0.018
	Ratio R	0.66	0.19		0.70	0.13	

Given the p-values less than 0.05 and the means values, we can reject H_{0a}, and accept that applying the RGT increased the number of explicit alternatives, for both single- and double-grid data points. Regarding H_{0b}, we cannot find any influence of RGT, for single grids, due to the high p-value. However, we can reject H_{0b} for the sample with double grids. Additionally, we reject the third null hypothesis (H_{0c}), as the p-values are low.

The results strongly suggest that one grid contains more explicit alternatives and higher ratio of explicit rankings than the equivalent description from a basic architectural document. Two grids seem to contain not only more explicit alternatives and rankings, but also more concerns.

4.2 Post Questionnaires

From post questionnaires, we learned that participants had a bachelor degree in Computer Science or a related field (e.g., Information Technology). Half of the participants had an average of around 2 years of work experience in software industry. Students needed an average of 24 minutes for each grid session, with a standard deviation of 11 minutes.

We asked students to rate some statements on a Likert scale from 0 to 4 (strongly disagree, disagree, neutral, agree, strongly agree) that referred to their understanding of the study. We learned that participants understood the presentation on the RG technique. They also understood the directions for creating the grids, and perceived the Idiogrid tool [6] as easy to use. Students perceived the first grid as easier to do, than the second one. A possible explanation for the difference may be that students applied RGT on the most familiar topic first, followed by the less familiar one. Participants indicated that they clearly understood the tasks. Also, they partly enjoyed the assignment.

Overall, post questionnaire results suggest that participants did not face significant issues in using RGT in a self-administrated manner, after a short introduction to it. We believe that the smooth learning curve and low time cost may facilitate RGT adoption by practitioners.

5 Discussion

RGT tends to elicit a higher number of alternatives, compared to basic architectural documents, which usually mention 2-3 alternatives. Additionally, RGT delivers 100%

of possible rankings for a decision, due to the systematic steps for capturing decisions. In contrast, architectural documents seem to make explicit only around 70% of rankings, as participants used no systematic technique for capturing decisions.

On average, one grid contained around 6 concerns, similar to decisions from the architectural documents. However, double grids contained around 9 concerns. Combining RGT-elicited concerns by 2 out of the 5 members of a team increased significantly the number of explicit concerns. Therefore, we believe RGT may be useful for architectural reviews, to help uncover more concerns from stakeholders.

We found out that participants spent an average of 24 minutes to capture a decision with RGT. In our previous study [10], participants captured a decision in 57 minutes, on average. We believe the difference is mainly due to the approach for eliciting concerns. As reported by Grice et al. [5], we also noticed that sentence completion seems to be more user friendly than triadic elicitation. However, we speculate that the triadic elicitation approach asks the expert to reflect more, with the potential to unearth more in-depth tacit AK than the sentence completion approach. We consider both approaches are valid and useful, as they have complementing qualities.

Regarding validity threats, Edwards et al. [3] offer criteria for evaluating a study that uses RGT, such as supplied vs. elicited elements or concerns. Our study used full individual repertory grids, as both elements (alternatives) and constructs (concerns) were elicited from each participant. This is especially well-suited for exploratory situations, like capturing tacit knowledge [3]. However, we decided to use a less known approach for construct elicitation, although more user-friendly: sentence completion [5]. To address the risks of using a less-established approach, we piloted it before the study, to make sure that sentence completion provides useful outputs.

Regarding the external validity of our study, we consider the study participants as representative for inexperienced software architects. Half of the graduated students had around 2 years of working experience. However, we cannot generalize our results to experienced architects. Moreover, we do not know if the decision descriptions in the architectural documents from students are representative for the industry. Concerning internal validity, the main issue is the history of the decision descriptions from the architectural documents. Students thought about the decisions, and created their descriptions as team work, while grids as individual work. We partly addressed this risk, by dividing the grids based on the number of students in the same group, who worked on the same decision topic. Additionally, the structure of the basic decision template influences the resulting decision descriptions. Different templates may result in different metrics and different information. Therefore we cannot claim that RGT provides better results when compared to any type of decision documentation. However, students repeatedly refined their decisions' descriptions in the architectural documents, as part of the course. In contrast, the students did not have time to refine their grids.

6 Conclusions and Future Work

The goal of our study was reducing AK vaporization. To achieve it, we used RGT on some architectural decisions. We also investigated how RGT compares to a basic approach to documenting architectural decisions. Specifically, we analyzed metrics on important parts of a decision: alternatives, concerns, and rankings. Also, content analysis of architectural documentation provided measurements for descriptions of decisions. We learned that grids seem to capture more alternatives and more rankings, compared to architectural documents.

We consider that although tool support for RGT exists, it is not geared towards capturing AK, leaving room for future improvements. Moreover, we need to refine metrics for AK vaporization, to help evaluating approaches for reducing it. Additionally, we plan to investigate the reuse of captured AK through RGT in order to obtain a good return on investment for the spent effort.

Acknowledgments. This research has been partially sponsored by NWO SaS-LeG, contract no. 638.000.000.07N07. We thank study participants, David Ameller, Uwe van Heesch and Pavel Bulanov for their help.

References

1. de Boer, R.C., Farenhorst, R.: In Search of 'Architectural Knowledge. In: Proc. Third Workshop on SHAring and Reusing architectural Knowledge Architecture, Rationale, and Design Intent, pp. 71–78 (2008)
2. Bosch, J.: Software architecture: The next step. In: Brauer, W. (ed.) ICALP 1985. LNCS, vol. 194, pp. 194–199. Springer, Heidelberg (1985)
3. Edwards, H.M., McDonald, S., Young, S.M.: The repertory grid technique: Its place in empirical software engineering research. Information and Software Technology 51, 785–798 (2009)
4. Fransella, F., Bell, R., Bannister, D.: A Manual for Repertory Grid Technique, 2nd edn. Wiley, London (2004)
5. Grice, J., Burkley, E., Burkley, M., Wright, S.: J: A sentence completion task for eliciting personal constructs in specific domains. Personal Construct Theory & Practice 1, 60–75 (2004)
6. Grice, J.W.: Idiogrid: software for the management and analysis of repertory grids. Behavior Research Methods 34, 338–341 (2002)
7. Jansen, A.: Architectural design decisions, PhD thesis, University of Groningen, Netherlands (2008)
8. Kruchten, P., Capilla, R., Dueñas, J.C.: The decision view's role in software architecture practice. IEEE Software 26, 36–42 (2009)
9. Kruchten, P., Lago, P., van Vliet, H.: Building up and reasoning about architectural knowledge. Quality of Software Architectures, 43–58 (2006)
10. Tofan, D., Galster, M., Avgeriou, P.: Capturing Tacit Architectural Knowledge Using the Repertory Grid Technique (NIER Track). In: Proceedings of the 33rd International Conference on Software Engineering, Honolulu, USA, pp. 916–919 (2011)