Enhancing Provenance Representation with Knowledge Based on NFR Conceptual Modeling: A Softgoal Catalog Approach

Sérgio Manuel Serra da Cruz[1,2,3(✉)] and André Luiz de Castro Leal[1]

[1] UFRRJ – Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, Brazil
[2] PPGMMC/UFRRJ – Programa de Pós Graduação em Modelagem Matemática e Computacional, Seropédica, RJ, Brazil
[3] PET-SI/UFRRJ – Programa de Educação Tutorial - Sistemas de Informação, Seropédica, RJ, Brazil
andrecastr@gmail.com, serra@ufrrj.br

Abstract. This work explores the organization of the provenance as a catalog of non-functional requirement (NFR). It considers provenance as a quality factor that should be incorporated since the early stages of software development as softgoals. The aim of this research is to introduce a systematic approach to design a provenance catalog using consolidated software engineering techniques. The study is an effort to depict provenance as patterns supported by Softgoal Interdependency Graphs (SIG) and Goal-Question-Operationalization method (GQO), a reusable framework that makes explicit characterization, decomposition, relationships and operationalization of elements that can be satisfied during the software design.

Keywords: Provenance · Non-functional requirements · Softgoal catalog · NFR patterns

1 Introduction

In Software Engineering (SE), one kind of requirement is called non-functional requirement. NFR is difficult to capture, organize, reuse and test; therefore, they are usually evaluated subjectively. NFR are known as constraints or quality requirements [1, 2] and are treated as *softgoals* [3]; they are targets that do not need to be addressed in an absolutely way but in a good enough sense [6]. The systematic treatment for NFR in early stages of software development may introduce positive contributions and increase software quality. The conceptual modeling for quality considering *provenance* as NFR is still underexplored either in the SE or Database domains. This is important because the quality achieved by data provenance has a clear proximity with software traceability. Both subjects are considered hot topics, offering potential benefits to data management and software development respectively.

Traceability and *provenance* handling consists of storing metadata that enables to reconstruct these chains of operations at different levels of abstraction. Due to the similarities between traceability and provenance [12], we advocate that the provenance

can also be considered as NFR in software development. There several representations of provenance focused on data [4, 7–9] and very few works of provenance focused on the software process [12, 13]. Data provenance authors use taxonomies, recommendations or ontologies to describe the elements involved in the conceptualization, classification and hierarchical structure of distinct kinds of provenance metadata. However, our research, differently from related works propose a new approach based on reusable catalogs (conceptual models) not only to represent provenance as a quality factor, but also to aid reducing the gap between software specification, its operationalizations and the diversity of data provenance descriptors generated by its execution.

The aim of this work is to present the steps to map provenance as NFR catalogs, using a systematic approach based on NFR framework [5], NFR patterns [6] and NRF catalogs [5, 10]. The NFR framework and the NFR patterns provide a solid theoretical foundation for treating NFR, with appropriate representation schemas and rules. In particular, the NFR pattern focuses on the reuse of NFR knowledge [3, 5]. NFR patterns may be decomposed to create/compose more precise and unambiguous patterns to build larger ones or be instantiated to create occurrence patterns using existing ones as templates.

2   Modeling Provenance as a NFR Catalog

Our proposal is one of the first to represent provenance as a quality factor within a catalog based on NFR framework and NFR patterns. We stress that the modeling effort is not a simple representation based on hierarchies of provenance or data provenance standards. Just the contrary, The NFR catalog was modeled taking into account the decomposition of softgoals to be addressed or achieved by (business or scientific) systems that require different kinds of provenance. Besides, our contribution also exposes the links and impacts between the software softgoals. We introduce a novel perception of provenance, describing it as a quality that must be satisfied to enhance the software traceability, enabling the construction of verifiable chains of operations in software systems to produce pieces of data with higher quality and embedded with data provenance descriptors.

The development of a Provenance NFR Catalog used several patterns defined by Supakkul et al. [6]: (i) *Objective Patterns* used to capture the definition of NFRs in terms of specific (soft)goals to be achieved; (ii) *Problem Patterns* captured knowledge of problems or obstacles to achieve goals; (iii) *Alternatives Patterns* (operationalizations) used to capture different means, solutions, and requirements mappings; (iv) *Selection Patterns* used to choose the best alternative considering their side-effects. To elaborate the provenance NFR Catalog we defined a set of three modeling steps.

First Step - The conceptual model was conceived to follow the Objective Pattern. The result is a Provenance SIG (not depicted here due to space restrictions). An SIG is a graph that shows two elements of Objective Patterns. First element is the Identification Pattern, where Provenance is modeled as the root of the graph. The second element is the Decomposition Pattern with relations, like 'Capturable', 'Classifiable' were presented. Such relations were based on the provenance taxonomy proposed by Cruz et al. [7].

The Provenance SIG was focused on the positive or negative contribution of the relations represented by links of the type HELP, HURT, BREAK and MAKE and also decompositions, operationalizations and argumentations represented by the links OR/ AND [5].

Second Step - In this step we defined three patterns: GrupoIdentification, Questions and Alternatives. The definition of such categories is important because they help designers to define the questions and further select the operationalizations during the software development process. After these definitions, it was possible specify the QuestionIdentification [10] and combine them with the GroupIdentification. The questions were answered according to the list of operationalizations for the softgoals (Alternative Patterns). Their impact on other NFR softgoals (previously defined in the SIG graph) were evaluated and then linked with questions as alternative responses. The operationalizations were represented at the lowest level of the SIG graph as leafs associated with NFR softgoals by contribution links of the type ANSWER.

Third Step - After defining the above mentioned patterns; it was possible to use SE standardized document like GQO [10] to organize and represent the knowledge achieved by the previous steps. The result of such effort was a conceptual model with the knowledge about provenance in a framework that can be used in (business or scientific) systems or even be shared, reused and evolved by third-party.

3 Conclusion

In this work, we introduce an original proposal about treating provenance of software development as a quality factor of (business or scientific) systems. Our research provides systematic approach based on conceptual modeling to represent provenance as NFR. We stress that our study is supported by consolidated methods of SE that do not substitute, but may compliment, traditional data provenance standards and specifications. We also agree with [11, 14] on the need for further empirical research on the use of NFRs and SIG during requirements engineering. As future work, we will expand the catalog through larger number of softgoals and operationalizations and evaluate it in different domains.

References

1. Sommerville, I., Sawyer, P., Viller, S.: Viewpoints for requirements elicitation: a practical approach, In: 3rd IEEE International Conference on Requirements Engineering, pp. 74–81 (1998)
2. Abran, A., Bourque, P., Dupuis, R., MooreDonald, J.W.: SWEBOK: Guide to the Software Engineering Body of Knowledge. IEEE Press, Piscataway (2004)

3. Chung, L.: Non-functional requirements. Department of Computer Science, The University of Texas at Dallas. http://www.utd.edu/~chung/RE/NFR-18–4-on-1.pdf

4. Simmhan, Y., Plale, B., Gannon, D.: A survey of data provenance in e-science. SIGMOD Rec. **34**(3), 31–36 (2005)

5. Chung, L., Nixon, B.A., Yu, E., Mylopoulos, J.: Non-functional Requirements in Software Engineering. Kluwer Academic Publishers, Boston (1999)

6. Supakkul, S., Hill, T., Chung, L., Than, T.T., Leite, J.C.S.P.: An NFR pattern approach to dealing with NFRs. In.: 18th IEEE International Requirements Engineering Conference, Sydney, vol. 18. pp. 179–188 (2010)

7. Cruz, S.M.S., Campos, M.L.M., Mattoso, M.: Towards a taxonomy of provenance in scientific workflow management systems. In: Proceedings of the SERVICES 2009 Congress, pp. 259–266. Los Angeles (2009)

8. Zhao, J. Bizer, C. Gil, Y. Missier, P.. Sahoo S.: Provenance requirements for the next version of RDF. In: Proceedings of the W3C Workshop - RDF Next Steps, Palo Alto (2010)

9. Moreau, L.: The foundations for provenance on the web. Found. Trends Web Sci. **2**, 99–241 (2010)

10. Serrano, M., Leite, J.C.S.P.: Capturing transparency-related requirements patterns through argumentation. In: 1st International. Workshop on Requirements Patterns (RePa), pp. 32–41 (2011)

11. Leal, A.L.C., Sousa, H.P., Leite, J.C.S.P.: Modelo orientado à meta para estabelecer relações de contribuição mútua entre Proveniência, Transparência e Confiança. In: XVII Workshop on Requirements Engineering (WER14), Pucón, Chile (2014). (in portuguese)

12. Asuncion, H.U., Asuncion, A.U., Taylor, R.N.: Software traceability with topic modeling. In: 32nd International Conference on Software Engineering (ICSE), pp. 95–104. Cape Town (2010)

13. Barbero, M., Didonet, M., Del Fabro, J.B.: Traceability and provenance issues in global model management. In: 3rd ECMDA-Traceability Workshop (2007)

14. Leal, A.L.C., Cruz, S.M.S.: Transparência em Experimentos Científicos Apoiados Em Proveniência: Uma Perspectiva para Workflows Científicos Transparentes. In: 2nd WTRANS-SBSI (2014)