**Variability Modeling in the Real - An Empirical Journey from Software Product Lines to Software Ecosystems**

Dissertation, 2013

Thorsten Berger

Variability modeling is one of the key disciplines to cope with complex variability in large software product lines. It aims at creating, evolving, and configuring variability models, which describe the common and variable characteristics, also known as features, of products in a product line. Since the introduction of feature models more than twenty years ago, many variability modeling languages and notations have been proposed both in academia and industry, followed by hundreds of publications on variability modeling techniques that have built upon these theoretical foundations.

Surprisingly, there are relatively few empirical studies that aim at understanding the use of such languages. What variability modeling concepts are actually used in practice? Do variability models applied in real-world look similar to those published in literature? In what technical and organizational contexts are variability models applicable?

We present an empirical study that addresses this research gap. Our goals are i) to verify existing theoretical research, and ii) to explore real-world variability modeling languages and models expressed in them. We study concepts and semantics of variability modeling languages conceived by practitioners, and the usage of these concepts in real, large-scale models. Our aim is to support variability modeling research by providing empirical data about the use of its core modeling concepts, by identifying and characterizing further concepts that have not been as widely addressed, and by providing realistic assumptions about scale, structure, content, and complexity of real-world variability models. We believe that our findings are of relevance to variability modeling researchers and tool designers, for example, those working on interactive product configurators or feature dependency checkers. Our extracted models provide realistic benchmarks that can be used to evaluate new techniques.

Recognizing the recent trend in software engineering to open up software platforms to facilitate inter-organizational reuse of software, we extend our empirical discourse to the emerging field of software ecosystems. As natural successors of successful product lines, ecosystems manage huge variability among and within their software assets, thus, represent a highly interesting class of systems to study variability modeling concepts and mechanisms. Our studied systems comprise eleven highly configurable software systems, two ecosystems with closed platforms, and three ecosystems relying on open platforms. Some of our subjects are among the largest successful systems in existence today. Results from a survey on industrial variability modeling complement these subjects.

Our overall results provide empirical evidence that the well-researched concepts of feature modeling are used in practice, but also that more advanced concepts are needed. We observe that assumptions about variability models in the literature do not hold. Our study also reveals that variability models work best in centralized variability management scenarios, and that they are fragile and have to be controlled by a small team. We also identify a particular type of dependencies that is increasingly used in open platforms and helps sustain the growth of ecosystems. Interestingly, while enabling distributed variability, these dependencies rely on a centralized and stable vocabulary. Finally, we formulate new hypotheses and research questions that provide direction for future research.