



The Board of Doctoral Study
Faculty of Mathematics and Physics
Charles University Prague
Ke Karlovu 3
12116 Prague 2

Prague, June 14, 2013

Viliam Šimko, Doctoral Thesis – Advisor's Reference

The submitted doctoral thesis concentrates upon two main topics: (i) formal verification of textual use-cases and (FOAM method) and (ii) statistical classification in requirements engineering (Prediction framework). The work has been done partially in the scope of the project GACR P103/11/1489, which aims at creating methods for verification of component systems using natural language specification. The thesis is focused on three main goals: (i) to formalize the behavior of textual use-cases in order to capture control flow and temporal dependencies among use-case steps, (ii) to design a method for verification of use-cases using existing model-checkers and to evaluate scalability and usability of the method, and (iii) to combine linguistic and software-engineering artifacts to support further statistical analysis.

The thesis starts with a motivation example describing a problem with sequencing of steps within a textual use-case specification. The author emphasizes that the majority of requirements is specified informally using natural language and states detailed goals of the thesis (Chapter 1). Then, the history of the presented "FOAM" tool is summarized (Chapter 2). In Chapter 3, the author establishes the formal basis for the presented FOAM method. A formal definition of annotated textual use-cases, a use-case model (UCM), a restricted version of a use-case model (rUCM) and the a use-case behavior automaton (UCBA) is presented following by a set of inference rules specifying the mapping between the structures. The text then continues by explaining the verification approach using the NuSMV model-checker and discusses the difficulties encountered. Next, a formal proof is presented concerning the expressive power of the method. Finally, the method is evaluated from the scalability (Section 3.5) and also from the usability point of view (Section 3.6). An implementation of the method is also provided which demonstrates verification of a textual use-case specification. The second major part of the thesis discusses an automated method for deriving a prototype domain model from text (Chapter 4). The chapter starts by a brief overview of techniques related to natural language processing and statistical classification, relevant to the rest of the chapter. Next, four phases of the method are explained in detail – preprocessing, feature selection, training and elicitation. Finally, the

classification models used by the method are evaluated, the performance results are discussed (Section 4.5). Chapter 5 compares the thesis with related work.

The thesis is based on a number of peer-reviewed papers published at international conferences. In particular these include the EUROMICRO SEAA'12 conference (proceedings by IEEE), FACS'11 conference (proceedings by Springer) and SNPD'10 conference (proceedings by Springer).

In my view, Viliam Šimko proved the ability of performing high-quality research with internationally relevant results. This thesis only confirms this fact. Thus, I strongly recommend the thesis for defense and to grant the Doctor degree to Viliam Šimko.



RNDr. Petr Hnětynka, Ph.D.
Advisor