Design and Evaluation of a Method to Reduce the Lexical Ambiguity of Requirement Specifications

Stephen Boyd (B.Eng(Comp.Sys))

A Thesis Submitted for the Degree of

DOCTOR OF PHILOSOPHY

University of Technology, Sydney
Faculty of Engineering and Information Technology

2009

Certificate of Authorship/Originality

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree, except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student

Production Note:
Signature removed prior to publication.

Acknowledgement

The author would like to thank three people for their support throughout this research:

1. I would like to thank my primary supervisor Prof. Didar Zowghi from the University of Technology, Sydney for her regular face-to-face discussions, problem solving advice, literature recommendations, and detailed review comments throughout the years of my research.

2. I would like to thank my secondary supervisor A/Prof. Vincenzo Gervasi from the University of Pisa, Italy for his in-depth review and accurate comments on my design solution and evaluation method.

3. Most importantly, I would like to thank my wife Alia for her understanding, patience, guidance, and continual support.

Preface

The author has previously published sections of this thesis in conference proceedings:

- Boyd, S., Zowghi, D., and Farroukh, A., 2005, 'Measuring the Expressiveness of a Constrained Natural Language: An Empirical Study', *Proceedings of the 13th International Conference on Requirements Engineering (RE'05)*, Paris, France, pp.339-349.
- Boyd, S., Zowghi, D., and Gervasi, V., 2007, 'Optimal-Constraint Lexicons for Requirement Specifications', *International Working Conference on Requirements Engineering Foundations for Software Quality (REFSQ '07)*, Trondheim, Norway, pp.203-217.

The author has been industry-supervisor for four engineering Capstone Project Reports that relate to this PhD thesis:

- Adi-wijaya, A., 2003, 'BOYDA: Object Oriented Requirements Engineering', University of Technology, Sydney, Unpublished Capstone Project Report.
- Farroukh, A., 2005, 'To define and use expressiveness to empirically derive the verbs of BOYDA for a particular domain', Faculty of Engineering, University of Technology, Sydney, Unpublished Capstone Project Report.
- Saeed, M., 2005, 'To define an unambiguous, disciplined and repeatable process for the coupling of BOYDA with an existing system modelling language', University of Technology, Sydney, Unpublished Capstone Project Report.
- Selvarajan, R., 2005, 'To develop a software tool that implements the BOYDA language and BOYDA modelling rules', University of Technology, Sydney, Unpublished Capstone Project Report.

Table of Contents

Abstract

The requirements engineering process has been criticised for its immaturity. Firstly, in the context of safety-critical systems, missing, misunderstood, and erroneous requirements have been attributed as the cause of many safety-system faults; and secondly, in the context of project success factors, many IT projects have identified requirement defects as a primary cause of being over-time or over-budget. Ambiguity is a requirement defect that is commonly associated with challenged IT projects, however there are but few empirical studies on how ambiguity can be reduced or eliminated from requirement specifications.

Eliminating the ambiguity inherent within a requirement specification is the seemingly unattainable ambition of the systems engineering zealot. This is because ambiguity is considered an unavoidable side-effect of using natural language, and most requirement specifications are written in natural language. One proposed solution to the ambiguity problem is to express requirements in Controlled Natural Language (CNL). CNLs enforce grammatical and/or lexical constraints to reduce the inherent ambiguity of natural language without sacrificing correctness, readability, or expressiveness. There is, however, a view in the literature that CNLs are overly restrictive and unnatural to read and write. Furthermore, the design and development of CNLs is both labour-intensive and time-intensive.

This thesis describes how a requirements specification can be automatically re-expressed in a way that significantly reduces its lexical ambiguity, without significantly reducing its correctness or conventionality. The thesis specifically focuses on lexical ambiguity, since this is the form of ambiguity most attributable to the lexicon used to express the specification. The term re-expression is used to distinguish this approach from that of CNLs, since the lexicon is not static, but is optimally selected on a word-by-word basis such that lexical ambiguity is minimised, whilst correctness and conventionality are maximised. Fundamental to the optimal word selection is a new concept: replaceability(W_1, W_2), which is the degree to which word W_1 can replace word W_2. The replaceability equation developed within this thesis is a function of semantic similarity, polysemy, frequency, and lexical width.

We implement a software prototype, and execute it on an existing industry-specification. A controlled experiment is used to measure the effects of the re-expression in terms of correctness, conventionality, and lexical ambiguity. Data are collected from project stakeholders using a questionnaire-style approach, and hypothesis testing is used to decide whether or not the optimal re-expression has significantly reduced lexical ambiguity without significantly reducing correctness or conventionality.