A COMPUTATIONAL FRAMEWORK FOR PREDICTING SOFTWARE QUALITY-IN-USE FROM SOFTWARE REVIEWS

Issa Atoum

Doctor of Philosophy
(Computer Science)
2015
A COMPUTATIONAL FRAMEWORK FOR PREDICTING SOFTWARE QUALITY-IN-USE FROM SOFTWARE REVIEWS

Issa Atoum

A thesis submitted

In fulfillment of the requirement for the degree of

Doctor of Philosophy (Computer Science)

Faculty of Computer Science and Information Technology

UNIVERSITI MALAYSIA SARAWAK
Dedications

I would like to dedicate this thesis to my parents (Ali and Moneera), wife Deema, brothers (Mohammed, Ahmed, Sultan, Sameer, and Samer), sisters (Mona, Manwa, Ibtisam, Amal, Samar, Sawsan, and Wafa), daughters (Asma, and Ghaida), sons (Khaled, Abed Al Rahman, and Ibrahim), and mother-in-law for their sincere love, confidence, and unselfish support, and to my fellow students at Universiti Malaysia Sarawak.

Issa Atoum
Acknowledgements

My sincerest appreciation and thanks go to my thesis advisors: Dr. Bong Chih How and Prof. Dr. Narayanan Kulathuramaiyer for their sincere patience, support, and guidance.

This work would not have been possible without the support of many people. Therefore, I would like to thank HE Minister of Education Malaysia for supporting this work by grant ERGS/ICT07 (01)/1018/2013 (15), the chancellor of Universiti Malaysia Sarawak for giving me the opportunity to do my research at the University and for supporting my research with Zamalah Graduate Scholarship. I would like also to thank the Dean and the members of the Post Graduate committee at Faculty of Computer Science and Information Technology for their guidance and help.

I would like to thank the annotators who annotated the dataset. Namely; Wendy Tan Wei Syn, Joseph Ramanair, Dr. Jecky Misieng, Ng Sing Yii, and Ama Chree Tamunonoemi Taribo. I would also like to thank the Examination Committee, the reviewers, Dr. Halikul Lenando, Dr. Shapiee Abdul Rahman, Dr. Bali Ranaivo Malançon, PG committee, all lecturers in FCSIT, and the IT support team for their support, comments, or feedback.
Abstract

Software *Quality-in-Use* (QinU) lies in the eyes of its users. QinU has gained its importance in e-government, mobile-based, and web applications. Currently, QinU is measured using either ISO standard (e.g. ISO 25010) or customized model approaches. These approaches tend to be incomplete and suffer from problems of user’s task definition sizing. Therefore, QinU measurement by these approaches has complexity resulting from quantifying QinU systematically.

This thesis proposes a computational and novel QinU Framework (QinUF) to measure QinU competently consuming software reviews. The significance of the framework is that it combines the semantic analysis and sentiment analysis research areas. In semantic similarity area, we proposed a novel Weighted Sentence Similarity Measure (WSSM) and developed an algorithm to predict a review-sentence QinU *topic* (QinU characteristic or software aspect). In the sentiment analysis area, we proposed an algorithm to classify and aggregate software review-sentences into QinU *topics*.

Experiments showed that the QinUF was able to predict software QinU *topics* on the fly, with high performance compared to selected *topic* prediction methods. Moreover, results of built *use* cases showed that employing minimal set of QinU *features* (properties) enable users to acquire software easily. As for future research, it is recommended to extend QinUF to support additional QinU characteristics, enhance sentiment orientation, specialize the framework to a certain software domain, and implement the framework in a large-scale system.

**Keywords:** Quality-in-Use; Sentiment Analysis; Text Similarity; ISO 25010; Quality-in-Use Framework; Weighted Sentence Similarity Measure.
ABSTRAK

Kualiti Perisian-alam-Penggunaan (QinU) terletak di mata pengguna perisian tersebut. QinU telah memperolehi kepentingannya dalam e-kerajaan, berasaskan konsep mudah alih dan aplikasi web. Pada masa kini, QinU diukur sama ada menggunakan ISO (contohnya ISO 25010) atau pendekatan model yang disesuaikan. Pendekatan-pendekatan ini cenderung untuk menjadi tidak lengkap dan mengalami masalah pentakrifan tugas saiz pengguna. Oleh itu, pengukuran QinU dengan pendekatan ini mempunyai kerumitan yang disebabkan oleh penjumlahan QinU secara sistematik.

Tesis ini mencadangkan satu Kerangka QinU pengkomputeran dan novel (QinUF) untuk mengukur QinU yang cekap dengan menggunakan penilaian perisian. Kepentingan rangka kerja ini adalah untuk menggabungkan analisis dan analisis sentimen kawasan kajian semantik. Di kawasan persamaan semantik, pengkaji telah mencadangkan wajaran Langkah Hukuman Persamaan novel (WSSM) dan membangunkan satu algoritma untuk meramalkan kajian-ayat QinU topik (QinU ciri atau perisian aspek). Dalam bidang analisis sentimen, pengkaji telah mencadangkan satu algoritma untuk mengelaskan dan meringkaskan perisian kajian-ayat ke topik QinU.

Eksperimen menunjukkan bahawa QinUF dapat meramalkan topik perisian QinU dengan cepat, serta prestasi yang tinggi berbanding dengan kaedah topicprediction yang dipilih. Selain itu, hasil daripada kes penggunaan yang dibina menunjukkan bahawa menggunakan set minimum ciri QinU (sifat-sifat) membolehkan pengguna untuk memperoleh perisian dengan mudah. Bagi kajian lanjutan, adalah disyorkan supaya memperluaskan QinUF untuk menyokong ciri-ciri QinU tambahan, meningkatkan orientasi sentimen, mengkhususkan rangka kerja untuk satu domain perisian tertentu, dan melaksanakan rangka kerja dalam sistem berskala besar.
Publications


# Table of Contents

Dedications ........................................................................................................................................... i

Acknowledgements ................................................................................................................................. ii

Abstract ...................................................................................................................................................... iii

ABSTRAK .................................................................................................................................................... iv

Publications .................................................................................................................................................. v

Table of Contents ...................................................................................................................................... vi

List of Tables ............................................................................................................................................. xi

List of Figures .............................................................................................................................................. xiii

List of Abbreviations ................................................................................................................................. xv

## Chapter 1 - Introduction .......................................................................................................................... 1

1.1. Background ........................................................................................................................................... 1

1.2. Problem Statement ............................................................................................................................... 4

1.2.1. Quality-in-Use Measurement Critiques ........................................................................................ 4

1.2.2. QinU Benchmarking Dataset ........................................................................................................ 5

1.2.3. Software Acquisition Needs .......................................................................................................... 6

1.2.4. QinU Scoring .................................................................................................................................... 6

1.3. Objectives .............................................................................................................................................. 8

1.4. Contributions ....................................................................................................................................... 9

1.5. Methodology ....................................................................................................................................... 10

1.6. Scope .................................................................................................................................................. 11

1.7. Thesis Outline ................................................................................................................................... 13
Chapter 2 - Literature Review .................................................................................15

2.1 Overview ........................................................................................................15

2.2 ISO Standard Models ....................................................................................16
  2.2.1 The ISO/IEC 25010:2011 Standard Model .............................................16

2.3 Customized Models .......................................................................................20

2.4 Computational Approaches ..........................................................................21

2.5 Summary of the Studied QinU Models .........................................................22

2.6 Challenges .....................................................................................................24

2.7 Conclusion .....................................................................................................27

Chapter 3 - Proposed Quality-in-Use Framework ...............................................29

3.1 Introduction ....................................................................................................29

3.2 QinU Measurement Components ..................................................................30

3.3 Proposed QinU Framework (QinUF) .............................................................31
  3.3.1 Walkthrough Example on QinUF .........................................................34

3.4 Detailed QinUF Components ........................................................................36
  3.4.1 QinU Topic Prediction ............................................................................37
  3.4.2 Polarity Classification ............................................................................42
  3.4.3 QinU Scoring ..........................................................................................43

3.5 QinUF Context of Use ....................................................................................46

3.6 Conclusion .....................................................................................................47

Chapter 4 - Proposed Weighted Sentence Similarity Measure (WSSM) ...............48
4.1. Introduction ...........................................................................................................................................48

4.2. Proposed Semantic Methods ..................................................................................................................49

4.3. Joint Weighted Word Similarity Measure (JWWS) ..................................................................................49

4.4. Proposed Weighted Sentence Similarity Measure (WSSM) ....................................................................52
   4.4.1. Weighted Sentence Similarity Measure Algorithm ........................................................................54

4.5. Proposed Pearson Over Root Mean Squared Scaled Error (PRMSSE) .................................................56

4.6. Results and Discussion ..........................................................................................................................58
   4.6.1. Benchmark Dataset .........................................................................................................................58
   4.6.2. Pearson and Spearman Correlations ..............................................................................................60
   4.6.3. Absolute Scaled error of Sentence Pairs Similarity ......................................................................62
   4.6.4. Results of the Proposed Semantic Similarity Evaluation Measure ..............................................64

4.7. Conclusion ..............................................................................................................................................66

Chapter 5 - Building a Gold Standard ...........................................................................................................68

5.1. Overview ...............................................................................................................................................68

5.2. Software Reviews Contents’ Problems ...................................................................................................69
   5.2.1. Grammatically Incorrect Sentences .............................................................................................70
   5.2.2. Software Reviews Language Uses ...............................................................................................70
   5.2.3. Rich and Poor Sentence Semantics .............................................................................................71

5.3. Software Reviews Selection ...................................................................................................................71

5.4. Experts Selection ....................................................................................................................................74

5.5. Annotation Scheme ................................................................................................................................74

5.6. Data Reconciliation ................................................................................................................................76
Chapter 6 - Framework Components’ Evaluation and Discussion

6.1 Overview

6.2 Evaluation Approach

6.3 Evaluation Metric

6.3.1 Number of Features

6.3.2 Topic Prediction Methods

6.3.3 Performance Criteria

6.4 Summary of Performance Measures

6.5 Dataset

6.6 Results and Discussion

6.6.1 F-Measure Performance

6.6.2 Performance on Receiver Operating Curve (ROC)

6.7 Conclusion

Chapter 7 - Practical Evaluation Using Use Cases

7.1 Introduction

7.2 Use Case 1: General Review – Most Accurate

7.3 Use Case 2: General Review – Most Inaccurate

7.4 Use Cases of Real-Life Scenario – Browser Selection

7.5 Conclusion
Chapter 8 - Conclusions and Future Work .......................................................................................... 111

8.1 Conclusions .................................................................................................................................. 111
  8.1.1 QinU Measurement Components ............................................................................................ 112
  8.1.2 Novel Overarching QinU Framework (QinUF) ........................................................................ 112
  8.1.3 Configurable and Proven Framework ....................................................................................... 113
  8.1.4 QinU Dataset .......................................................................................................................... 114
  8.1.5 Novel Word and Sentence Similarity Measures (JWSM, WSSM) ........................................... 114

8.2 Future Work .................................................................................................................................. 115
  8.2.1 QinUF Extension .................................................................................................................... 115
  8.2.2 Sentence Sentiment Orientation ............................................................................................. 115
  8.2.3 Framework Specialization ...................................................................................................... 116
  8.2.4 Framework Implementation ................................................................................................... 116
  8.2.5 QinU Trends .......................................................................................................................... 116

Bibliography ..................................................................................................................................... 117

Appendixes ....................................................................................................................................... 138

Appendix A - Training Material ......................................................................................................... 138
Appendix B - Sentence Similarity Measures Comparison ................................................................. 145
Appendix C - Experiments With Google Search Engine .................................................................. 150
Appendix D - Top Ten Topic-Wise Features for the Proposed Method WSSM ............................. 152
Appendix E - Customized Quality-in-Use Models ......................................................................... 153
Appendix F - Semantic Similarity ................................................................................................... 158
Appendix G - Sentiment Analysis Related Works ......................................................................... 178
List of Tables

Table 2.1 Definitions of quality-in-use characteristics .......................................................... 18
Table 2.2 Comparison of quality-in-use models versus the ISO 25010.................................... 23
Table 3.1 QinUF walkthrough example .................................................................................. 34
Table 4.3 Pearson and Spearman correlations with respect to human ratings on STS-65 dataset.61
Table 5.1 Distribution of software reviews over software categories ....................................... 73
Table 5.2 Distribution of software sentences over software categories .................................... 73
Table 5.3 Kappa agreements for the gold standard dataset ....................................................... 79
Table 6.1 Evaluation metric .................................................................................................... 87
Table 6.2 Topic prediction methods settings ............................................................................ 88
Table 6.3 Average F-measure of different topic prediction methods ........................................ 93
Table 6.4 Average Topic-Wise AUC of different topic prediction methods ................................. 95
Table 7.1 Use Case 1 QinU topic predictions and scores. ......................................................... 101
Table 7.2 Topics detected for Use Case 1 ............................................................................. 103
Table 7.3 Use Case 2 QinU topic predictions and scores. ......................................................... 104
Table 7.4 Topics detected for Use Case 2 ............................................................................. 106
Table 7.5 Explanation of low performing review by sentence. ................................................ 107
Table A.1 Definitions of quality-in-use characteristics. ......................................................... 140
Table A.2 Annotation Checklist .............................................................................................. 140
Table B.1 Comparison between sentence similarity measures. .............................................. 145
Table B.2 STS-65 sentence pairs .......................................................................................... 147
Table C.1 Top ten words for each QinU topic extracted using LDA over Google SE ............ 151
Table D.1 Top ten features for each QinU topic using WSSM ............................................... 152
Table F.1 Summary of word and sentence similarity approaches................................................ 164

Table F.1 Comparison of word pairs similarity given same path and depth of (1:10, 4:7) ........... 175

Table F.2 Comparison of word pair’s similarity given equal IC values in Brown Corpus.......... 175
List of Figures

Figure 1.1 Research methodology.................................................................10
Figure 2.2 Structure of the Quality Measurement division..............................19
Figure 3.1 Conceptual Model of QinU Measurement.......................................30
Figure 3.2 Proposed QinU Framework ..........................................................33
Figure 4.2 Absolute Scaled Error of WSSM-S, WSSM-T, and Google Trigram method..........63
Figure 4.3 Methods performance over the proposed PRMSSE ......................65
Figure 5.1 Sentence Annotation Scheme .....................................................76
Figure 5.2 Data Reconciliation ..................................................................77
Figure 6.1 Low performance similarity measures .........................................91
Figure 6.2 High performance similarity measures .......................................91
Figure 6.3 Distribution of topics (features = 55) ......................................94
Figure 6.4 Multiclass ROC curve when number of features =55 ...............94
Figure 6.5 AUC over WSSM F-measures and number of features .............97
Figure 6.6 Multiclass ROC curve over features .........................................98
Figure 7.1 QinU using QinUF and gold standard ......................................102
Figure 7.2 QinU using QinUF and gold standard on UseCase2 ..................106
Figure 7.3 Real-life software QinU scoring (browser selection) ....................109
Figure A.1 Annotation system login screen.............................................141
Figure A.2 Annotation system main screen.............................................142
Figure A.3 Annotation system entry screen.............................................143
Figure A.4 Annotation system update screen.......................................144
Figure C.1 Performance of WSSM using fixed list of features extracted from Google. ....151
Figure F.4 Vector space classification into three .................................................................170

Figure F.1 Fragment of WordNet showing nodes in an is-a relationship. ...............................173
## List of Abbreviations

<table>
<thead>
<tr>
<th>Acronym / Synonym</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Q2U</td>
<td>Quality, Quality-in-Use, Usability and User experience</td>
</tr>
<tr>
<td>ASE</td>
<td>Absolute Scaled Error</td>
</tr>
<tr>
<td>BNC</td>
<td>British National Corpus</td>
</tr>
<tr>
<td>BPQRM</td>
<td>The Business Process Quality Reference-Model</td>
</tr>
<tr>
<td>CMMI</td>
<td>Capability Maturity Model Integration</td>
</tr>
<tr>
<td>CQMLs</td>
<td>Comprehensive Quality Model Landscapes</td>
</tr>
<tr>
<td>DEMATEL</td>
<td>Decision-Making Trial and Evaluation Laboratory</td>
</tr>
<tr>
<td>DTW</td>
<td>Dynamic Time Warping</td>
</tr>
<tr>
<td>HAL</td>
<td>Hyperspace Analogues to Language</td>
</tr>
<tr>
<td>HMM-LDA</td>
<td>Hidden Markov Model of LDA</td>
</tr>
<tr>
<td>IBM</td>
<td>The International Business Machines Corporation</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>ISO 25022</td>
<td>Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Measurement of quality in use</td>
</tr>
<tr>
<td>Acronym / Synonym</td>
<td>Meaning</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ISO 25023</td>
<td>Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Measurement of system and software product quality</td>
</tr>
<tr>
<td>ISO 9000</td>
<td>The ISO 9000 family addresses various aspects of quality management and contains some of ISO’s best known standards</td>
</tr>
<tr>
<td>ISO 9241-11:1998</td>
<td>Guidance on Usability</td>
</tr>
<tr>
<td>ISO 9241-210:2008</td>
<td>Human-centred design process for interactive systems</td>
</tr>
<tr>
<td>ISO/IEC 14598-1:1999</td>
<td>Information technology -- Software product evaluation -- Part 1: General overview. This standard has been revised by: ISO/IEC 25040:201</td>
</tr>
<tr>
<td>ISO/IEC 25010:2011</td>
<td>Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models</td>
</tr>
<tr>
<td>ISO/IEC 25024</td>
<td>Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Measurement of data quality</td>
</tr>
<tr>
<td>Acronym / Synonym</td>
<td>Meaning</td>
</tr>
<tr>
<td>------------------</td>
<td>---------</td>
</tr>
<tr>
<td>ISO/IEC 9126:2001</td>
<td>Software engineering — Product quality was an international standard for the evaluation of software quality. It has been replaced by ISO/IEC 25010:2011</td>
</tr>
<tr>
<td>LCS</td>
<td>Least Common Sub summer</td>
</tr>
<tr>
<td>LRT</td>
<td>likelihood ratio test</td>
</tr>
<tr>
<td>LSA</td>
<td>Latent Semantic Analysis</td>
</tr>
<tr>
<td>LSI</td>
<td>Latent Semantic Indexing</td>
</tr>
<tr>
<td>MAP</td>
<td>Maximum A Posteriori</td>
</tr>
<tr>
<td>MISRA</td>
<td>The Motor Industry Software Reliability Association</td>
</tr>
<tr>
<td>NSID</td>
<td>Normalized Search engine Index Distance</td>
</tr>
<tr>
<td>OANC</td>
<td>Open America National Corpus</td>
</tr>
<tr>
<td>PMI-IR</td>
<td>Point-Wise Mutual Information and Information Retrieval</td>
</tr>
<tr>
<td>PRMSSE</td>
<td>Pearson over Root Mean Squared Scaled Error</td>
</tr>
<tr>
<td>QinU</td>
<td>Quality-in-Use</td>
</tr>
<tr>
<td>Acronym / Synonym</td>
<td>Meaning</td>
</tr>
<tr>
<td>------------------</td>
<td>---------</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>RMSSE</td>
<td>Root Mean Squared Scaled Error</td>
</tr>
<tr>
<td>SAM</td>
<td>Sentiment Aspect Match</td>
</tr>
<tr>
<td>SAP</td>
<td>Systems Applications Products</td>
</tr>
<tr>
<td>SemEval</td>
<td>Semantic Evaluation</td>
</tr>
<tr>
<td>SIQinU</td>
<td>Strategy for understanding and Improving Quality-in-useUse</td>
</tr>
<tr>
<td>SMAC</td>
<td>Sentiment Match plus Aspect Coverage</td>
</tr>
<tr>
<td>SME</td>
<td>Service-based Mobile Ecosystem</td>
</tr>
<tr>
<td>SO/IEC 25021:201</td>
<td>Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Quality measure elements</td>
</tr>
<tr>
<td>SQuaRE</td>
<td>Software Product Quality Requirements and Evaluation</td>
</tr>
<tr>
<td>SQUID</td>
<td>Software QUality In Development</td>
</tr>
<tr>
<td>SRA</td>
<td>Semantic Role Annotation</td>
</tr>
<tr>
<td>STSS</td>
<td>Short Text Semantic Similarity</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>SWET-QUM</td>
<td>Semantic Web Exploration Tools Quality-in-Use Model</td>
</tr>
<tr>
<td>TF-IDF</td>
<td>Term Frequency –Inverse Document Frequency</td>
</tr>
<tr>
<td>WSSM-S</td>
<td>Standard Weighted Sentence Similarity Measure</td>
</tr>
<tr>
<td>WSSM-T</td>
<td>Tuned Weighted Sentence Similarity Measure</td>
</tr>
</tbody>
</table>
Chapter 1 - Introduction

1.1. Background

With a large amount of software published online it is essential for users to find the software that satisfy their stated or implied requirements. Users often require an easy to use, resource efficient and risk free software. Typically, users spend a lot of time reading surveys or online reviews, trying to find the software that fulfills their requirements. Accordingly, they repeatedly seek adequate software quality by their own viewpoints. Such viewpoints are jointly called the software Quality-in-Use (QinU).

Software QinU has several benefits to users. For example, a good software system design allows users to operate efficiently. Consequently, it improves users’ productivity, by allowing them to concentrate on their requirements rather than mastering software usage. Moreover, a well-designed software system will reduce training, system errors, and gain users’ trust. However, because of time and resource constraints, users may not view all reviews. Hence, it is important to measure QinU quantitatively in order to compare different types of software, thus allowing users to acquire the best software quality in their own perspectives.

Currently, there are two approaches to measure the software QinU: The International Organization for Standardization (ISO) models (ISO/IEC, 2004, 2011) and the customized model (e.g. Oliveira et al., 2014; Osman & Osman, 2013) approaches. Both approaches tend to be deficient to measure QinU. The ISO standard models tend to be difficult to customize (Mordal et al., 2013) and suffer from various critiques (Deissenboeck et al., 2009) while customized models tend to cover specific software domains (Alnanih et al., 2014; Hsu & Lee, 2014). Nevertheless, the consideration of this
thesis is the ISO models because they contain the landmark QinU specifications and are widely accepted.

The ISO/IEC 25010:2011 (ISO/IEC, 2011) defines two dimensions; the software product quality and the software QinU. The former is related to properties intrinsic to the product such as ‘line of code’ whereas the latter is related to the human interaction with software such as ‘navigation’. According to the ISO standard the QinU is defined as “the degree to which a product or system can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency, freedom from risk and satisfaction in specific contexts of use.” (ISO/IEC, 2011, p. 8). Therefore, the QinU is related to human use of the product which makes its measurement challenging. Throughout this thesis the QinU will refer to the ISO standard specification ISO/IEC 25010:2011 (referred to as ISO 25010 hereafter, or just the ISO standard).

According to the ISO 25010, the QinU model has five characteristics, namely; effectiveness, efficiency, freedom from risk, satisfaction and context coverage. For example, the effectiveness and the efficiency characteristics are related to the user job completion (e.g. ‘designing a logo’) and user’s system resources expenditure (e.g. RAM space) respectively. A software quality characteristic can have common software properties grouped in one characteristic. For example, efficiency characteristic may include performance, reliability, and compatibility properties. Hence, an adequate software QinU requires measuring such properties. However, since QinU represents quality from the human’s viewpoint, the human-part of QinU makes its measurement challenging. The challenge arises due to the difficulty of quantifying human viewpoints using available QinU models. That being the case, software reviews from experienced users (humans) play an important role for software acquisition decision and it has real economic values for targeted products (Ghose
& Ipeirotis, 2011). Therefore, processing these reviews could reveal underlying user’s viewpoints (i.e. QinU).

For example given the sentence: “This software is pretty fast, takes less memory.” several QinU properties can be inferred from this sentence. The sentence is talking about using software resources thus it is mapped to the QinU efficiency characteristic (called a topic throughout this thesis). The reason is that the words fast and memory are representative features for efficiency in this regard which maps to the ISO standard definition of efficiency (ISO/IEC, 1998, p. 6).

This thesis processes software reviews using Natural Language Processing (NLP) techniques in order to estimate software QinU scores. The NLP used techniques are based on semantic similarity to extract common QinU characteristics and sentiment analysis to extract opinions on these characteristics. Accordingly, we propose a QinU Framework (QinUF). Basically, the framework works as follows. First, the framework classifies software sentences into QinU characteristics (called topics) using a proposed sentence similarity measure. Then, the framework classifies the sentences to their polarity orientation values (positive, negative, and neutral). After that, the framework aggregates the sentences’ polarities into their QinU scores grouped by their topics. Finally, the framework obtains the overall estimated software QinU score.

In the next section, we present the problem statement. Next, we summarize the research objectives. Then, we highlight the research contributions. After that, we describe the research methodology. Finally, we discuss the research scope and present the thesis outline.
1.2. Problem Statement

This research identifies three main problems.

1.2.1. Quality-in-Use Measurement Critiques

Quality-in-Use measurement is very essential for users because it comprehends their needs. At the same time, it can guide software providers to enhance their products and provide customer satisfaction. Currently, there are two different approaches to measure the software QinU: the ISO standard models (ISO/IEC, 2004, 2011), and the customized models (e.g. Oliveira et al., 2014; Osman & Osman, 2013).

However, these types of models are challenging. The major challenge of QinU models is in its task measurement. To measure QinU according to its underlying models, it is a prerequisite to agree with users on tasks (such as ‘open a file’ or ‘press a button’) that they should execute in order to complete software function. The series of such tasks performs a user-needed function such as ‘pay a bill’. Conversely, users are rarely involved in the software development cycle and often have dynamic requirements over time (Ishikawa, 1985). So, they are not requested to perform or design software tasks to measure QinU. The worst, task measurement embraces the variety of tasks from one software function to another, and from software to another. Hence, QinU measurement is not directly quantified.

On the first hand, the ISO standard models has many critiques. They are unclear in their purposes. The purpose of quality model can be in characterization, understanding, evaluation or prediction (Deissenboeck et al., 2009). They are also hard to customize (Alnanih et al., 2014; Deissenboeck et al., 2009; Kläs et al., 2009), thus the aggregation of the QinU evaluation scores is challenging.