CONTENTS

Acknowledgements iii
Abstract v
Resumen vii
Resum ix
Contents xi
List of Figures xv
List of Tables xix

Part I Introduction 1

Chapter 1. Introduction 3
  1.1. Motivation .............................................................. 4
  1.2. Dissertation Objectives ........................................... 6
  1.3. Dissertation Contribution ........................................ 10
  1.4. Dissertation Overview ............................................ 11
  1.5. Research Methodology ............................................ 13
  1.6. Dissertation Structure ............................................ 15

Part II Investigation Problem 17

Chapter 2. Background 19
  2.1. Overview of the Chapter .......................................... 20
  2.2. Model Driven Development ....................................... 20
    2.2.1 Models in MDD ................................................ 21
    2.2.2 Domain Specific Language in MDD ......................... 21
  2.3. Software Maintenance Tasks ................................... 22
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4. Machine Learning</td>
<td>23</td>
</tr>
<tr>
<td>2.5. Runtime Example</td>
<td>24</td>
</tr>
<tr>
<td>2.5.1 Railway Domain</td>
<td>24</td>
</tr>
<tr>
<td>2.5.2 Train Control and Management Language</td>
<td>25</td>
</tr>
<tr>
<td>3.1. Overview of the chapter</td>
<td>30</td>
</tr>
<tr>
<td>3.2. Search Process</td>
<td>30</td>
</tr>
<tr>
<td>3.3. Related Works</td>
<td>32</td>
</tr>
<tr>
<td>3.3.1 Related works regarding the kind of software artifacts</td>
<td>34</td>
</tr>
<tr>
<td>3.3.2 Related works regarding the ML techniques</td>
<td>35</td>
</tr>
<tr>
<td>3.4. Research Questions for the Systematic Review</td>
<td>37</td>
</tr>
<tr>
<td>Part III Treatment Design</td>
<td>39</td>
</tr>
<tr>
<td>Chapter 4. Ontological Evolutional Encoding</td>
<td>41</td>
</tr>
<tr>
<td>4.1. Overview of the chapter</td>
<td>42</td>
</tr>
<tr>
<td>4.2. Input and output artifacts</td>
<td>43</td>
</tr>
<tr>
<td>4.3. Stages of the ontological evolutionary encoding</td>
<td>45</td>
</tr>
<tr>
<td>4.3.1 Ontological Encoding</td>
<td>45</td>
</tr>
<tr>
<td>4.3.2 Evolutionary Encoding</td>
<td>47</td>
</tr>
<tr>
<td>4.3.3 Feature Selection</td>
<td>48</td>
</tr>
<tr>
<td>4.4. Summary of the ontological evolutionary encoding</td>
<td>49</td>
</tr>
<tr>
<td>Chapter 5. Fitness Function</td>
<td>51</td>
</tr>
<tr>
<td>5.1. Overview of the chapter</td>
<td>52</td>
</tr>
<tr>
<td>5.2. Training phase</td>
<td>53</td>
</tr>
<tr>
<td>5.2.1 Input and output artifacts</td>
<td>54</td>
</tr>
<tr>
<td>5.2.2 Steps of training phase</td>
<td>55</td>
</tr>
<tr>
<td>5.3. Testing phase</td>
<td>57</td>
</tr>
<tr>
<td>5.3.1 Input and output artifacts</td>
<td>58</td>
</tr>
<tr>
<td>5.3.2 Steps of testing phase</td>
<td>60</td>
</tr>
<tr>
<td>5.3.3 Classifier</td>
<td>61</td>
</tr>
</tbody>
</table>
5.4. Summary of the fitness function .......................... 62

Chapter 6. Evolutionary Algorithm  65
6.1. Overview of the chapter .................................. 66
6.2. Input and output artifacts ................................ 67
6.3. Stages of the evolutionary algorithm ................. 69
   6.3.1 Initialization .......................................... 69
   6.3.2 Genetic Operations .................................... 70
   6.3.3 Fitness Function ....................................... 74
6.4. Overview of the FRAME approach .................... 75

Part IV  Treatment Validation .............................. 77

Chapter 7. Evaluation Design ................................. 79
7.1. Overview of the chapter .................................. 80
7.2. Real Case: Test Cases and Oracle ...................... 82
7.3. Approaches under Evaluation ......................... 84
   7.3.1 TLR-FRAME: FRAME approach .................. 85
   7.3.2 TLR-Linguistic: Linguistic Rule-Based approach .... 86
   7.3.3 TLR-IR: Information Retrieval approach .. 88
   7.3.4 TLR-FNN: Feedforward Neural Network approach ... 89
   7.3.5 TLR-RNN: Recurrent Neural Network approach .... 91
   7.3.6 TLR-LtoR: Learning to Rank approach .... 93
7.4. Comparison and Measure ................................ 94
7.5. Threats to validity ..................................... 96

Chapter 8. Results of the Evaluation .................. 101
8.1. Overview of the chapter .................................. 102
8.2. Reported results ....................................... 102
8.3. Statistical Analysis .................................... 103
   8.3.1 Statistical Test .................................... 103
   8.3.2 Post Hoc Analysis .................................. 104
   8.3.3 Effect Size ......................................... 105
8.4. Research Questions for the Evaluation ........................................ 108
8.5. Discussion ................................................................................. 109
     8.5.1 Prerequisites and Properties .............................................. 109
     8.5.2 Advantages of TLR-FRAME ............................................. 112

Part V Conclusion .............................................................................. 117

Chapter 9. Conclusion and Ongoing Research .................................... 119
    9.1. Overview of the chapter ......................................................... 120
    9.2. Research Questions .............................................................. 120
          9.2.1 Results of the Objective 1 ........................................... 120
          9.2.2 Results of the Objective 2 ........................................... 121
          9.2.3 Results of the Objective 3 ........................................... 122
    9.3. Ongoing Research ................................................................. 123
          9.3.1 Feature Location and Bug Localization ....................... 123
          9.3.2 Ontological Evolutionary Encoding: Granularity .......... 124
          9.3.3 Fitness Function: Deep Learning ................................ 125
          9.3.4 Domain Ontology: Metamodel ..................................... 125
          9.3.5 Knowledge Base: Size and Model Fragments ............. 126
    9.4. Concluding Remark .............................................................. 126

Bibliography ..................................................................................... 129

Part VI Publications ........................................................................... 145

Chapter 10. Publications ................................................................. 147
    10.1. REVE’17 Paper (First page) .............................................. 148
    10.2. ER’17 Paper (First page) ................................................. 150
    10.3. CoopIS’17 Paper (First page) ........................................... 152
    10.4. MODELS’18 Paper (First page) ....................................... 154
    10.5. JSS’20 Paper (First page) ................................................. 156
# List of Figures

## Chapter 1. Introduction
1.1 Example of a product model and a model fragment for fragment retrieval in software maintenance tasks on models ............ 5
1.2 Design problem behind RQ2: Characterization of model fragments 7
1.3 Design problem behind RQ3: Assessing model fragments regarding a query .......................................................... 8
1.4 Design problem behind RQ4: Extraction of model fragments from a model .......................................................... 9
1.5 Overview of the work performed as part of the dissertation ... 12
1.6 Design cycle of the research methodology followed in this dissertation ................................................................. 14

## Chapter 2. Background
2.1 Example of a TCML model and model fragment ................. 26

## Chapter 3. State of the Art
3.1 Overview of the search process ...................................... 32
3.2 Number of articles by year ........................................... 33
3.3 Number of articles by software maintenance task according to their main software artifact ........................................ 34
3.4 Number of articles by software maintenance task according to the ML technique applied ........................................ 36

## Chapter 4. Ontological Evolutionary Encoding
List of Figures

4.1 Overview of the ontological evolutionary encoding in the FRAME approach ....................... 42
4.2 Example of a domain ontology ......................... 43
4.3 Example of a sample in the knowledge base ............. 44
4.4 Example of a feature vector in the training set .......... 45
4.5 Example of the inputs and output for the ontological encoding stage ................................ 46
4.6 Example of the input and output for the evolutionary encoding stage .................................. 48
4.7 Example of feature selection based on a mask .......... 49
4.8 Overview of the ontological evolutionary encoding approach. ...................................... 50

Chapter 5. Fitness Function
5.1 Overview of the process to evaluate and rank model fragments 53
5.2 Example of k-fold method ................................. 57
5.3 Example of queries: requirement and feature descriptions .... 58
5.4 Example of two model fragments from the same model .......... 59
5.5 Example of a feature vector in the testing set ............. 60
5.6 Example of a model fragment ranking .................. 60
5.7 Overview of the ontological evolutionary encoding approach. ...................................... 63

Chapter 6. Evolutionary Algorithm
6.1 Overview of the evolutionary algorithm in the FRAME approach 67
6.2 Example of a model, taken from a real-world train .......... 68
6.3 Examples of Model Fragment Encoding ........................ 71
6.4 Example of genetic operations ................................ 72
6.5 Overview of the FRAME approach .......................... 76
Chapter 7. Evaluation Design

7.1 Setup of the evaluation ........................................ 81
7.2 Distribution of fitness values in the knowledge base ........ 83
7.3 Example of requirement-to-object-model rules ............... 86
7.4 Example of Traceability Link Recovery using Latent Semantic
    Indexing .............................................................. 88
7.5 Feedforward Neural Network .................................... 90
7.6 Elman Simple Recurrent Neural Network ..................... 92
LIST OF TABLES

Chapter 3. State of the Art
3.1 Terms for the search string regarding PICO criteria 31
3.2 Articles grouped by Software Maintenance Task 33
3.3 Articles grouped by software maintenance task and software artifact 35
3.4 Articles grouped by software maintenance task and ML technique 37

Chapter 7. Evaluation Design
7.1 Threats to Validity 96

Chapter 8. Results of the Evaluation
8.1 Mean Values and Standard Deviations for Precision, Recall, F-Measure, and Matthews Correlation Coefficient (MCC) for the approaches 102
8.2 Quade test statistic and $p – Values$ 103
8.3 Holm’s Post Hoc $p – Values$ 104
8.4 $\hat{A}_{12}$ statistic for each pair of approaches 106
8.5 Required artifacts for each approach 109
8.6 Artifacts whose properties have an impact on the results 110