A REVERSE ENGINEERING APPROACH OF A DECISION SUPPORT SYSTEM FOR FORENSIC DNA ANALYSIS

AHMAD FAIZ BIN GHAZALI

MASTER OF SCIENCE
UNIVERSITI MALAYSIA TERENGGANU
2012

1100087774

Perpustakaan Sultanah Nur Zahirah
Universiti Malaysia Terengganu (UMT)

PERPUSTAKAAN SULTANAH NUR ZAHIRAH UMT

tesis
T 58.62 .A3 2012

1100087774
A reverse engineering approach of a decision support system for forensic dna analysis / Ahmad Faiz Ghazali.

PERPUSTAKAAN SULTANAH NUR ZAHIRAH
UNIVERSITI MALAYSIA TERENGGANU (UMT)
21030 KUALA TERENGGANU

1100087774

Lihat sebelah

HAK MILIK
PERPUSTAKAAN SULTANAH NUR ZAHIRAH UMT

A REVERSE ENGINEERING APPROACH OF A DECISION SUPPORT SYSTEM FOR FORENSIC DNA ANALYSIS

AHMAD FAIZ BIN GHAZALI

Thesis Submitted in Fulfillment of the Requirement
for the Degree of Master of Science in the
Faculty of Science and Technology
Universiti Malaysia Terengganu

August 2012

TABLE OF CONTENTS

Page

DEDICATION v
ABSTRACTS vi
ABSTRAK viii
ACKNOWLEDGEMENTS x
APPROVAL xi
DECLARATION xiii
LIST OF TABLES xiv
LIST OF FIGURES xvi
LIST OF ABBREVIATIONS xix

CHAPTER

1. INTRODUCTION 1
   1.1 Background of the Study 1
   1.2 Statement of the Problems 4
   1.3 Motivation and Significance of the Study 6
   1.4 Objectives of the Study 7
   1.5 Thesis Organization 8

2. LITERATURE REVIEW 9
   2.1 Decision Support System (DSS) 9
   2.2 Forensic Science (FS) and Decision Support 18
   2.3 Bioinformatics and Forensic DNA Analysis (FDA) 21
   2.4 Statistics 25
   2.5 Statistical Methods in Forensic Science (FS) 27
   2.6 Bayesian Networks (BN) 29
   2.7 Likelihood Ratio (LR) 40
   2.8 Paternity Test 43
   2.9 Motherless case of Paternity Index (PI) 44
   2.10 Hypothetical case to determine PI using mathematical formula 46
   2.11 Software Engineering (SE) and Reverse Engineering (RE) 57
      2.11.1 Reverse Engineering (RE) of User Interfaces 60
      2.11.2 Reverse Engineering (RE) to Understand Processes 62
      2.11.3 Reverse Engineering (RE) to Understand Data 63
   2.12 Data Visualization 65
   2.13 Overall Comparisons of Related Works 67
   2.14 Conclusion 72

3. RESEARCH METHODOLOGY 73
   3.1 Phases of Methodology 73
   3.2 Mathematical Models 75
      3.2.1 Paternity Index 78
      3.2.2 Combined Paternity Index 79

3.2.3 Probability of Paternity 79
3.3 The Proposed Architecture 82
3.4 The Proposed Probabilistic Graphical Model 83
3.5 The Proposed Approach 90
3.6 The Developed Algorithms 96
3.7 The Flow of Processes in the Software 99
3.8 System Sequence Diagram and Use Case Diagram for Main Engine 104
3.9 Window Navigation Design 115

4. IMPLEMENTATION 116
4.1 Programming languages used for software development 116
4.2 Reverse Engineering of National DNA Database (NDNAD) 118
4.3 Reverse Engineering of Combined DNA Index System (CODIS) 119
4.4 Reverse Engineering of Forensic DNA Databank of Malaysia (FDDM) 120

5. EVALUATION AND DISCUSSION 129
5.1 Introduction to Evaluation 130
5.2 Methods of Evaluation 130
5.3 Software Testing 130
5.4 Data Testing of Forensic DNA Analysis (FDA) with Real Case Study 131
5.5 Statistical Results of Forensic DNA Analysis (FDA) 132
5.6 Validation of Mathematical Formulas Programmed 154
5.7 Usability Study 155
5.8 Results and Discussion of the Usability Study 156
5.9 Heuristics Evaluation 159
5.10 Pilot Test 160
5.11 Real Test 160

6. CONCLUSION 166
6.1 Summary 166
6.2 Advantages of the Proposed Architecture 169
6.3 Disadvantages of the Proposed Architecture 170
6.4 Contributions 177
6.4.1 Contributions to decision-making processes 177
6.4.2 Contributions to Web-based Decision Support System (DSS) 178
6.4.3 Contributions to Forensic Science (FS) 179
6.4.4 Contributions to process modeling 180
6.4.5 Contributions to academics and innovations for country 181
6.5 Future Works 183

REFERENCES 185
APPENDICES 195
Appendix A: Result for Plagiarism Checking 195
Appendix B: Recommendation Letter 196
Appendix C: Questionnaire 197
Appendix D1: Case 1 203
Appendix D2: Case 2 204
Appendix D3: Case 3 205
Appendix D4: Case 4 206
Appendix D5: Case 5 207

BIODATA OF THE AUTHOR 208

Perpustakaan Sultanah Nur Zahirah
Universiti Malaysia Terengganu (UMT)

DEDICATION

To my parents, teachers and lecturers

Abstract of thesis presented to the Senate of Universiti Malaysia Terengganu in fulfillment of the requirement for the degree of Master of Science.

A REVERSE ENGINEERING APPROACH OF A DECISION SUPPORT SYSTEM FOR FORENSIC DNA ANALYSIS

AHMAD FAIZ GHAZALI

August 2012

Main Supervisor: Associate Professor Noor Maizura Mohamad Noor, Ph.D.

Co-Supervisor : Professor Md. Yazid Mohd Saman, Ph.D.
Associate Professor Zafarina Zainuddin, Ph.D.

Faculty : Science and Technology

Decision Support System (DSS) can gather various sources of information, choose relevant and related knowledge intelligently, and structure the decision process efficiently. The problem is how to develop DSS for forensic science (FS)? How to perform computerised forensic DNA analysis (FDA) in FS? Is it possible to reverse engineer the existing system and then reengineered them with enhancements? In this study, methods to be used are discovered after reverse engineering of the user interfaces from the existing system. Research activities are divided into three phases; design of architecture; develop prototype of Web-based DSS based on the proposed architecture; and evaluate the prototype. The outcomes of this research are the type of analysis determined, system architecture designed and prototype developed. A DSS for FDA is implemented using Bayesian networks (BN) and likelihood ratios (LR) methods. Through reverse engineering processes of existing system, Forensic DNA Databank of Malaysia (FDDM), Architecture of Decision Support for Forensic DNA (ADeFor) are designed for a reengineered system called Forensic DNA Analysis Software (ForAS). Data validation using five real case studies proved that results obtained using ForAS is almost 100% accurate. Usability study with domain

experts shows that in scale from 0 to 9, with score 7.33 for data representation, 7.22 for system's learnability, and 6.73 for overall reaction of the system, the designed architecture and the developed prototype is proven reliable and helpful for FDA in FS.

Abstrak tesis yang dikemukakan kepada Senat Universiti Malaysia Terengganu sebagai memenuhi keperluan untuk ijazah Sarjana Sains.

PENDEKATAN KEJURUTERAAN TERBALIK BAGI SISTEM SOKONGAN KEPUTUSAN UNTUK ANALISIS FORENSIK DNA

AHMAD FAIZ GHAZALI

Ogos 2012

Penyelia Utama : **Profesor Madya Noor Maizura Mohamad Noor, Ph.D.**

Penyelia Bersama : **Profesor Md. Yazid Mohd Saman, Ph.D.**
Profesor Madya Zafarina Zainuddin, Ph.D.

Fakulti : **Sains dan Teknologi**

Sistem sokongan keputusan (DSS) dapat mengumpulkan pelbagai sumber maklumat, memilih ilmu yang berkaitan dan relevan secara pintar, dan menstrukturkan proses pembuatan keputusan dengan efisien. Masalahnya adalah bagaimanakah cara untuk membangunkan DSS untuk sains forensik (FS)? Bagaimanakah cara untuk melakukan analisis forensik DNA (FDA) menggunakan komputer? Bolehkah kejuruteraan terbalik daripada sistem yang sedia ada dibina semula dengan lebih banyak manfaat? Dalam kajian ini, kaedah yang patut digunakan hanya dapat dikenalpasti selepas kejuruteraan terbalik daripada antaramuka sistem yang sedia ada telah dilakukan. Aktiviti penyelidikan dapat dibahagikan kepada tiga fasa; rekabentuk senibina yang telah direkayasa semula, pembangunan prototaip DSS berasaskan Web yang dibina berdasarkan senibina; dan penilaian prototaip. Hasil penyelidikan ini termasuklah penemuan cara untuk melaksanakan analisis, senibina sistem yang direkabentuk dan prototaip yang dibangunkan. DSS untuk FDA diimplementasi menggunakan kaedah rangkaian Bayesian (BN) dan nisbah kebolehjadian (LR). Melalui proses kejuruteraan terbalik daripada sistem yang sedia ada, "Forensic DNA Databank of Malaysia" (FDDM), "Architecture of Decision

Support for Forensic DNA" (ADeFor) direkabentuk untuk sistem yang direkayasa semula yang dipanggil "Forensic DNA Analysis Software" (ForAS). Pengesahan data menggunakan lima kes sebenar membuktikan bahawa hasil yang diperolehi menggunakan ForAS adalah hampir 100% tepat. Kajian kepenggunaan dengan pakar forensik menunjukkan bahawa daripada skala 0 hingga 9, dengan skor 7.33 untuk persembahan data, 7.22 untuk tahap kemudahan bagi mempelajari sistem, dan 6.73 untuk reaksi keseluruhan terhadap sistem, senibina yang direkabentuk dan prototaip yang dibangunkan terbukti boleh dipercayai dan berguna untuk FDA di dalam FS.