

TABLE OF CONTENT

| | |
|--------------------------------|------|
| APPROVAL PAGE..... | iii |
| STATEMENT | iv |
| FOREWORD | vi |
| TABLE OF CONTENT | viii |
| LIST OF TABLES | xi |
| LIST OF FIGURES..... | xii |
| ABSTRACT..... | xiv |
| INTISARI | xv |
| CHAPTER I..... | 1 |
| INTRODUCTION | 1 |
| 1.1 Research Background..... | 1 |
| 1.2 Research Problem | 4 |
| 1.3 Research Scope | 4 |
| 1.4 Research Objective | 4 |
| 1.5 Research Advantages | 4 |
| 1.6 Research Methodology | 5 |
| 1.7 Thesis Organization..... | 6 |
| CHAPTER II..... | 8 |
| LITERATURE REVIEW | 8 |
| CHAPTER III..... | 12 |
| THEORETICAL BASIS | 12 |
| 3.1 Program Comprehension..... | 12 |

| | | |
|----------------------------|--|-----------|
| 3.2 | Object-Oriented Programming | 13 |
| 3.2.1 | Classes | 13 |
| 3.2.2 | Variables and Methods..... | 14 |
| 3.2.3 | Constructors..... | 14 |
| 3.2.4 | Access Control Modifiers | 15 |
| 3.3 | Stereotypes | 15 |
| 3.4 | srcML..... | 16 |
| 3.5 | Data Mining | 18 |
| 3.5.1 | Clustering Techniques..... | 18 |
| 3.6 | K-Prototypes Clustering..... | 19 |
| 3.6.1 | Cost Function | 23 |
| 3.6.2 | Similarity Measure | 24 |
| 3.7 | Silhouette Coefficient..... | 25 |
| 3.8 | Sum of Squared Errors..... | 26 |
| CHAPTER IV | | 28 |
| ANALYSIS AND DESIGN | | 28 |
| 4.1 | General Analysis | 28 |
| 4.2 | Feature Extraction Design | 30 |
| 4.2.1 | Input Model Design | 30 |
| 4.2.2 | srcML Representation Design | 33 |
| 4.2.3 | XPath Queries Design | 34 |
| 4.3 | Clustering Design | 35 |
| 4.4 | Experiment Design | 36 |
| 4.5 | Evaluation Design | 37 |
| CHAPTER V | | 38 |

| | |
|--|-----------|
| IMPLEMENTATION | 38 |
| 5.1 Specification | 38 |
| 5.2 Feature Extraction Implementation | 38 |
| 5.2.1 srcML Representation..... | 38 |
| 5.2.2 XPath Queries | 41 |
| 5.3 Clustering Implementation | 50 |
| 5.4 Evaluation Implementation | 51 |
| 5.4.1 Sum of Squared Errors..... | 52 |
| 5.4.2 Silhouette Coefficient | 52 |
| CHAPTER VI | 55 |
| RESULT AND DISCUSSION | 55 |
| 6.1 Feature Extraction Result | 55 |
| 6.2 Clustering Evaluation | 56 |
| 6.2.1 Clustering Validity with Silhouette Measure | 57 |
| 6.2.2 Clustering Validity with SSE | 58 |
| 6.3 Discussion | 59 |
| CHAPTER VII | 65 |
| CONCLUSION AND FUTURE WORKS..... | 65 |
| 7.1 Conclusion | 65 |
| 7.2 Future Works | 65 |
| REFERENCES | 67 |