

Table of Contents

I	Prelude	1
1	Introduction	3
1.1	Contributions	4
1.1.1	Clone Stability	4
1.1.2	Clone Authorship	5
1.1.3	Clones and Programmer Performance	5
1.2	Previously Published Content	5
1.3	Thesis Outline	6
II	Software Clones	9
2	Definition and Terminology	11
2.1	Definition	12
2.2	Terminology	13
3	Detecting Clones	21
3.1	Detection	21
3.1.1	Textual Comparison	22
3.1.2	Token Comparison	23
3.1.3	Syntactical Comparison	25
3.1.4	Metric Comparison	27
3.1.5	Dependency Comparison	28
3.1.6	Other techniques	28
3.2	Evolution	29
3.2.1	Incremental Detection	29
3.2.2	Tracking Clones	30

3.2.3	Evolution Patterns	32
3.3	Clone Detection for this Thesis	33
3.3.1	Detection and Tracking Approach	34
3.3.2	Improving Precision	34
3.3.3	Detection Thresholds	36
4	Causes and Effects of Clones	37
4.1	Causes	38
4.1.1	Technical Reasons	38
4.1.2	Organizational Reasons	40
4.1.3	Personal Reasons	41
4.2	Effects	42
4.2.1	Program Changeability	44
4.2.2	Program Correctness	48
4.2.3	Program Comprehension	53
5	Managing Clones	57
5.1	Preventive	57
5.2	Compensative	58
5.3	Corrective	60
III	Empirical Studies on the Effects of Clones	61
6	Clone Stability	63
6.1	Research Questions	64
6.2	The Original Study Compared to Ours	66
6.3	Study Setup	67
6.3.1	Clone Detection	67
6.3.2	Subject Systems	68
6.3.3	Data Collection	74
6.3.4	Measurement	76
6.4	Results	79
6.4.1	Validation of the original studies' results	79
6.4.2	Generalizability	82
6.4.3	The Relationship of Detection Parameters and Stability	82
6.4.4	Deletion Characteristics	90

6.4.5	Distribution Over Time	93
6.5	Threats to Validity	101
6.5.1	Internal validity	101
6.5.2	Construct Validity	102
6.5.3	External Validity	103
6.6	Discussion	104
6.7	Further Studies on Clone Stability	106
6.7.1	Modification Frequency	106
6.7.2	Proportional Stability Metrics	108
6.8	Conclusion	113
7	Clone Authorship	115
7.1	Research Questions	116
7.2	Authorship Detection	117
7.2.1	Authorship vs. Ownership	117
7.2.2	Existing Techniques	118
7.2.3	Token-based Authorship Detection	119
7.3	Study Setup	124
7.3.1	Clone Detection	124
7.3.2	Subject Systems	125
7.4	Results	128
7.4.1	Clone Authors	129
7.4.2	Change Frequency	129
7.4.3	Change Consistency	132
7.4.4	Intentions and Bugs	132
7.5	Discussion	134
7.6	Threats to Validity	135
7.6.1	Internal Validity	135
7.6.2	Construct Validity	136
7.6.3	External Validity	136
7.7	Conclusion	137
8	Clones and Programmer Performance	139
8.1	Research Questions	141
8.2	Experimental Design	141
8.2.1	Hypotheses and Variables	142
8.2.2	Design	143

8.2.3 Subjects 144

8.2.4 Objects 145

8.2.5 Instrumentation 152

8.3 Execution 153

8.4 Analysis 154

8.4.1 Descriptive Statistics 155

8.4.2 Hypothesis Testing 163

8.5 Interpretation & Discussion 165

8.5.1 Time 165

8.5.2 Correctness 166

8.6 Threats to Validity 168

8.6.1 Internal validity 168

8.6.2 Construct Validity 170

8.6.3 External Validity 171

8.7 Conclusion 172

IV Finale 173

9 Conclusion 175

Appendices 181

A Clone Stability Charts 183

B Experiment Replication Manual 195

B.1 Introduction 195

B.2 Contents of the Replication Set 196

B.3 Shopping List 197

B.4 Preparing an Execution 197

B.4.1 Requirements for Participation 197

B.4.2 Preparing the materials 198

B.5 What data are collected 199

B.6 Recorded Data not Analyzed in the Paper 200

B.7 Executing the Experiment 201

B.7.1 Sessions 201

B.7.2 Phase 1: Arrival 201

B.7.3 Phase 2: Introduction 201

B.7.4	Phase 3: Distribution of Material	201
B.7.5	Phase 4: Programming Tasks	202
B.7.6	Phase 5: Result Collection	202
B.7.7	Phase 6: End	203
B.8	Data Analysis	203
B.8.1	Logs	203
B.8.2	Correctness	203
B.8.3	Analysis Scripts	204
B.9	How to keep the experiment sound	204
B.10	Known Issues & Questions Frequently Asked by Participants	204
C	Experiment Handout	207
	List of Figures	211
	List of Tables	215
	Bibliography	217