



University of Zagreb

FACULTY OF ORGANIZATION AND INFORMATICS

Matija Novak

**NEKI JAKO JAKO JAKO JAKO JAKO
JAKO VELIIIIKI NASLOV**

DOCTORAL THESIS

Varaždin, 2019



Sveučilište u Zagrebu

FAKLULTET ORGANIZACIJE I INFORMATIKE

Matija Novak

**SOME VERY VERY VERY VERY VERY
VERY LOOOOOOONG TITLE**

DOKTORSKI RAD

Varaždin, 2019.



University of Zagreb

FACULTY OF ORGANIZATION AND INFORMATICS

Matija Novak

**NEKI JAKO JAKO JAKO JAKO JAKO
JAKO VELIIIIKI NASLOV**

DOCTORAL THESIS

Supervisors:

professor NAmE Surname, PhD

professor NAmE Surname, PhD

Varaždin, 2019

ACKNOWLEDGMENTS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

PROŠIRENI SAŽETAK

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

Some „asdf” [2]

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra

metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

KEYWORDS

key1, key2, ke3

CONTENTS

1	Introduction	1
2	Related work	3
3	Methodology	4
3.0.1	Top authors	4
3.1	Research constraints	5
4	Similarity detection tools	7
5	Comparison measures	10
5.1	Precision, Recall and F-beta	11
6	Preprocessing techniques	12
7	Result analysis	14
7.1	SOCO dataset analysis	14
7.1.1	Results for D1 assignments	15
8	Future work	19
9	Conclusion	20
	Appendix A Contrast codings	21
	Appendix B Shapiro-Wilk normality test	22
	Appendix C Model comparisons	24
	Appendix D Constrast effect sizes	26
	Appendix E Interaction graphs	28
	Bibliography	30

LIST OF FIGURES

3.1	Number of articles in databases per year	5
3.2	Number of articles in databases per year	6
7.1	F1 score for SOCO D1 assignment with 3*IRQ	15
7.2	D1 assignments - residuals	15
7.3	F1 mean comparison for SOCO D1	16
E.1	Interaction graphs for SOCO D1 - part 1	28
E.2	Interaction graphs for SOCO D1 - part 2	29

LIST OF TABLES

3.1	Authors with most articles in SLR and most citations	5
4.1	JPlag-java and SIM-java calibration dataset similarities	8
4.2	SIM-java calibrated with JPlag-java as base tool ^a	9
6.1	JPlag-text and SIM-text PPTest result	13
7.1	SOCO dataset structure for experiment	14
7.2	ANOVA results for SOCO D1	16
7.3	Contrasts results for SOCO D1	17
7.4	Simple effect analysis result for SOCO D1	18
A.1	Tool contrast codings	21
A.2	SOCO technique contrast codings	21
C.1	Common Code Remove (MLM) comparison for SOCO D1	25
D.1	Contrasts effect sizes for SOCO D1	26
D.2	Simple effect analysis effect sizes for SOCO D1	27

GLOSSARY

CCR Common Code Remove.

MLM Common Code Remove.

PPT PreProcessing Technique.

SOCO Common Code Remove.

INTRODUCTION

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

Cites examples [1, 2, 6] some footnote¹ level.

The overall goal of this research is to :

G1 : *Goal 1*

¹footnote text

G2 : *Goal 2.*

From the goals the following hypotheses emerge:

H1 : *Hip 1*

H2 : *Hip 2*

Since G2 is partially fulfilled by H2 the rest is covered by the following research question:

Q1 : *Queston*

The rest of the thesis is structured as follows. In Chapter 2, related work is described including short sections describing ... Chapter 3 ... Conclusion 9 concludes. [3, 5]

Glossary PreProcessing Technique (PPT) and plural PPTs and Common Code Remove (CCR)

RELATED WORK

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

METHODOLOGY

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

3.0.1 Top authors

Another way of finding relevant papers was to search for main authors in the field. A....

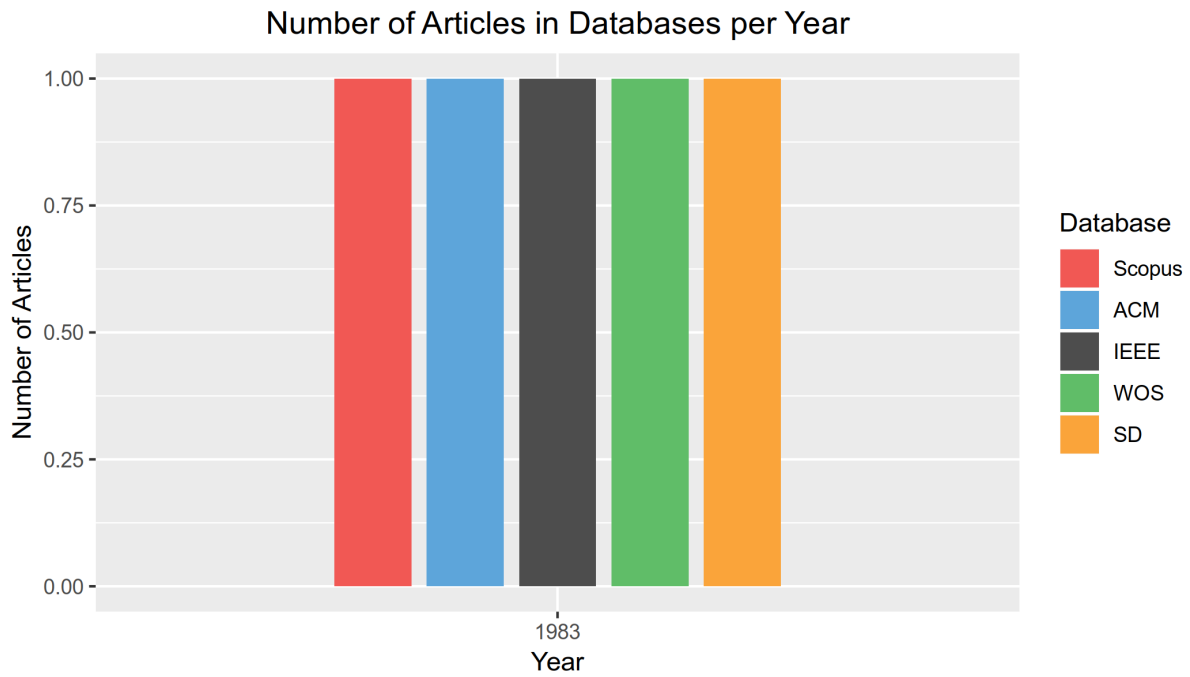


Figure 3.1: Number of articles in databases per year

Table 3.1: Authors with most articles in SLR and most citations

Author	Number of articles	Number of top cited articles	Articles
J, J	7	3	[1]
C, C	6	2	[1]

3.1 Research constraints

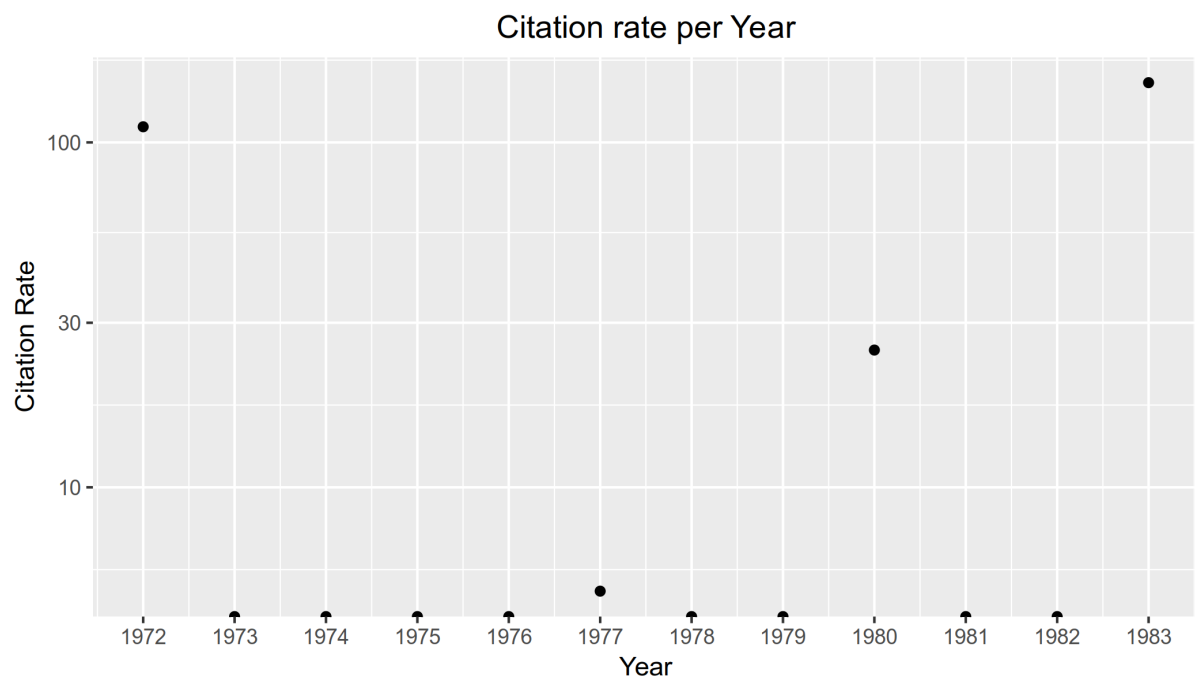


Figure 3.2: Number of articles in databases per year

CHAPTER 4

SIMILARITY DETECTION TOOLS

Table 4.1: JPlag-java and SIM-java calibration dataset similarities

Case name	Similarity			
	Java version		Text version	
	JPlag (t=3)	SIM (r=22)	JPlag (t=8)	SIM (r=4)
manual - 0% example A	10.0	0.0	0.0	6.0
manual - 0% example B	11.4	0.0	0.0	2.3
manual - 50% Copy	64.1	58.5	40.7	56.5
manual - 50% Simple Obfuscation	67.5	67.0	48.8	64.0
manual - 50% Complex Obfuscation	64.1	58.0	39.9	57.0
manual - 100% Copy	100.0	100.0	99.7	100.0
manual - 100% Simple Obfuscation	98.3	100.0	96.7	100.0
manual - 100% Complex Obfuscation	84.8	92.5	79.1	98.5
SOCO 0 - N - (084-258)	45.7	14.5	0.0	4.7
SOCO 1 - N - (000-001)	22.1	7.5	4.1	30.6
SOCO 2 - N - (002-003)	28.9	0.0	4.0	10.2
SOCO 3 - P - (003-004)	77.6	54.0	49.6	61.5
SOCO 4 - P - (107-112)	100.0	100.0	33.3	75.0
SOCO 5 - N - (052-077)	50.0	0.0	4.2	17.5
SOCO 6 - N - (011-037)	38.5	0.0	0.0	3.0
SOCO 7 - P - (062-064)	87.4	85.5	77.7	85.5
SOCO 8 - N - (144-192)	57.1	1.0	0.0	15.3
SOCO 9 - N - (037-093)	41.4	0.0	0.0	12.0

Note:

N in the case name marks a non-plagiarised case

P in the case name marks a plagiarised case

Table 4.2: SIM-java calibrated with JPlag-java as base tool ^a

JPlag-java (t)	SIM -java (r - best)	CDS
1	5	6
2	6	7
3	7	8
4	8	9
5	9	9
6	10	7
7	11	8
8	12	9
9	13	10
10	14	10
11	15	8
12	16	9
13	17	10
14	18	11
15	19	11

^a Footnote

COMPARISON MEASURES

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

Performance Index is calculated as

$$Performance_index = D_{max} * \frac{1}{(tp + fn) * (1 - P_L)}$$

or

$$Performance_Index = MAX_{1-esm} \left(\left(tp@esm - esm * \frac{tp+fn}{\frac{N}{2}} \right) * \frac{1}{(tp+fn) * (1 - \frac{tp+fn}{\frac{N}{2}})} \right)$$

where: N is number of submissions; tp – true positive essential matches, fn – false negative essential matches; esm is number of essential matches; tp@esm – true positives essential matches at the number of essential matches.

5.1 Precision, Recall and F-beta

Precision and Recall are calculated as follows:

$$P = \frac{tp}{tp+fp}, \in [0, 1] \quad R = \frac{tp}{tp+fn}, \in [0, 1]$$

F-beta is calculated as follows:

$$F_\beta = \frac{(\beta^2 + 1) * P * R}{(\beta^2 * P) + R}, \in [0, 1]$$

PREPROCESSING TECHNIQUES

Table 6.1: JPlag-text and SIM-text PPTest result

Technique Name	JPlag Text				SIM Text			
	E1	E2	E3	E4	E1	E2	E3	E4
No Preprocessing	1	2	3	4	5	6	7	8
Techniques								
Remove Comments (RC)	2	3	4	5	6	7	8	9
Remove White Spaces (RWS)	3	4	5	6	7	8	9	10
Normalise (NOR)	4	5	6	7	8	9	10	11
Common Code Remove (CCR)	5	6	7	8	9	10	11	12
Template Exclusion (TE)	6	7	8	9	10	11	12	13
Combinations								
All (TE-RC-CCR-NOR-RWS)	7	8	9	10	11	12	13	14
RC-NOR-RWS	8	9	10	11	12	13	14	15
NOR-RC	9	10	11	12	13	14	15	16
RC-NOR	10	11	12	13	14	15	16	17
TE-CCR	11	12	13	14	15	16	17	18

RESULT ANALYSIS

This Chapter provides the complete results of this research. Before the results are presented there is a description of some preparations that are important for the analysis.

7.1 SOCO dataset analysis

Preparation for statistical analysis

The full model used was:

$$\begin{aligned} & lmer(F1 \sim Tool + Technique + Tool : Technique + \\ & (1|Participant) + (1|Tool : Participant) + (1|Technique : Participant), \\ & data = SOCO.Dn, REML = FALSE) \end{aligned}$$

The null model was:

$$\begin{aligned} & lmer(F1 \sim (1|Participant) + (1|Tool : Participant) + (1|Technique : Participant), \\ & data = SOCO.Dn, REML = FALSE) \end{aligned}$$

The full model for the simple effects analysis was:

$$\begin{aligned} & lmer(F1 \sim ToolTechniqueCombo + \\ & (1|Participant) + (1|Tool : Participant) + (1|Technique : Participant), \\ & data = SOCO.Dn, REML = FALSE) \end{aligned}$$

and the null model was the same as written above. Recall that the term *Participant* represents one subset assignment (e.g., Dn-1, Dn-2, Dn-3, etc.) from the D1, D2, D3 or D4 group of assignments.

Table 7.1: SOCO dataset structure for experiment

Collection	Assignment		Number of			
	original	new	Files	Plagiarized files	Matches	Plagiarised matches
Test	A1	D1	188	86	17,578	54
Test	A2	D2	175	75	15,225	47
Test	B1	D3	218	124	23,653	73

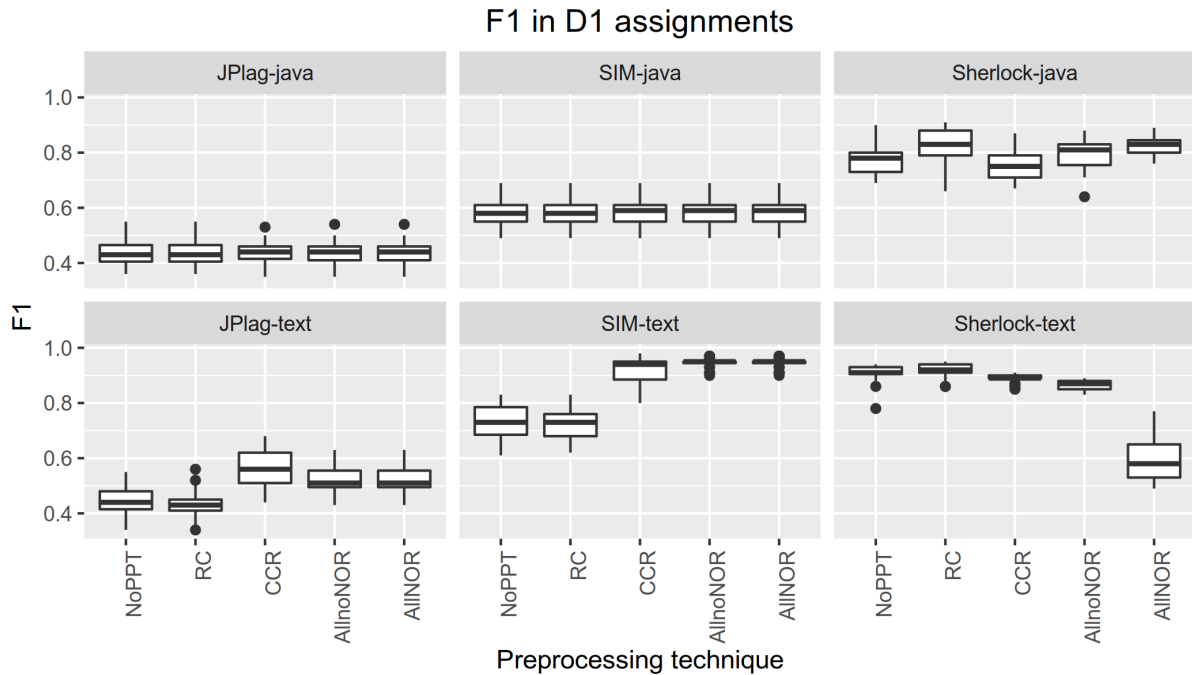


Figure 7.1: F1 score for SOCO D1 assignment with 3*IRQ

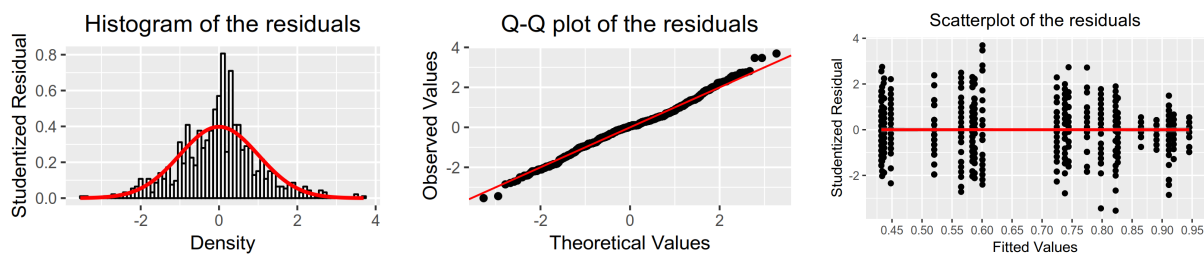


Figure 7.2: D1 assignments - residuals

To perform the bootstrap hypothesis testing the function *boot* from package *boot* was used, and to get confidence intervals the function *boot.ci* from package *boot* was used or function *bootMer* from the *lme4* package. Models were compared using the *PBmodcomp* function from the *pbkrtest* package. The full list of packages used in R is presented in Appendix ??.

7.1.1 Results for D1 assignments

D1 assignments are created from the A1 assignment in the Common Code Remove (SOCO) dataset.

To confirm ...

Since

To (Table 7.2).

Figure 7.3 .

By looking at Figure 7.3 .

Table 7.2: ANOVA results for SOCO D1

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	p.boot
Tool	4.07	0.814	5	155	818.0	0.0000	0.0001
Technique	0.34	0.085	4	124	85.9	0.0000	0.0001
Tool:Technique	3.95	0.198	20	620	198.7	0.0000	0.0001

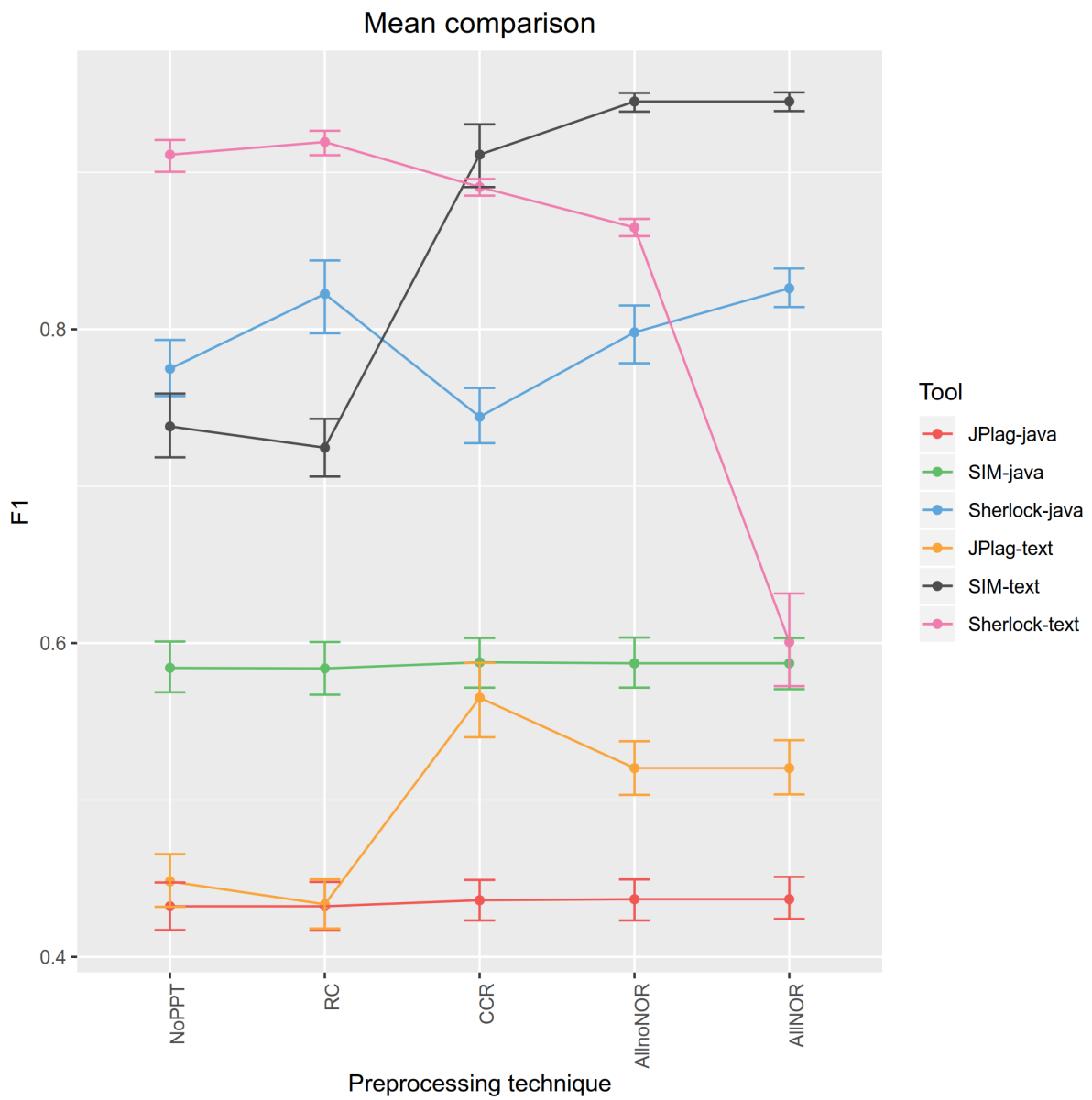
**Figure 7.3:** F1 mean comparison for SOCO D1

Table 7.3: Contrasts results for SOCO D1

Name	Estimate	SE	df	t.value	Pr(> t)	p.boot
(Intercept)	0.667	0	31	228.7363	0.0000	0.0001
Tool.TextvsJava	-0.062	0	155	-23.7553	0.0000	0.0001
TT.SHvsOthers	-0.054	0	155	-20.6218	0.0000	0.0001
TJ.SHvsOthers	-0.094	0	155	-35.9487	0.0000	0.0001
TT.SIMvsJPlag	-0.178	0	155	-39.1278	0.0000	0.0001
TJ.SIMvsJPlag	-0.076	0	155	-16.6443	0.0000	0.0001
NoPPTvsPPT	0.005	0	124	8.8276	0.0000	0.0001
SinglevsCombo	0.001	0	124	0.5975	0.5513	0.8008
RCvsCCR	0.018	0	124	10.8250	0.0000	0.0001
AnoNvsAN	-0.020	0	124	-11.6699	0.0000	0.0001
Tool.TextvsJava:NoPPTvsPPT	-0.003	0	620	-5.4427	0.0000	0.0001
TT.SHvsOthers:NoPPTvsPPT	0.013	0	620	25.1444	0.0000	0.0001
TJ.SHvsOthers:NoPPTvsPPT	-0.001	0	620	-2.5992	0.0096	0.0073
TT.SIMvsJPlag:NoPPTvsPPT	-0.008	0	620	-9.1208	0.0000	0.0001
TJ.SIMvsJPlag:NoPPTvsPPT	0.000	0	620	0.1080	0.9140	0.5662
Tool.TextvsJava:SinglevsCombo	0.005	0	620	4.0800	0.0001	0.0221
TT.SHvsOthers:SinglevsCombo	0.041	0	620	35.5109	0.0000	0.0001
TJ.SHvsOthers:SinglevsCombo	-0.004	0	620	-3.8591	0.0001	0.0004
TT.SIMvsJPlag:SinglevsCombo	-0.027	0	620	-13.2677	0.0000	0.0001
TJ.SIMvsJPlag:SinglevsCombo	0.000	0	620	0.1611	0.8721	0.5987
Tool.TextvsJava:RCvsCCR	-0.030	0	620	-18.3619	0.0000	0.0001
TT.SHvsOthers:RCvsCCR	0.031	0	620	19.1510	0.0000	0.0001
TJ.SHvsOthers:RCvsCCR	0.014	0	620	8.3837	0.0000	0.0001
TT.SIMvsJPlag:RCvsCCR	-0.014	0	620	-4.8688	0.0000	0.0001
TJ.SIMvsJPlag:RCvsCCR	0.000	0	620	0.0000	1.0000	0.5130
Tool.TextvsJava:AnoNvsAN	0.024	0	620	14.8934	0.0000	0.0001
TT.SHvsOthers:AnoNvsAN	0.044	0	620	26.9264	0.0000	0.0001
TJ.SHvsOthers:AnoNvsAN	-0.005	0	620	-2.8603	0.0044	0.0034
TT.SIMvsJPlag:AnoNvsAN	0.000	0	620	0.0000	1.0000	0.5259
TJ.SIMvsJPlag:AnoNvsAN	0.000	0	620	0.0000	1.0000	0.5265

Note:

TT - ToolText, TJ - ToolJava, SH - Sherlock, AnoN - AllnoNOR, AN - AllNOR

Table 7.4: Simple effect analysis result for SOCO D1

Name	Estimate	SE	df	t.value	Pr(> t)	p.boot
(Intercept)	0.667	0	31.0	228.7363	0.0000	0.0001
Textvs,Java	-0.062	0	155.0	-23.7553	0.0000	0.0001
TT.SHvsOthers	-0.054	0	155.0	-20.6218	0.0000	0.0001
TT.SIMvsJPlag	-0.178	0	155.0	-39.1278	0.0000	0.0001
TJ.SHvsOthers	-0.094	0	155.0	-35.9487	0.0000	0.0001
TJ.SIMvsJPlag	-0.076	0	155.0	-16.6443	0.0000	0.0001
TT.SH.NoPPTvsPPT	-0.018	0	743.6	-14.5161	0.0000	0.0001
TT.SH.SinglevsCombo	-0.086	0	743.6	-30.2497	0.0000	0.0001
TT.SH.RCvsCCR	-0.014	0	743.6	-3.5650	0.0004	0.0003
TT.SH.AllnoNORvsAllNOR	-0.132	0	743.6	-32.8057	0.0000	0.0001
TT.JPlag.NoPPTvsPPT	0.012	0	743.6	9.7027	0.0000	0.0001
TT.JPlag.SinglevsCombo	0.010	0	743.6	3.6821	0.0002	0.0453
TT.JPlag.RCvsCCR	0.066	0	743.6	16.3428	0.0000	0.0001
TT.JPlag.AllnoNORvsAllNOR	0.000	0	743.6	0.0000	1.0000	0.5204
TT.SIM.NoPPTvsPPT	0.029	0	743.6	22.5342	0.0000	0.0001
TT.SIM.SinglevsCombo	0.064	0	743.6	22.3474	0.0000	0.0001
TT.SIM.RCvsCCR	0.093	0	743.6	23.1923	0.0000	0.0001
TT.SIM.AllnoNORvsAllNOR	0.000	0	743.6	0.0000	1.0000	0.5241
TJ.SH.NoPPTvsPPT	0.005	0	743.6	3.5974	0.0003	0.0580
TJ.SH.SinglevsCombo	0.014	0	743.6	5.0416	0.0000	0.0018
TJ.SH.RCvsCCR	-0.039	0	743.6	-9.7336	0.0000	0.0001
TJ.SH.AllnoNORvsAllNOR	0.014	0	743.6	3.4849	0.0005	0.0699
TJ.JPlag.NoPPTvsPPT	0.001	0	743.6	0.5067	0.6125	0.7739
TJ.JPlag.SinglevsCombo	0.001	0	743.6	0.4532	0.6506	0.7326
TJ.JPlag.RCvsCCR	0.002	0	743.6	0.4807	0.6309	0.7469
TJ.JPlag.AllnoNORvsAllNOR	0.000	0	743.6	0.0000	1.0000	0.5271
TJ.SIM.NoPPTvsPPT	0.000	0	743.6	0.3547	0.7229	0.7017
TJ.SIM.SinglevsCombo	0.001	0	743.6	0.2266	0.8208	0.6289
TJ.SIM.RCvsCCR	0.002	0	743.6	0.4807	0.6309	0.7522
TJ.SIM.AllnoNORvsAllNOR	0.000	0	743.6	0.0000	1.0000	0.5198

Note:

TT - ToolText, TJ - ToolJava, SH - Sherlock

FUTURE WORK

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

CONCLUSION

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

CONTRAST CODINGS

To have the planned comparisons

Table A.1: Tool contrast codings

Tool	Text			Java		
	JPlag	SIM	Sherlock	JPlag	SIM	Sherlock
TextvsJava	1	1	1	-1	-1	-1
Text.SherlockvsOthers	0	0	0	1	1	-2
Java.SherlockvsOthers	1	1	-2	0	0	0
Text.SIMvsJPlag	0	0	0	1	-1	0
Java.SIMvsJPlag	1	-1	0	0	0	0

Table A.2: SOCO technique contrast codings

Technique	NoPPT	RC	CCR	AllnoNOR	AllNOR
NoPPTvsPPT	-4	1	1	1	1
SinglevsCombo	0	-1	-1	1	1
RCvsCCR	0	-1	1	0	0
AllnoNORvsAllNOR	0	0	0	-1	1

SHAPIRO-WILK NORMALITY TEST

Results of Shapiro-Wilk normality test for SOCO dataset D1 assignmenets group:

```
## $`Normal distribution`
## [1] "JPlag-java - NoPPT with W=0.9646739, p=0.3856066;"
## [2] "JPlag-java - RC with W=0.9646739, p=0.3856066;"
## [3] "JPlag-java - CCR with W=0.981933, p=0.8640068;"
## [4] "JPlag-java - AllnoNOR with W=0.9736222, p=0.6233963;"
## [5] "JPlag-java - AllNOR with W=0.9736222, p=0.6233963;"
## [6] "SIM-java - NoPPT with W=0.9684525, p=0.4775116;"
## [7] "SIM-java - RC with W=0.9703781, p=0.5295313;"
## [8] "SIM-java - CCR with W=0.9645588, p=0.3830315;"
## [9] "SIM-java - AllnoNOR with W=0.963926, p=0.3691145;"
## [10] "SIM-java - AllNOR with W=0.963926, p=0.3691145;"
## [11] "Sherlock-java - NoPPT with W=0.9622875, p=0.3349699;"
## [12] "Sherlock-java - RC with W=0.9397757, p=0.08134755;"
## [13] "Sherlock-java - CCR with W=0.9488192, p=0.1448137;"
## [14] "Sherlock-java - AllnoNOR with W=0.9316043, p=0.04848914;"
## [15] "Sherlock-java - AllNOR with W=0.9490796, p=0.1472336;"
## [16] "JPlag-text - NoPPT with W=0.968684, p=0.4835915;"
## [17] "JPlag-text - RC with W=0.9604874, p=0.3005428;"
## [18] "JPlag-text - CCR with W=0.9628915, p=0.3472413;"
## [19] "JPlag-text - AllnoNOR with W=0.9712717, p=0.554704;"
## [20] "JPlag-text - AllNOR with W=0.9712717, p=0.554704;"
## [21] "SIM-text - NoPPT with W=0.9618008, p=0.3253471;"
## [22] "SIM-text - RC with W=0.9781271, p=0.7587939;"
## [23] "Sherlock-text - AllNOR with W=0.9072469, p=0.01098593;"
## $`Non-normal distribution`
## [1] "SIM-text - CCR with W=0.8525712, p=0.0005750178;"
## [2] "SIM-text - AllnoNOR with W=0.790909, p=3.519414e-05;"
## [3] "SIM-text - AllNOR with W=0.790909, p=3.519414e-05;"
## [4] "Sherlock-text - NoPPT with W=0.6764819, p=5.32229e-07;"
## [5] "Sherlock-text - RC with W=0.8960805, p=0.005762239;"
## [6] "Sherlock-text - CCR with W=0.8812919, p=0.002538942;"
## [7] "Sherlock-text - AllnoNOR with W=0.8855633, p=0.003204142;"
```

Results of Shapiro-Wilk normality test for SOCO dataset D2 assignmenets group:

```

## $`Normal distribution`
## [1] "JPlag-java - NoPPT with W=0.9646739, p=0.3856066;"
## [2] "JPlag-java - RC with W=0.9646739, p=0.3856066;"
## [3] "JPlag-java - CCR with W=0.981933, p=0.8640068;"
## [4] "JPlag-java - AllnoNOR with W=0.9736222, p=0.6233963;"
## [5] "JPlag-java - AllNOR with W=0.9736222, p=0.6233963;"
## [6] "SIM-java - NoPPT with W=0.9684525, p=0.4775116;"
## [7] "SIM-java - RC with W=0.9703781, p=0.5295313;"
## [8] "SIM-java - CCR with W=0.9645588, p=0.3830315;"
## [9] "SIM-java - AllnoNOR with W=0.963926, p=0.3691145;"
## [10] "SIM-java - AllNOR with W=0.963926, p=0.3691145;"
## [11] "Sherlock-java - NoPPT with W=0.9622875, p=0.3349699;"
## [12] "Sherlock-java - RC with W=0.9397757, p=0.08134755;"
## [13] "Sherlock-java - CCR with W=0.9488192, p=0.1448137;"
## [14] "Sherlock-java - AllnoNOR with W=0.9316043, p=0.04848914;"
## [15] "Sherlock-java - AllNOR with W=0.9490796, p=0.1472336;"
## [16] "JPlag-text - NoPPT with W=0.968684, p=0.4835915;"
## [17] "JPlag-text - RC with W=0.9604874, p=0.3005428;"
## [18] "JPlag-text - CCR with W=0.9628915, p=0.3472413;"
## [19] "JPlag-text - AllnoNOR with W=0.9712717, p=0.554704;"
## [20] "JPlag-text - AllNOR with W=0.9712717, p=0.554704;"
## [21] "SIM-text - NoPPT with W=0.9618008, p=0.3253471;"
## [22] "SIM-text - RC with W=0.9781271, p=0.7587939;"
## [23] "Sherlock-text - AllNOR with W=0.9072469, p=0.01098593;"
## $`Non-normal distribution`
## [1] "SIM-text - CCR with W=0.8525712, p=0.0005750178;"
## [2] "SIM-text - AllnoNOR with W=0.790909, p=3.519414e-05;"
## [3] "SIM-text - AllNOR with W=0.790909, p=3.519414e-05;"
## [4] "Sherlock-text - NoPPT with W=0.6764819, p=5.32229e-07;"
## [5] "Sherlock-text - RC with W=0.8960805, p=0.005762239;"
## [6] "Sherlock-text - CCR with W=0.8812919, p=0.002538942;"
## [7] "Sherlock-text - AllnoNOR with W=0.8855633, p=0.003204142;"

```

MODEL COMPARISONS

The models that were compared are:

- NullModel

$$\begin{aligned} & \text{Imer}(F1 \sim (1|Participant) + (1|Tool : Participant) + (1|Technique : Participant), \\ & \quad data = SOCO.D_n, REML = FALSE) \end{aligned}$$

- ToolModel:

$$\begin{aligned} & \text{Imer}(F1 \sim Tool + \\ & \quad (1|Participant) + (1|Tool : Participant) + (1|Technique : Participant), \\ & \quad data = SOCO.D_n, REML = FALSE) \end{aligned}$$

- MainEffectsModel:

$$\begin{aligned} & \text{Imer}(F1 \sim Tool + Technique + \\ & \quad (1|Participant) + (1|Tool : Participant) + (1|Technique : Participant), \\ & \quad data = SOCO.D_n, REML = FALSE) \end{aligned}$$

- InteractionModel or FullModel:

$$\begin{aligned} & \text{Imer}(F1 \sim Tool + Technique + Tool : Technique + \\ & \quad (1|Participant) + (1|Tool : Participant) + (1|Technique : Participant), \\ & \quad data = SOCO.D_n, REML = FALSE) \end{aligned}$$

Table C.1: MLM comparison for SOCO D1

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)	p.boot
NullModel	5	-1421.5	-1397.3	715.7	-1431.5	NA	NA	NA	NA
ToolModel	10	-1988.1	-1939.7	1004.0	-2008.1	576.6	5	0.0000	0.0001
MainEffectsModel	14	-2033.9	-1966.2	1030.9	-2061.9	53.8	4	0.0000	0.0001
InteractionModel	34	-3361.6	-3197.2	1714.8	-3429.6	1367.8	20	0.0000	0.0001

CONTRAST EFFECT SIZES

Table D.1: Contrasts effect sizes for SOCO D1

ContrastName	EffectSize	CI.LB	CI.UB
(Intercept)	1.00	NA	NA
Tool.TextvsJava	0.89	0.88	0.91
TT.SHvsOthers	0.86	0.84	0.89
TJ.SHvsOthers	0.94	0.94	0.96
TT.SIMvsJPlag	0.95	0.95	0.96
TJ.SIMvsJPlag	0.80	0.78	0.84
NoPPTvsPPT	0.62	0.26	0.69
SinglevsCombo	0.05	0.00	0.19
RCvsCCR	0.70	0.33	0.75
AnoNvsAN	0.72	0.35	0.77
Tool.TextvsJava:NoPPTvsPPT	0.21	0.13	0.28
TT.SHvsOthers:NoPPTvsPPT	0.71	0.64	0.74
TJ.SHvsOthers:NoPPTvsPPT	0.10	0.02	0.18
TT.SIMvsJPlag:NoPPTvsPPT	0.34	0.26	0.41
TJ.SIMvsJPlag:NoPPTvsPPT	0.00	0.00	0.09
Tool.TextvsJava:SinglevsCombo	0.16	0.08	0.23
TT.SHvsOthers:SinglevsCombo	0.82	0.77	0.84
TJ.SHvsOthers:SinglevsCombo	0.15	0.07	0.22
TT.SIMvsJPlag:SinglevsCombo	0.47	0.38	0.52
TJ.SIMvsJPlag:SinglevsCombo	0.01	0.00	0.09
Tool.TextvsJava:RCvsCCR	0.59	0.52	0.64
TT.SHvsOthers:RCvsCCR	0.61	0.53	0.65
TJ.SHvsOthers:RCvsCCR	0.32	0.24	0.38
TT.SIMvsJPlag:RCvsCCR	0.19	0.11	0.26
TJ.SIMvsJPlag:RCvsCCR	0.00	0.00	0.09
Tool.TextvsJava:AnoNvsAN	0.51	0.43	0.56
TT.SHvsOthers:AnoNvsAN	0.73	0.67	0.77
TJ.SHvsOthers:AnoNvsAN	0.11	0.03	0.19
TT.SIMvsJPlag:AnoNvsAN	0.00	0.00	0.09
TJ.SIMvsJPlag:AnoNvsAN	0.00	0.00	0.09

Note:

TT - ToolText, TJ - ToolJava, SH - Sherlock, AnoN - AllnoNOR,
AN - AllNOR

Table D.2: Simple effect analysis effect sizes for SOCO D1

ContrastName	EffectSize	CI.LB	CI.UB
(Intercept)	1.00	NA	NA
TextvsJava	0.89	0.88	0.91
TT.SHvsOthers	0.86	0.84	0.89
TT.SIMvsJPlag	0.95	0.95	0.96
TJ.SHvsOthers	0.94	0.94	0.96
TJ.SIMvsJPlag	0.80	0.78	0.84
TT.SH.NoPPTvsPPT	0.47	0.41	0.53
TT.SH.SinglevsCombo	0.74	0.71	0.78
TT.SH.RCvsCCR	0.13	0.06	0.20
TT.SH.AllnoNORvsAllNOR	0.77	0.74	0.80
TT.JPlag.NoPPTvsPPT	0.34	0.27	0.40
TT.JPlag.SinglevsCombo	0.13	0.06	0.21
TT.JPlag.RCvsCCR	0.51	0.46	0.57
TT.JPlag.AllnoNORvsAllNOR	0.00	0.00	0.08
TT.SIM.NoPPTvsPPT	0.64	0.59	0.68
TT.SIM.SinglevsCombo	0.63	0.59	0.68
TT.SIM.RCvsCCR	0.65	0.61	0.69
TT.SIM.AllnoNORvsAllNOR	0.00	0.00	0.08
TJ.SH.NoPPTvsPPT	0.13	0.06	0.20
TJ.SH.SinglevsCombo	0.18	0.11	0.25
TJ.SH.RCvsCCR	0.34	0.27	0.40
TJ.SH.AllnoNORvsAllNOR	0.13	0.06	0.20
TJ.JPlag.NoPPTvsPPT	0.02	0.00	0.09
TJ.JPlag.SinglevsCombo	0.02	0.00	0.09
TJ.JPlag.RCvsCCR	0.02	0.00	0.09
TJ.JPlag.AllnoNORvsAllNOR	0.00	0.00	0.08
TJ.SIM.NoPPTvsPPT	0.01	0.00	0.09
TJ.SIM.SinglevsCombo	0.01	0.00	0.09
TJ.SIM.RCvsCCR	0.02	0.00	0.09
TJ.SIM.AllnoNORvsAllNOR	0.00	0.00	0.09

Note:

TT - ToolText, TJ - ToolJava, SH - Sherlock

CHAPTER E

INTERACTION GRAPHS

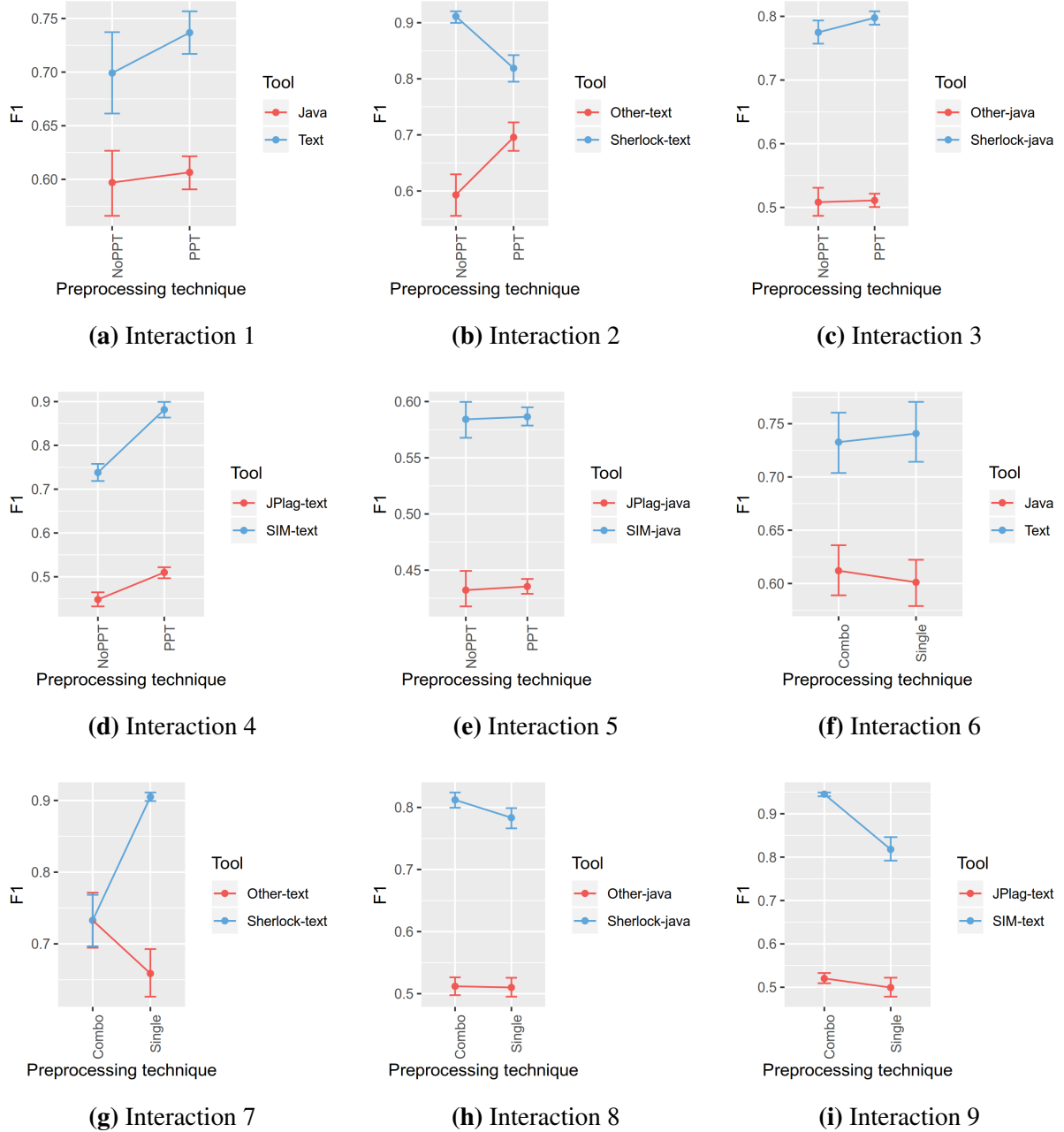


Figure E.1: Interaction graphs for SOCO D1 - part 1

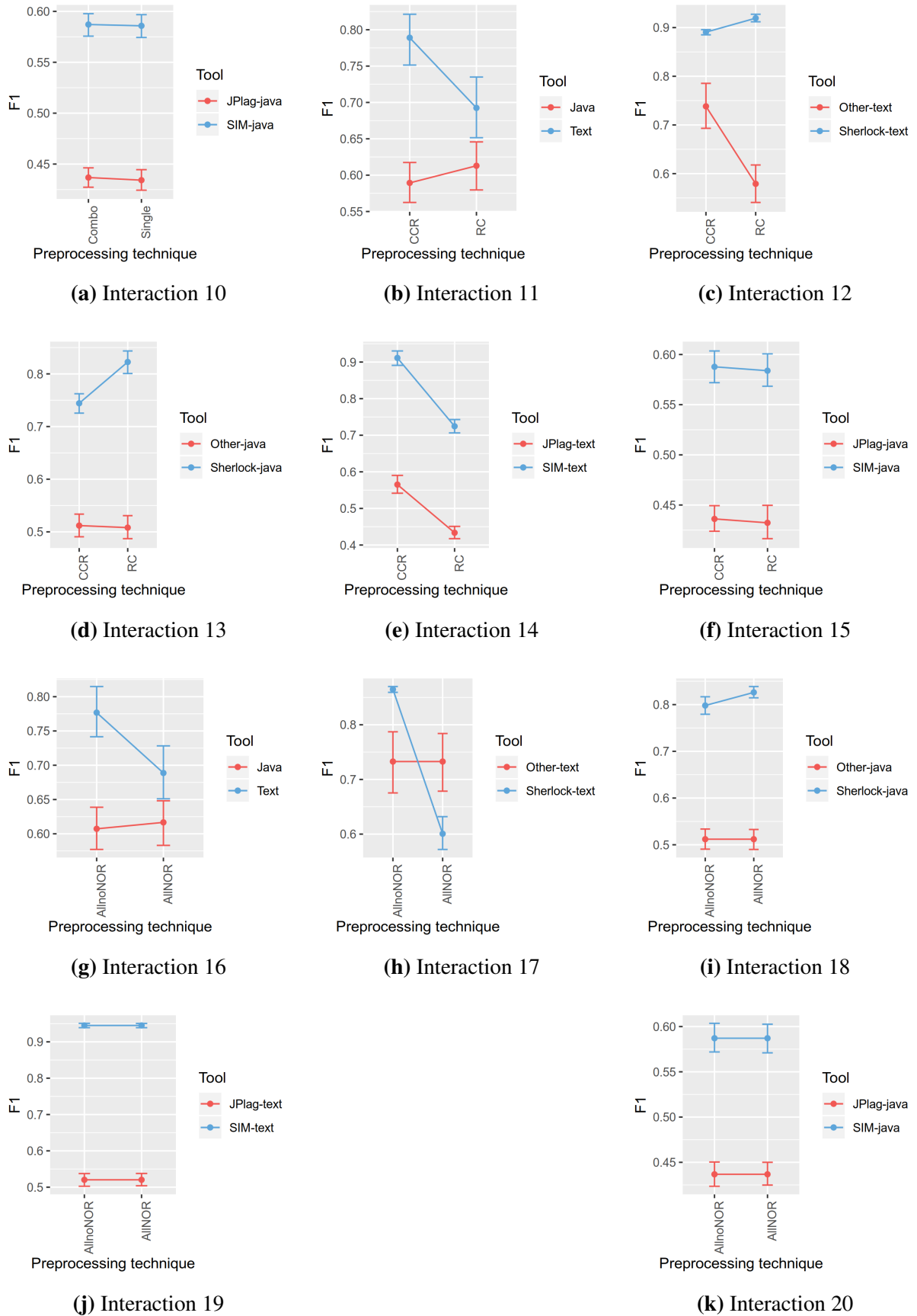


Figure E.2: Interaction graphs for SOCO D1 - part 2

BIBLIOGRAPHY

- [1] E. L. Deci and R. M. Ryan, “A motivational approach to self: Integration in personality”, *Perspectives on motivation*, vol. 38, no. 237, pp. 237–288, 1990.
- [2] D. O. Hebb and W. Thompson, “The social significance of animal studies”, in *G. Lindzey & E. Aronson, Handbook of social psychology*, 3rd ed, New York: Random House, 1985, pp. 729–774, ISBN: 978-0-8985-9720-2.
- [3] R. J. Leach, “Using metrics to evaluate student programs”, *ACM SIGCSE Bulletin*, vol. 27, no. 2, pp. 41–43, 1995, ISSN: 0097-8418. DOI: 10.1145/201998.202010.
- [5] K. J. Ottenstein, “An algorithmic approach to the detection and prevention of plagiarism”, *ACM SIGCSE Bulletin*, vol. 8, no. 4, pp. 30–41, 1976, ISSN: 0097-8418. DOI: 10.1145/382222.382462.
- [6] S. Saxon, “Comparison of plagiarism detection techniques applied to student code”, Part II . Computer Science project, Trinity College, Cambridge, UK, Tech. Rep., 2000.

CURICULUM VITE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

...

Bibliography

1. M. Novak, M. Novak, and M. Novak, "Using", *International Journal*, vol. 1, no. 4, pp. 147–171, 2018, ISSN: 21-321. DOI: 10.4018/101010101

2. D. O. Hebb and W. Thompson, "The social significance of animal studies", in *G. Lindzey & E. Aronson, Handbook of social psychology*, 3rd ed, New York: Random House, 1985, pp. 729–774, ISBN: 978-0-8985-9720-2