UNIVERSITY OF MELBOURNCE

COMP90019 RESEARCH PROJECT

25 CREDITS

De-identification and Colander method

*Author: Supervisor:*

Ao Li Karin Verspoor

*A thesis submitted in fulfilment of the requirements*

*for the degree of Master of Information Technology*

Student ID: 798657

Department of Computing and Information System

May 30, 2018

**Abstract**

**Key words:**

Declaration

I certify that

- this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person where due reference is not made in the text.

- where necessary I have received clearance for this research from the University’s Ethics Committee and have submitted all required data to the Department.

- the thesis is \_\_\_\_\_ words in length (excluding text in images, table, bibliographies and appendices).

Acknowledgements

Thanks to my kindly supervisor Professor Karin Verspoor for her professional assistance, insightful guidance and patient support through my project. She offers the great help to me when I meet some challenges.

Thanks to my dear parents for their fully understand which keeps me positive.

Thanks to my lovely friends for their encouragement which makes me strong.

Contents

[List of Figures v](#_Toc509221317)

[List of Tables vi](#_Toc509221318)

[1 Introduction vii](#_Toc509221319)

[2 Project Description viii](#_Toc509221320)

[3 Performance Analysis ix](#_Toc509221321)

[4 Conclusion x](#_Toc509221322)

# List of Figures

# List of Tables

# 1 Introduction

1.1 Background

1.2 Overview of Pre-processing Approaches

1.2.1 De-identification Approach

1.2.2 Colander Approach

# 2 Project Description

2.1 Experiment Settings

2.2 Main Characteristics of Selected Approaches

# 3 Performance Analysis

3.1 De-identification Approach Analysis

3.1.1 External Dictionary Performance Analysis

3.1.2 Comparison between External Dictionary and Online Open Source Dictionary

3.2 Colander Approach Analysis

# 4 Conclusion

# Reference