FLORIDA POLYTECHNIC UNIVERSITY

THESIS TITLE GOES HERE

by

Levi C. Nicklas

A Thesis Submitted to the Faculty of the

DEPARTMENT OF COMPUTER SCIENCE

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

Florida Polytechnic University

2021

THESIS TITLE GOES HERE

by

Levi C. Nicklas

A Thesis Submitted to the Faculty of the

DEPARTMENT OF COMPUTER SCIENCE

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

Florida Polytechnic University

2021

Approved by:

	Signature	**Date**
Dr. Reinaldo Sanchez-Arias (Committee Chair, Advisor)	_____	_____
Dr. Grisselle Centeno (Committee Member)	_____	_____
Dr. Harish Chintakunta (Committee Member)	_____	_____
Dr. Tom Dvorske Vice Provost of Academic Affairs (Graduate Division)	_____	_____

Dedications go here

ACKNOWLEDGMENTS

Acknowledgments go here

Contents

List of Figures ii

List of Tables iii

List of Algorithms iv

Abstract 1

1 Introduction 1

2 Literature Review 3

2.1 Introduction 3

2.1.1 Subsection Testing 3

3 Research work 4

3.1 Introduction 4

4 Discussion of results 5

5 Conclusion 6

6 Future Work 7

Appendices 8

A My first appendix 9

List of Figures

List of Tables

List of Algorithms

Abstract

Abstract goes here

Chapter 1: Introduction

Introduction Chapter Here

Text mining is the process of extracting insight from text documents using computational and statistical methods. A common task in text mining is document clustering, where the natural language of the documents is reformatted and used to group documents into clusters of similar topics or text content. Popular methods to cluster text include k-means, hierarchical methods, and topic models. Most of these methods utilize term frequency-inverse document frequency measures or bag-of-words methods to do clustering, though these methods often divide words and thus may lose meaning, or context, surrounding the word of concern. In an effort to cluster documents while still preserving the relationship between words, text can be modeled with graphs which better preserve the context surrounding a word. Text can be represented with bigram graphs, where the edge between two vertices is the result of two words appearing adjacent in the text. This idea can be extended to more general n-grams, and further extended to skip-n-grams, resulting in even richer representations of the text. The similarity of two of graphical representations can then be assessed using modern methods called graph kernels. Graph kernels produce a measure of similarity between graphs, thus allowing for further machine learning to take place. With a measure of similarity between two graphs, we can do an assortment of clustering methods.

Chapter 2: Literature Review

Literature review chapter goes here

2.1 Introduction

Introduction section of this chapter goes here

2.1.1 Subsection Testing

subection content goes here

Chapter 3: Research work

Contents of research work go here.

3.1 Introduction

Content of the introduction section

Rest of sections go here

Chapter 4: Discussion of results

Discussions and results go here

Chapter 5: Conclusion

Conclusion goes here.

Chapter 6: Future Work

Future work goes here.

Appendices

Appendix A: My first appendix