AN EFFICIENT SIMILARITY BASED SEARCH ENGINE FOR MATHEMATICAL CONTENT IN LaTeX MARKUP

by

Wei Zhong

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Electrical and Computer Engineering

Spring 2015

AN EFFICIENT SIMILARITY BASED SEARCH ENGINE FOR MATHEMATICAL CONTENT IN LaTeX MARKUP

by

Wei Zhong

Approved: _____

Fouad E. Kiamilev, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Kenneth E. Barner, Ph.D.
Chair of the Department of Electrical and Computer Engineering

Approved: _____

Babatunde A. Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: _____

James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

Thank you to my family who supports me the most from every perspective through out my graduate academic education. Thank you to my advisor Hui Fang who offers me the opportunity to develop my idea further and supports me in many other ways. I am also grateful to all InfoLab members for their kind help. And to my instructors or friends, not previously mentioned, who have influenced me or helped along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The structured sense of mathematical language, as well as many its semantic properties, makes general text retrieval models (e.g. bag of words model) deficient to provide good search result.

In this paper, we have addressed the problems of measuring the similarity degree between mathematical formulae and proposed an approach to parse mathematical content (particularly in LaTeX markup) into a tree representation, index syntactic information from math formulae, and a search method to capture some level of query-document subgraph isomorphism, combined with a pruning method to speed search. We build our own proof-of-concept prototype search engine to demonstrate the ideas mentioned in this paper, and thus are able to present some evaluation results. Through this paper, we have showed another possibility to tackle a number of the problems in math-aware search.

Chapter 1

INTRODUCTION

This is the information for the first chapter, Chapter 1. Copy the base file, chap1.tex, for additional chapter needed such as chap2.tex, chap3.tex, etc. Modify the main base file to include each additional chapter file.

1.1 intro

This is the information for the first section of the first chapter.

1.1.1 intro2

This is the information for the first section of the first chapter.

1.1.1.1 intro3

This is the information for the first section of the first chapter.

1.1.2 intro4

This is the information for the first section of the first chapter.

intro5

This is the information for the first section of the first chapter.

intro6

This is the information for the first section of the first chapter.

All elements	
1^{st}	2^{nd}
1	2
3	4

Table 1.1: Table Example

One of my chapter headings[1] is [2] at the bottom of a page and I'd like to drop it down to start off the next page. Any suggestions would be much appreciated.

$$a + \pi^2$$

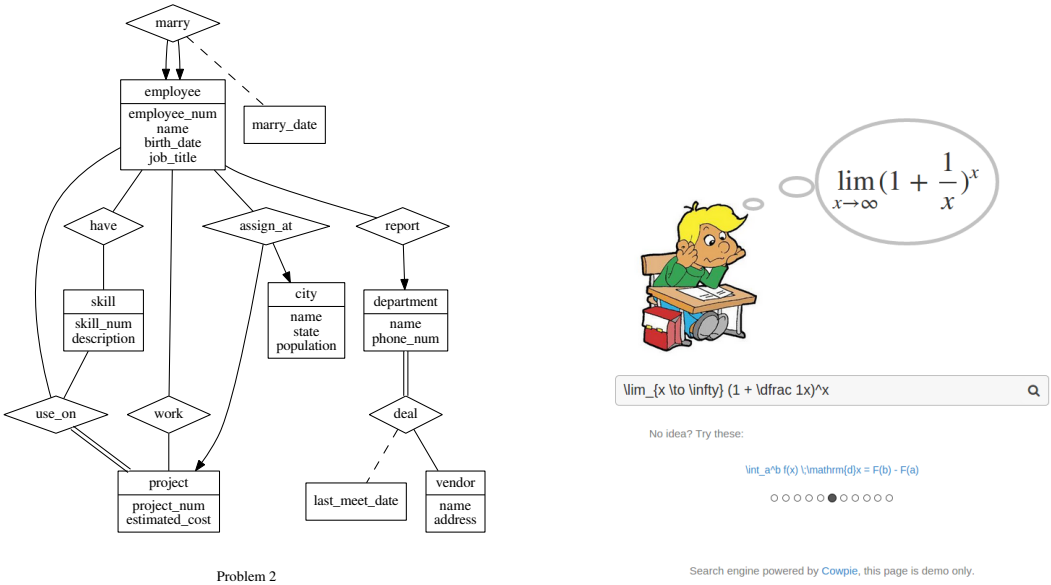

Figure 1.1: Figure Example

BIBLIOGRAPHY

[1] Author and Author. *Title of Book*. Publisher.

[2] Author. Title of article. *Title of Journal*, month.

Appendix A

TITLE OF APPENDIX A

This is the information for the first appendix, Appendix A. Copy the base file, appA.tex, for each additional appendix needed such as appB.tex, appC.tex, etc. Modify the main base file to include each additional appendix file.

If there is only one appendix, then modify the main file to only use app.tex instead of appA.tex.

Appendix B

TITLE OF APPENDIX B