

University of Sheffield

Size Matters: Acquiring Vague Spatial Size Information from Textual Sources



Alexander White

Supervisor: Dr Robert Gaizauskas

A report submitted in partial fulfilment of the requirements
for the degree of Computer Science with a Year in Industry BSc in Computer Science

in the

Department of Computer Science

November 21, 2019

Declaration

All sentences or passages quoted in this document from other people's work have been specifically acknowledged by clear cross-referencing to author, work and page(s). Any illustrations that are not the work of the author of this report have been used with the explicit permission of the originator and are specifically acknowledged. I understand that failure to do this amounts to plagiarism and will be considered grounds for failure.

Name: Alexander White

Signature:

Date:

Abstract

One or two sentences providing a basic introduction to the field, comprehensible to a scientist in any discipline. Two to three sentences of more detailed background, comprehensible to scientists in related disciplines. One sentence clearly stating the general problem being addressed by this particular study. One sentence summarising the main result (with the words “here I show” or their equivalent). Two or three sentences explaining what the main result reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge. One or two sentences to put the results into a more general context. Two or three sentences to provide a broader perspective, readily comprehensible to a scientist in any discipline.

Contents

1	Introduction	1
1.1	Aims and Objectives	1
1.2	Overview of the Report	1
2	Literature Survey	3
2.1	A Section that Contains Some References	3
2.2	A Section that Contains Some Maths	3
2.3	A Section that Contains a Figure	3
2.4	A Section that Contains a Table	3
2.5	Summary	3
3	Methodology	5
3.1	Requirements	5
3.2	Analysis	5
3.3	Design	5
3.4	Implementation	5
4	Results	6
4.1	Experiment 1	6
4.2	Experiment 2	6
5	Discussion	7
6	Conclusions	8
	Appendices	10
A	An Appendix of Some Kind	11
B	Another Appendix	12

List of Figures

2.1	A simple figure in \LaTeX . Reproduced from http://tinyurl.com/nqtrlj5 with the permission of the copyright owner.	4
-----	--	---

List of Tables

2.1	A simple table in L ^A T _E X.	3
-----	--	---

Chapter 1

Introduction

Why its a problem

This disseration project takes on the problem of determining the general sizes of different objects. Humans are talented at estimating sizes of objects based on common sense or memory and this skill helps us to determine distances of objects Talk about scraping data from web pages - where are we getting it from?

How effective this method will be and its limitations

1.1 Aims and Objectives

The aim of this dissertation is to make progress towards creating a database containing information about objects and their usual sizes. This can be broken down into three stages. Stage one is fulfilling our requirement of training data containing various objects and sizes. The aim is to train machine learning models to be able to identify objects and sizes within a sentence and determine if they are related. To collect enough data to adequately train these models we will use semi-supervised learning, which means that this stage will run simultaneously alongside stage two.

Stage two is building the named entity recogniton and relationship extraction models for both identifying objects and sizes in text, and also determining if they are related. To help with the semi-supervised learning we can build some very basic models to start collecting data. Using regular expressions to determine sizes, and part-of-speech tagging to determine nouns, we can build models with poor accuracy but that will help in collecting data that can be refined into the final training set.

The final stage of the project will be to collect all the results into a database. The accuracy of the collected data can be improved if objects have been found multiple times. We can look at previously found sizes of the object to determine if this new measurement is accurate. If we introduce an object hierarchy to determine if objects are related then we can also use similar objects to estimate realistic sizes.

1.2 Overview of the Report

This dissertation will begin with a literature survey.

This will be followed by the Methodology stage

This will be followed by the Implementation stage
The results

Chapter 2

Literature Survey

2.1 A Section that Contains Some References

According to seminal research by Hawthorn et al. (2001), lorem ipsum dolor sit amet. However, this result was already known in the 1990s (Wieman & Hollberg 1991, Arnold et al. 1998).

2.2 A Section that Contains Some Maths

$$M = \frac{1}{T} \sum_{t=1}^T e(t) / \max_t [e(t)] \quad (2.1)$$

This is shown in Equation 2.1 and is repeated here $M = \frac{1}{T} \sum_{t=1}^T e(t) / \max_t [e(t)]$.

2.3 A Section that Contains a Figure

See Figure 2.1.

2.4 A Section that Contains a Table

A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

Table 2.1: *A simple table in L^AT_EX.*

This is shown in Table 2.1.

2.5 Summary

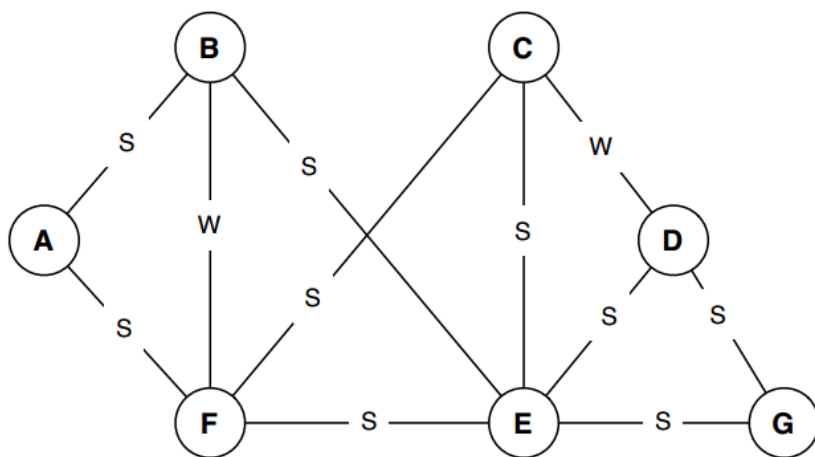


Figure 2.1: A simple figure in \LaTeX . Reproduced from <http://tinyurl.com/nqtrlj5> with the permission of the copyright owner.

Chapter 3

Methodology

3.1 Requirements

3.2 Analysis

3.3 Design

3.4 Implementation

Chapter 4

Results

4.1 Experiment 1

4.2 Experiment 2

Chapter 5

Discussion

Chapter 6

Conclusions

Bibliography

Arnold, A. S., Wilson, J. S. & Boshier, M. G. (1998), ‘A simple extended-cavity diode laser’, *Review of Scientific Instruments* **69**(3), 1236–1239.

URL: <http://link.aip.org/link/?RSI/69/1236/1>

Hawthorn, C. J., Weber, K. P. & Scholten, R. E. (2001), ‘Littrow configuration tunable external cavity diode laser with fixed direction output beam’, *Review of Scientific Instruments* **72**(12), 4477–4479.

URL: <http://link.aip.org/link/?RSI/72/4477/1>

Wieman, C. E. & Hollberg, L. (1991), ‘Using diode lasers for atomic physics’, *Review of Scientific Instruments* **62**(1), 1–20.

URL: <http://link.aip.org/link/?RSI/62/1/1>

Appendices

Appendix A

An Appendix of Some Kind

Appendix B

Another Appendix