

UNIVERSITY OF SOUTHAMPTON

# **A Context-Sensitive Relevance-Based Intelligent Data-Ranking Agent**

by

Thomas J. Bell

A project report submitted for the award of  
MEng Electronic Engineering

in the  
School of Electronics and Computer Science

October 2013

*“Write a funny quote here.”*

If the quote is taken from someone, their name goes here

UNIVERSITY OF SOUTHAMPTON

# *Abstract*

School of Electronics and Computer Science

MEng Electronic Engineering

by Thomas J. Bell

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

# *Acknowledgements*

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>vii</b>
<b>Abbreviations</b>	<b>viii</b>
<b>Physical Constants</b>	<b>ix</b>
<b>Symbols</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Problem . . . . .	1
1.2 Project Objective . . . . .	1
1.3 Goals . . . . .	1
1.4 Unique Features . . . . .	2
<b>2 Background Research</b>	<b>3</b>
2.1 Existing Data-Ranking Implementations . . . . .	3
2.1.1 Google Now . . . . .	3
2.2 Text analytics services . . . . .	3
Viralheat . . . . .	3
AlchemyAPI . . . . .	4
Semantria . . . . .	4
Saplo . . . . .	4
Wingify . . . . .	4
DatumBox . . . . .	4
2.3 Data-mining . . . . .	4
2.4 Topic Analysis . . . . .	4
2.5 Context-Sensitive Scoring Algorithms . . . . .	5
2.6 Ranking Algorithms . . . . .	5

---

<b>3</b>	<b>Specification</b>	<b>6</b>
3.1	A Section . . . . .	6
3.1.1	A Subsection . . . . .	6
3.2	Another Section . . . . .	6
<b>4</b>	<b>Design Detail</b>	<b>7</b>
4.1	Data Acquisition . . . . .	7
4.2	Scoring Algorithm . . . . .	7
4.3	Ranking Algorithm . . . . .	7
<b>5</b>	<b>Planning and Progress</b>	<b>8</b>
5.1	A Section . . . . .	8
5.1.1	A Subsection . . . . .	8
5.2	Another Section . . . . .	8
<b>6</b>	<b>Testing Strategy and Results</b>	<b>9</b>
6.1	A Section . . . . .	9
6.1.1	A Subsection . . . . .	9
6.2	Another Section . . . . .	9
<b>7</b>	<b>Critical Evaluation</b>	<b>10</b>
7.1	A Section . . . . .	10
7.1.1	A Subsection . . . . .	10
7.2	Another Section . . . . .	10
<b>8</b>	<b>Conclusion</b>	<b>11</b>
8.1	A Section . . . . .	11
8.1.1	A Subsection . . . . .	11
8.2	Another Section . . . . .	11
<b>9</b>	<b>Further Work</b>	<b>12</b>
9.1	A Section . . . . .	12
9.1.1	A Subsection . . . . .	12
9.2	Another Section . . . . .	12
<b>A</b>	<b>An Appendix</b>	<b>13</b>
	<b>Bibliography</b>	<b>14</b>

# List of Figures

# List of Tables



# Abbreviations

**LAH** List Abbreviations **Here**

# Physical Constants

Speed of Light  $c = 2.997\,924\,58 \times 10^8 \text{ ms}^{-\text{s}}$  (exact)

# Symbols

$a$	distance	m
$P$	power	W ( $\text{Js}^{-1}$ )
$\omega$	angular frequency	$\text{rads}^{-1}$

*For/Dedicated to/To my...*

# Chapter 1

## Introduction

Context-sensitive techniques, data fusion and ranking agents are frequently used in industry applications and consumer software products. Combined, these techniques are useful for ranking social media according to a users preferences and context. This report describes an agent for the classifying, scoring and ranking of data according to its context-sensitive relevance to a user.

### 1.1 The Problem

Social media, productivity tools and internet-based information are abundant on mobile devices, leading to users being overwhelmed with information, despite only a small amount of it being of any interest to a particular individual at any given moment. This calls for a means by which such data can be ranked or filtered according to its importance, interest or relevance.

### 1.2 Project Objective

The objective of this project is to produce a scalable and highly modular context-sensitive mobile-content relevance-based intelligent ranking agent, to order social media, productivity and other web-based information according to its time- and situation-specific relevance to a user.

### 1.3 Goals

The following are core goals which this project sets out to achieve.

1. Develop a scoring algorithm by which to judge the relevance of an item of data based upon a users
  - (a) Personality profile
  - (b) Historical data such as click-history
  - (c) Environment (conditional upon time-constraints)
2. To perform automatic remote topic analysis to judge the topic of an item of data
3. Develop a sorting/ranking algorithm to sort or insert scored items of data efficiently, into an ordered list
4. Develop a stable and robust data fusion technique to combine a range of data into a user-context and data-context object.
5. Abstract away this agent into an extensible Java API for use in
  - (a) Smartphone apps (Android)
  - (b) Web-apps (Spring MVC)
  - (c) Desktop applications (Java Swing etc.)
6. To develop a consumer smart phone app to demonstrate the working API which automatically ranks a users data according to its relevance

These are the criteria by which the extent of this project's success will be evaluated.

## 1.4 Unique Features

Many of the aspects of this project have never been seen combined into a single research project before and others (such as query-less context-sensitive scoring of data) have recieved little attention. This project is unique in its endeavour to combine data fusion techniques with context-sensitive scoring in the development of a commercially viable prototype.

## Chapter 2

# Background Research

This is some sample text.

### 2.1 Existing Data-Ranking Implementations

This is some sample text.

**Google Now** Google Now is a mobile app which combines Google's search feature with useful information which is deemed relevant to the user's environment such as weather, a map to get them home after a night out or nearby events.

**Viralheat** Viralheat is a web-based social media content aggregator and filter, used for commercial uses of social media. It allows the user to filter content from twitter, Facebook and others according to its sentiment (positive/negative). It's not available as a non-commercial social media aggregator and does not perform topic analysis for ranking.

**StreamLife** This app aggregates facebook content and tweets, but performs no topic/sentiment analysis or ranking, and provides no capability for including tasks, calendar appointments, SMS messages or emails.

## 2.2 Text analytics services

There are a good number of existing text analytics services available to the end user and the developer for a range of different types of analytics. These services include sentiment analysis, text categorisation, contextual targeting and a range of others. This section highlights some which are relevant to this study.

**AlchemyAPI** Alchemy provides an API which performs entity extraction, sentiment analysis, concept tagging, relation extraction and most notably, text categorisation (among others). It provides a free licence for up to 1,000 API calls per day. It provides all the core features which this project requires in terms of text analytics, however the text categorisation did not produce strong results for short amounts of text (such as tweets) and often failed to make any categorisation.

**Semantria** The best solution for sentiment analysis appears to be semantria. It gave consistently precise and accurate sentiment analysis for text containing more than 5 words.

**Saplo** Saplo was distinguished in its contextual analysis feature which allowed me to define a personalised textual context which could be matched against any type of text. This could have allowed me to define user-specific textual contexts against which to match data items, however it only allows users 2000 API calls per month on their free account.

**Wingify** This is a beta-stage contextual targeting API which can categorise text from a web page and extract key concepts. The online demo provided accurate results, yet there is not yet a public API available.

**DatumBox** DatumBox is a free machine learning API which performs sentiment analysis, subjectivity analysis, topic classification, language detection, readability detection, educational detection, document similarity analysis, and gender detection. Many of these features may be useful for ranking text based upon its relevance to a particular individual. It has a simple API using http POST requests and a JSON response.



## **2.3 Data-mining**

Outline: Facebook and Twitter API (phone, web and desktop), Android API, Android Calendar, Android Tasks, Android SMS, Android Sensors, Google Calendar and Tasks (web-based and desktop API).

## **2.4 Topic Analysis**

This is some sample text.

## **2.5 Context-Sensitive Scoring Algorithms**

This is some sample text.

## **2.6 Ranking Algorithms**

Insertion sort etc.

## Chapter 3

# Specification

This is some sample text.

### 3.1 A Section

This is some sample text.

#### 3.1.1 A Subsection

This is some sample text.

### 3.2 Another Section

This is some sample text.

## Chapter 4

# Design Detail

This is some sample text.

### 4.1 Data Acquisition

This is some sample text.

#### 4.1.1 Data Item Acquisition

#### 4.1.2 User Data Acquisition

### 4.2 Classification

TODO: Discuss lexicon based vs. learning based techniques - lexicon requires dictionary, but learning requires training.

### 4.3 Scoring Algorithm

This is some sample text.

#### **4.3.1 Data Item Scoring**

#### **4.3.2 User Data Scoring**

### **4.4 Ranking Algorithm**

This is some sample text.

## Chapter 5

# Planning and Progress

This is some sample text.

### 5.1 A Section

This is some sample text.

#### 5.1.1 A Subsection

This is some sample text.

### 5.2 Another Section

This is some sample text.

## Chapter 6

# Testing Strategy and Results

This is some sample text.

### 6.1 A Section

This is some sample text.

#### 6.1.1 A Subsection

This is some sample text.

### 6.2 Another Section

This is some sample text.

## **Chapter 7**

# **Critical Evaluation**

This is some sample text.

### **7.1 A Section**

This is some sample text.

#### **7.1.1 A Subsection**

This is some sample text.

### **7.2 Another Section**

This is some sample text.

## Chapter 8

# Conclusion

This is some sample text.

### 8.1 A Section

This is some sample text.

#### 8.1.1 A Subsection

This is some sample text.

### 8.2 Another Section

This is some sample text.



## **Chapter 9**

# **Further Work**

This is some sample text.

### **9.1 A Section**

This is some sample text.

#### **9.1.1 A Subsection**

This is some sample text.

### **9.2 Another Section**

This is some sample text.

## Appendix A

# An Appendix

This is an appendix.

# Bibliography