

Electronics and Computer Science
Faculty of Physical Sciences and Engineering
University of Southampton

Jamie Davies
November 19, 2013

Enhanced Content Analysis and Information
Retrieval Using Twitter Hashtags

A project report submitted for the award of
MEng Computer Science

Supervisor:
Dr. Nick Gibbins

Examiner:
Dr. Otto Octavius

Abstract

Microblogging service such as Twitter have become very popular in recent years. One of the key characteristics of such services is the use of ‘hashtags’ — topical/categorical annotations provided by the authors of the posts (tweets) themselves. There is a vast and diverse collection of hashtags for the users to choose from, resulting in the information provided by hashtags not being as accurate or as complete as it could be.

This project aims to create a system to support and enrich the information provided by hashtags. It will examine and classify the topics and concepts behind the hashtags, and in doing so, be able to suggest suitable new hashtags for tweets that are relevant to their content, allowing users to make a better choice of hashtags when writing their tweets. Furthermore, this system will be extended by enabling a user to search for a particular hashtag, and instead of only returning tweets containing that hashtag (as current systems do), it will also provide tweets that are contextually relevant to the search term but do not contain that given hashtag.

Contents

Abstract	1
1 Project Goals	3
2 Background and report of literature search	4
3 Report on Technical Progress	5
4 Plan of remaining work	6
References	7

1 Project Goals

12pt text for main body.

2 Background and report of literature search

12pt text for main body.

3 Report on Technical Progress

12pt text for main body.

4 Plan of remaining work

12pt text for main body.

References