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

Jamie Davies
November 21, 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

One of the key characteristics of Twitter and other microblogging platforms is the use of 'hashtags' — topical/categorical annotations provided by the authors of the posts (tweets) themselves. Hashtags are any combination of letters and digits preceded by a hash (#) symbol allowing users to use whichever hashtags they like when publishing their tweets, including creating their own new hashtags. Whilst this flexible system gives users the freedom to express themselves, it unfortunately also results in data from the hashtags being fragmented and inaccurate due to the vast and diverse possibilities of hashtags.

If users are presented with a choice of relevant hashtags when writing a tweet, they are more likely to publish tweets with accurate tag data. This project aims to create an intelligent hashtag suggestion tool to raise the information gain from hashtags. However, whilst such a system could improve the quality of the hashtag data for future tweets, tweets that have already been published will remain untouched by the system. Thus, the system will be extended to also retrofit hashtags to published tweets — allowing for tweets to appear in search results for a particular hashtag even if they don't actually contain the hashtag in question.

Contents

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