Trending topics using the Twitter-API

Marco Pock-Steen Fraile
Christian Palmhøj Nielsen
IT University of Copenhagen

December 4, 2012

Abstract

We have designed an algorithm for identifying trending topics in twitter..

1 Introduction

We wish to make use of the Twitter streaming API to implement our own version of trending topics using the Misra-Gries algorithm, and for the trending topics we would track how many unique tweets have been made per topic, using one of the algorithms for counting distinct elements. We could also count number of tweets for trending topics to compare actual tweets with unique tweets, but this is not algorithmically a challenge.

Challenges would include finding a suitable k for the Misra-Gries that is a good trade-off between memory space and making sure we take into account the tweets inbetween the actual trending topics that might "flood" them out of the array.

For the counting algorithm we similarly need to find a good hashing algorithm for tweets, and based on this find a good number for kth-minimum distances that gives us a realistic count for unique tweets.

Outline   The remainder of this article is organized as follows. Section 3 gives account of previous work. Our new and exciting results are described in Section 5. Finally, Section 6 gives the conclusions.

2   Twitter topics

3   Streaming algorithms

4   Algorithm design

4.1   Design

4.2   Running time

4.3   Satisfaction

5   Results

In this section we describe the results.

6   Conclusions

We worked hard, and achieved very little.