Automatic Extractive Twitter Hashtag Summarization

Hengodage, M.A. and Samankula, W.G.D.M.

modithaon@gmail.com, dileepa@northshore.edu.lk

During past years, social media content has been widely used for many purposes, fields. Large amounts of content are generated in these social media websites such as Facebook, Twitter, and LinkedIn in every day. Twitter has large sets of data which is limited to 140 words, unlike any other social media platforms. Twitter allows a huge number of users to contribute frequent short messages. The content is an extraordinarily large number of small textual messages that are posted by millions of users. Twitter has topics system which enables the user to check about public opinion using with particular hashtags. Twitter websites' content was focused in this research.

This trending topic changes daily, even it changes based on locations. Each country has a different set of trends. But the major problem is for recognizing these hashtags for the users who are not very active on twitter websites. Some hashtags self-explain themselves with the words being used but not all of them. This project gives the ability to summarize idea about particular hashtag.

A hashtag summarizer was implemented in order to address this problem. Creating hashtag summarizer may give the foundation to address previously mentioned issues about hashtags. This implemented system has some additional features other than hashtag summarization, for an instance user may use it to quickly check updates about some live event going on with its hashtag keyword. This solution provide some options to fine tune the results of summarization.

In the summarisation process, all the related posts for given hashtag are considered. Then emoji's and non-ascii characters are removed from the imported text data. The data is sent to the summarizer after cleaning up the characters. Then summarizer associates the words with their grammartical counterpart. Considering the frequency, each word is ranked with points system. Furthermore, by end of the sentences taking into the account, summariser splits the text into sentences. At the end of the process, sentences are ranked by sum of their words' points. Then, it returns the highly ranked sentences.

Keywords: *Automatic, Twitter, Hashtag summarizer, Social media*