Abstract

Millions of Internet users use social sites for sharing and storing data, blogs for connecting with people and sharing information. Twitter is one of the fastest growing social sites. Short-text messages are being posted and shared at a unique rate. Twitter collects millions of tweets, which contain lots of noise and redundancy, un-structured tweets. As redundancy leads to inconsistency and less accurate result, users are unable to understand current topics of discussion. Due to time constraints readers are unable to read each and every tweet, so user requires summary to understand important information on social media. To generate summary for large amount of data, a new summarization method is proposed, namely sequential summarization, which provides a topic detection of social media and generate ordered short
sub-summaries for a trending topic in order to convey the important information in few sentences. The system will implement two approaches, stream-based and semantic-based, for detecting the necessary and non-redundant subtopics within trending information.

References

- Dehong Gao, Wenjie Li, Xiaoyan Cai, Renxian Zhang and You Ouyang. "Sequential Summarization of Twitter Trending Topics" IEEE/ACM Trans on Audio, speech and language processing, vol. 22 No 2; Feb 2014

- S. Petrovic, M. Osborne, and V. Lavrenko. "Streaming first story detection with application to Twitter" in Proc. ACL-10, 2010
- Gulab R. Shaikh, Digambar M. Padulkar. "Template Based Abstractive Summarization of Twitter Topic with Speech Act" by Asst. Prof. , Department of CSE, VPCOE Baramati, Pune, India, India in June 2014.

Index Terms

Computer Science

Information Sciences
Keywords
Topic Detection  Social Media Data  Summarization  Event Detection