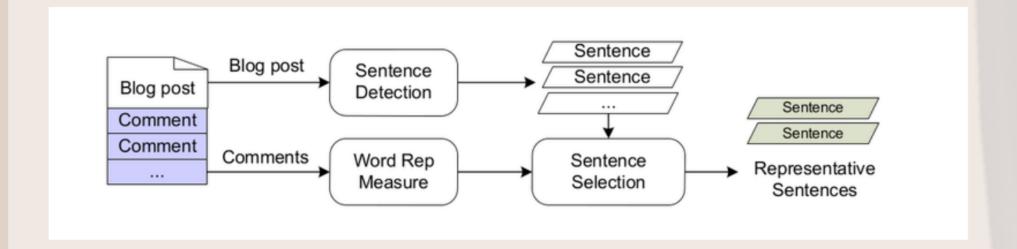# Comments Oriented
Blog Summarization

Motivation

A recent study on blog conversation showed that readers treat comments associated with a post as an inherent part of the post.

Comments gives the general opinion about the blog and this knowledge can then be used to extract most relevant sentences from the blog post.

Problem Statement

Given a blog consisting of a set of sentences and a set of comments associated with it, the task is to extract a subset of sentences from the blog such that it best represents the topics presented and discussed among its comments.

Word Representative Measure

Calculating WRM of each word for all the comments based on:

- Reader's Authority
- Popularity of a comment
    - Likes
    - Replies
- Topic Clustering
- Named Entity

Sentence Extraction

Each sentence is represented by a bag of words and its weightage is given by

$$W(Si) = Sum(W(Ti)) / |Si|$$

where,

$W(Si)$ = weightage of the sentence Si,
$W(Ti)$ = weightage of the term Ti as calculated previously.

Results

A number of top weighted sentences is shown as the summary.

Output Format:

- Inline Summary in Chrome Extension
- A Beamer PDF of the Summary to download

PREZI

Dataset

- Summarazation is done considering TechCrunch blogs as the dataset

Assumptions

- Proposed solution gives a good output summary for long blog posts and sufficient number of comments associated with it.

Future Work

- Dataset Expansion:
  Extending current implementation to various other blog websites

- Application of various semantic techniques for sentence extraction

- Eliminating redundant data/comments

Project Team

- Dilwar SK
  dilwarhossain.sk@students.iiit.ac.in

- Ritika Tanwani
  ritika.tanwani@students.iiit.ac.in

- Tejas Shah
  tejas.shah@students.iiit.ac.in

- Tohar Patel
  tohar.patel@students.iiit.ac.in

THANK YOU