Opinion Summarization of Web Comments

Martin Potthast and Steffen Becker

Comment Summarization

Comments can be found all over the Web, and on all kinds and types of objects.

Commenters express their opinion about the commented object, especially if the object is non-textual.

Reading comments is often very boring; to find good comments or to get an overview one has to read all of them.

In comment summarization, the task is to generate an overview of a set of comments with regard to opinion or other aspects of interest.

Text summarization is about extracting important sentences, which is frequently proposed for opinion mining in product reviews.

Comments, however, are difficult to handle:
- very short, often less than a sentence
- slang, and short message abbreviations
- incorrect spelling and grammar
- spam and multilingual text
- discussions in the comment threads

Therefore, treating a set of comments as a bag of opinion words is the most promising approach.

Comment Summarization for YouTube and Flickr

The technology is available as a Firefox and Chrome add-on:
http://www.webis.de/research/projects/opinioncloud

Offline:

An opinion dictionary was trained based on the General Inquirer, and about 10 million YouTube comments.

Rules for emoticons and other typical slang terms were set up.

Online:

A set of comments is summarized by a dictionary-based classification of each token into the classes positive, neutral, and negative.

The summary is visualized both as percentages of opinion terms, and as opinion word cloud.
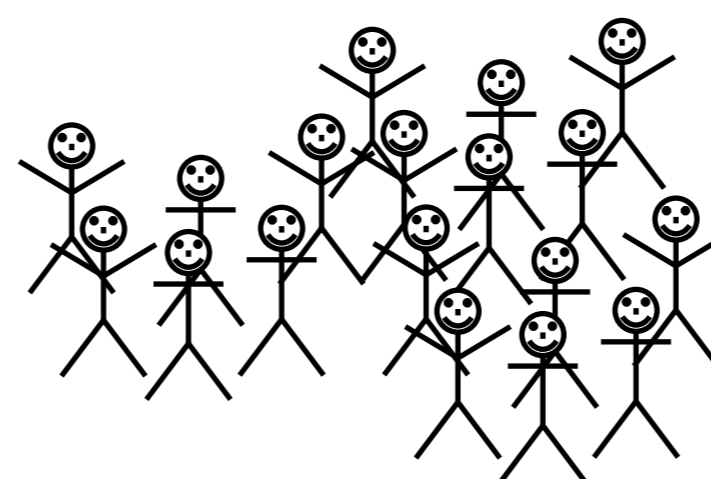
Comment-based Cross-media Retrieval

Object x — Object y

Comments D_x on x — Comments D_y on y

Representation x for x (bag of words / tags) — φ — Representation y for y

Web Technology and Information Systems — **Bauhaus-Universität Weimar**