# Exploratory Search on Social Media

Aaron Russ[1] and Michael Kaisser[2]

[1] DFKI, Berlin, Germany
*aaron.russ@dfki.de*
[2] AGT International, Berlin, Germany
*mkaisser@agtinternational.com*

**Abstract.** The rise of Social Media creates a wealth of information that can be very valuable for private and professional users alike. But many challenges surrounding this relatively new kind of information are yet unsolved. This is true for algorithms that efficiently and intelligently process such data, but also for methods of how users can conveniently access it and how results are displayed. In this paper we present a tool that lets users perform exploratory search on several Social Media sites in parallel. It gives users the opportunity to explore a topic space, and to better understand facets of current discussions.

**Key words:** Social Media, Exploratory Search

## 1 Introduction

Keeping a pulse on what is happening on Social Media sites is useful for many purposes, e.g. to stay up-to-date with the latest trends or to understand how people feel about issues, events, or products. Yet, most Social Media sites themselves provide only rudimentary search functionality, and as of yet there are not many tools available that provide a better experience. In this paper we present "Social Explorer", a tool designed to perform exploratory search on several Social Media sites in parallel. The tool itself can be tested at `http://esttest.dfki.de/`.

Social Explorer supports searching the latest public posts on Twitter and Facebook. It also features blog search capabilities, and can be used for standard web search via the Bing API. Search Results are clustered into subtopics, or alternatively a word cloud view can be displayed. Both the clusters and the word cloud view can be used to drill down into subtopics of the search result set. We use a custom-build post ranking algorithm that aims to display posts that are the most representative for the discussion at large (or a subtopic thereof) on top of the results list.

## 2 Related Work

There are at least three realms in which significant related work to Social Explorer can be found: publicly available tools on the Web, academic work and commercial products by Social Media Analytics companies.[1]

---

[1] For reasons of brevity, we will not cover the later here.
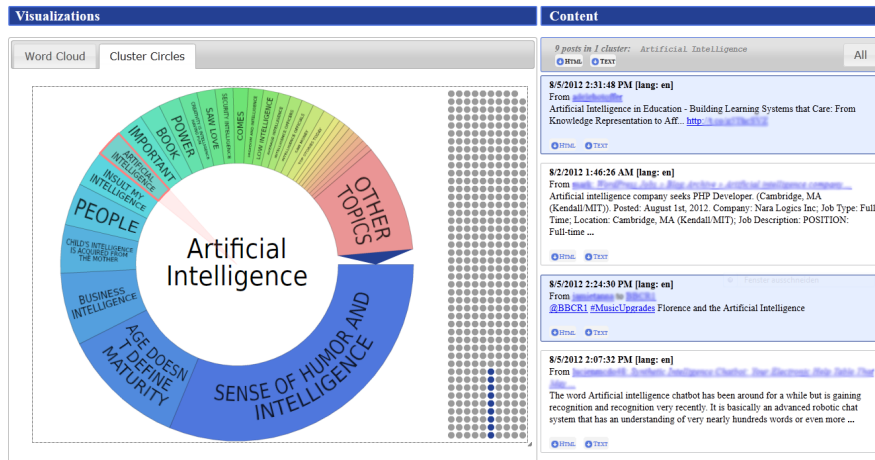
**Fig. 1.** Social Explorer's result view for the query "intelligence". To the left we see a clustering visualization of the result set. If the user selects the other tab, a word cloud view of the result set is shown. To the right the posts belonging to the currently selected subset are shown.

In terms of publicly available tools, Social Mention[2] provides a nice and powerful real-time social media search and analysis platform that aggregates content from various Social Media sources. Next to the traditional search result listing, Social Mention provides a faceted search interface that allows the user to filter the results according to "Sentiment", "Top Keywords", "Top Users", "Top Hashtags", and "Sources".

So far only a few academic papers describing projects dealing with how Social Media insights can be accessed and be presented to the user can be found. [3] present a dedicated exploratory search application for Twitter, TweetMotif[3]. Some features of our system can be found here as well, most notably the fact that posts in a result set are grouped by frequent significant terms. In such a way a faceted search interface is created in order to facilitate navigation and drilldown into subtopics. In [4], a modular and scalable framework for opinion mining on Social Media sources and a tool that was built using this architecture is presented.[4] It provides several data visualizations, and a wide range of options that allow the filtering of a result set, e.g. according to time, topic and sentiment.

---

[2] http://socialmention.com

[3] The demo itself is down, but the project's source code can be found here:
https://github.com/brendano/tweetmotif

[4] http://varovani.barcelonamedia.org/Orange/

# 3    Core Functionalities

In the following we will explain some of Social Explorer's key features in more detail.

## 3.1    Source Selection

Social Explorer allows the user to search a selection of Social Media sites either in parallel or individually. In a "Data Source Selection" tab, the default option will search for public posts on Twitter and Facebook, plus relevant blog posts found through Google's blog search API. These three sources can also be selected individually. Additionally, the tool can be run in standard web search mode, by using the public Bing API. Furthermore, a selection of stored documents can be searched. In the current demo, the last selection option contains tweets issued in the New York region in April 2012. Such content is internally stored in a Lucene [2] index, which allows for fast text search.

## 3.2    Clustering

For a given query, different discussions are often happening at the same time on Social Networks and different opinions on the subject are shared. The common approach of listing posts sorted by their timestamp makes it hard for the user to comprehend the bigger picture of a discussion, a typical case of not seeing the forest for the trees. To assist the user here, we employ a simple, well-known clustering tool, $Carrot^2$ [5]. In Figure 1 an example of its output can be seen prominently to the left. Each cluster in the cluster wheel can be clicked, the results it contains are then displayed to the right.

## 3.3    Word Cloud

We provide a second search result visualization to the user that, similar to clustering, serves as a high level summary about the ongoing discussions: A word cloud. For a given result set, it displays the 100 most frequent words where the size of a word corresponds to the word's frequency (stop words are excluded). Each word in the tag cloud can be clicked, which brings up a new tag cloud created just from the posts that contain this word. As with the cluster wheel, results that belong to the user's current selection are displayed to the right.

## 3.4    Post Ranking

The problem of how to rank Social Media posts differs from ranking documents or web sites in a few crucial ways: 1) Social Media posts are typically fairly short, this is especially true for tweets. 2) Social Media posts often get forwarded, which means that there are many identical or near-identical posts. 3) Trending topics and memes are spreading as fast as they die down, so different temporal result snapshots for a query usually provide very different results.

For these reasons, we implemented a custom ranking method, that is conceptually related to the word cloud view described above. The first position in the ranking is selected according to

$$\max_{P \in D} \left( \sum_{i=1}^{n} \text{freq}_{w_i}(D) \right) \qquad (1)$$

where

$$P : \text{Post consisting of terms } w_1 \ldots w_n$$
$$D : \text{Result set of all posts}$$
$$\text{freq}_{w_i}(D) : \text{Frequency of word } w_i \text{ with regard to } D$$

Thus we select the post $P$ whose words are in sum most frequent in the whole result set. $P$ is then removed from the result set $D$, and $\text{freq}_w(D)$ is artificially lowered for all words $w_1 \ldots w_n$ contained in $P$ by dividing it by a constant c, where we commonly use $c = 5$. This process is repeated until the top x positions are filled. This approch is inspired by [1], where a similar method was successfully applied for the "other" question task in TREC's QA track.

With this ranking algorithm we achieve the following: 1) A post that is representative for the discussion at large get selected as the top result. 2) At the following positions in the ranking we have posts that cover other aspects of the discussion. 3) Duplicates and near-duplicates of selected posts will be pushed down in the ranking, because the words they contain already have been downgraded.

## 4    Conclusions

We have presented "Social Explorer", an exploratory search tool that allows its users to discover a topic space, and to dive deeper into facets of the currently ongoing discussions on Social Media sites. We hope you like it.

## References

1. Michael Kaisser, Silke Scheible, and Bonnie Webber. Experiments at the University of Edinburgh for the TREC 2006 QA track. In *Proceedings of the 2006 Edition of the Text REtrieval Conference (TREC-06)*, 2006.
2. Michael McCandless, Erik Hatcher, and Otis Gospodnetić. *Lucene in Action, Second Edition*. Manning Publications Co., 2010.
3. Brendan O'Connor, Michel Krieger, and David Ahn. TweetMotif: Exploratory Search and Topic Summarization for Twitter. In *Proceedings of ICWSM 2010*, 2010.
4. Carlos Rodriguez-Penagos, Jens Grivolla, and Joan Codina-Filba. A Hybrid Framework for Scalable Opinion Mining in Social Media: Detecting Polarities and Attitude Targets. In *Proceedings of the Workshop on Semantic Analysis in Social Media*, 2012.
5. Jerzy Stefanowski and Dawid Weiss. Carrot2 and Language Properties in Web Search Results Clustering. In *Proceedings of AWIC 2003*, 2003.