Journal of Web Engineering, Vol. 0, No. 0 (2016) 000–000

Enhanced Exploration of Social Media Data with the Beomap: an Ad Hoc Topic Map

Martin Leginus
Department of Computer Science, Aalborg University, Selma Lagerlofs Vej 300
9220 Aalborg-East, Denmark
and
Falcon Social, H. C. Andersens Blvd. 27
1553 Copenhagen, Denmark
martin.leginus@gmail.com

ChengXiang Zhai
Department of Computer Science, University of Illinois at Urbana-Champaign, 201 North Goodwin Avenue
Urbana, IL 61801, United States
czhai@cs.uiuc.edu

Peter Dolog
Department of Computer Science, Aalborg University, Selma Lagerlofs Vej 300
9220 Aalborg-East, Denmark
dolog@cs.aau.dk

Jess Alfredsen
Falcon Social, H. C. Andersens Blvd. 27
1553 Copenhagen, Denmark
jess.alfredsen@gmail.com

Received (received date)
Revised (revised date)

Social media is ubiquitous. There is a need for intelligent retrieval interfaces that will enable a better understanding, exploration and browsing of social media data. To this end, we propose a novel two dimensional ad hoc topic map (called Beomap). The main novelty of Beomap is that it allows a user to define an ad hoc semantic dimension with a keyword query when visualizing topics in text data. This not only helps to impose more meaningful spatial dimensions for visualization, but also allows users to steer browsing and exploration of the topic map through ad hoc defined queries. We have applied the Beomap in exploring Twitter data, and evaluated the proposed Beomap in two ways, including an offline simulation and a user study. Results of the evaluations show that the new Beomap interface is better than a standard interactive interface. We also present a crowd-sourced user study analyzes whether topic maps with noun phrases are considered as more relevant than hashtags based Beomaps. Further, we implement visualization metrics for Beomap to enable sentiment and location based exploration. To illustrate how the sentiment and location is applied, use cases with discussions on specific queries are presented.

Keywords: Adaptive visualization; adaptive browsing; topic map; social media

Communicated by: to be filled by the Editorial

1. Introduction

To fully exploit the potential of social media data, there is a need for social analytics tools to enable more effective ways to understand, analyze and exploit social media information.

An especially useful way to help users explore the information space is to visualize topics. Interfaces for visual exploration of topics proposed in [5, 9] give users an understanding of underlying topic space and enable them to browse related topics. Further, these interfaces enable a user-centered exploration exploiting different machine learning techniques. These interfaces usually visualize relations between topics based on relevance scores but do not exploit other meta-data dimensions (e.g., popularity, sentiment or location) which might prevent the user from a more realistic understanding of underlying topic space. Moreover, it is not clear what a certain direction or position in the interface means e.g., what does it mean when a topic is presented in the top right corner of the interface. Self-organizing maps [10] satisfy this partially, however the dimension cannot be flexibly modified by the user. Further, self-organizing maps are computationally expensive as well as difficult to clearly interpret horizontal and vertical axes. The last but perphaps the most crucial deficiency is the limited ability of a user to reach ad hoc relevant topics regions in a flexible way i.e., user-driven exploration. The user-driven exploration is different from machine-learning user centered exploration as it allows the user to completely control an exploration process.

In this paper, we propose a way to solve these limitations with a novel two dimensional ad hoc topic map (see Figure 1) which we refer to as Beomap (*beo* in latin means to make happy). The idea is to allow a user to define the meaning of a dimension using an arbitrary query to visualize an ad hoc topic map, which enables user-guided exploration and browsing of the underlying social media topics space. Such a Beomap has many desirable properties and advantages. First, Beomap would be especially useful to help users interact with search results when there is a missing aspect in the query (e.g., most top ranked results missed one query aspect), which often is the case when the search results are not satisfactory. We refer to a missing aspect as a need to view topics (from both inside as well as outside of search results) and underlying documents from a particular aspect/perspective. Beomap extends a standard ranking which is usually presented with an ordered list of topics into a two-dimensional ranking. The second dimension expresses how well a topic relevant to the main query also matches an additional aspect or a perspective which can be flexibly expressed with a second ad hoc query. The positive effect of this two-dimensional ranking enables easy understanding of topics orientation within the topic map i.e., a particular position of the depicted topic represents its closeness to the main query as well as to the second ad hoc query. Hence, users are better able to understand underlying topic structure between individual topics, which might lead to serendipitous discoveries. Further, Beomap enables a user to explore and navigate the topic space through user-chosen visualization metrics. The ad hoc topic map can also be generated and visualized according to a particular predefined visualization metric e.g., recency, relevance, popularity or location based dimension. Due to this, Beomap provides the means for enhanced browsing into related relevant topics from underlying social media data as well as the means for improved understanding of the topics structure.
The proposed topic map visualization is general enough and it can be applied to other domains where topic keywords are available or can be derived for a particular resource (other social media data e.g., Instagram, Facebook; folksonomy data e.g., Delicious, Bibsonomy and other forms e.g., research publications, news articles etc.).

This paper is an extended version of the conference paper from [12]. We provide various extensions. First, we implement three visualization metrics for the Beomap. Two of these metrics combine relevance with sentiment polarity (positive in one case and negative in another case) and in the third variation, we combine location with relevance. Further, we extend the Beopmap to depict also noun phrases not only hashtags. We perform a crowd-

Fig. 1. Beomap is a two-dimensional topic map which supports ad hoc semantic dimensions that can be easily defined through a second query e.g. Nato. Further, it allows one to visualize a topic map with respect to the user defined meta-data metric e.g., Relevance. The topics in the top right corner are relevant for the main query as well as for the ad hoc dimension.

sourced study to analyze whether the Beopmap with noun phrases is more preferred over hashtag based Beomap. We further present two use cases for location aware exploration and sentiment based exploration of twitter data using Beomap.

To evaluate Beomap, we implemented BeomapApp a prototype system over social stream data from Twitter. The core component of the prototype system is a two-dimensional ad hoc topic map denoted as Beomap as well. To evaluate the proposed topic map, the offline browsing simulation of Beomap was performed which proved an enhanced retrieval of additional relevant tweets. We simulated various browsing strategies for Beomap and the baseline (see the detailed description of strategies in Section) assuming a user optimal behaviour i.e., an upper-bound scenario. The precision improvements of **BeomapBest**(the best strategy for Beomap) are significantly better than the **BLBest** (the best strategy for the baseline) (Wilcoxon signed rank test, $p < 0.00001$).The best precision versus user effort is attained with **BeomapMisAsp**, when a missing aspect of the query is used for Beomap generation. Moreover, the performed user study proves benefits of BeomapApp. Thirty one users consider the system as significantly more useful and are significantly more satisfied with the system than the baseline. Further, they perceive the system as a flexible tool for exploration, browsing and analysis of Twitter data as well as being easy to browse and interesting. Analyzed usage logs and acquired user comments about Beomap validate the benefits of two-dimensional ad hoc topic maps i.e., improved recall, more explored topics as well as positive comments about Beomap.

The structure of the paper is organized as follows. Section 2 positions our work with respect to existing information retrieval interfaces. Section 3 describes a novel two-dimensional topic map as well as several visualization metrics. Section 4 presents findings from the evaluation of Beomap (consisting of offline browsing simulation as well as user study). Section 5 describes location based and sentiment based exploration with Beomaps as use cased. Section 6 discusses the benefits and limitations of Beomap. Section 7 summarizes the paper achievements and outlines future work.

2. Related Work

VizLinc [1] helps users to get an understanding of underlying data. It allows the user to find patterns and relations between mentioned entities and facilitates users to narrow down the existing documents to a fraction of only relevant ones. The system allows a text and entity search. The retrieved results can be also geo visualized in the map. Further, recognized entities matching a query are visualized in the graph. Entities are clustered and their neighbours can be exploited with the n-hop network functionality. The limitations of VizLinc are: (1) a graph exploration is limited only to n-hop neighbours - not possible ad-hoc browsing; (2) no meaningful explanation of direction within the visualized graph; (3) the system is not adapted for social media data where a meta-data annotation might be more challenging. Bron et at. [4] proposed a subjunctive exploratory search interface which supports exploration of multiple views on a topic as well as enables a discovery of patterns in the data. The proposed interface combines two side-by-side versions of an exploratory search interface that are extended with visualizations in which characteristics of the result sets can be depicted and compared. The performed user study indicates that the subjunctive interface enables users to define more diverse queries and to retrieve more diverse documents than with the standard exploratory search interface. Despite the benefits of the system, it is not possible to visualize user chosen metrics except of term frequencies and time.
Apolo, an interactive user interface [5], helps users to make sense of large graph data. The aim

of the system is to support user-driven understanding of data. The underlying information landscape is adapted to user information need using Belief propagation algorithm. Users rated the system positively when used for sensemaking while browsing a scientific literature network. This work is in line with our motivation, to support user-driven exploration of the information landscape. The user-driven exploration in Beomap is achieved through a second ad hoc query dimension that allows users to regenerate freely a topic map according to user information need. The consequences of our design are a more realistic understanding of the underlying information landscape and the possibility of observing topic correlations between the main and second ad hoc query. Although, the adaptation is not the scope of this study, our system could be easily extended to support personalization. The system proposed by [9] enables users to direct their exploration of the information space. This interactive information retrieval interface not only allows users to steer the direction of the exploratory search but also models user information need with the reinforcement learning technique. The performed user study validates the contributions in terms of effectiveness (improved recall of relevant and novel information) of the interface when compared to the standard keyword based search interface. Similarly, in this work, we propose a system which allows users freely, in ad hoc fashion, to steer the browsing and exploration process. The benefits of our system in comparison to [9] are three-fold. First, our system allows one to visualize information space according to a user-selected ranking metric such as Relevance, Popularity, Location, Sentiment polarity and Recency. Second, the ad hoc nature of the proposed system allows users to steer exploration into the many regions that might not otherwise be reached. Further, Beomap interface gives a clear interpretation for direction and orientation in the visualized topic space i.e., a particular position and movement in the topic map represents relatedness to the main query or the missing aspect query.

3. Beomap

Beomap is a novel two-dimensional topic map with both directions controlled by a user. It addresses a major challenge in visualizing text information, i.e., a user is often lost in the space as there is no clearly defined meaning associated with an orientation (e.g., it is unclear what to expect if a user moves the cursor to the left or up). In this paper, we particularly explore a special Beomap useful for a search interface where the vertical dimension is relevance (consistent with a standard search interface where documents are ranked vertically). A second dimension is defined by a user in ad hoc fashion by placing a query and choosing an appropriate visualization metric. The aim of Beomap is to ”stretch” the standard linear relevance ordering to a two-dimensional ranking that would enable a user to examine visually the list based on how well topics match the second dimension, which can be defined flexibly by another query. Therefore, the second ad hoc dimension indicates whether a topic relevant to the main query is related to a missing aspect or a perspective. The second dimension can be visualized with a user-selected metric. Note that the second dimension is offered as an option, which the user can take as needed. Thus it is a very natural extension of the current search interface. A user can also easily go back to one dimension if needed. The possibility to change a visualization metric according to user needs provides new ways of user-driven exploration of underlying topic space and corresponding tweets. Hence, Beomap interface provides a new way of how to visualize user information need combined with arbitrary meta-data metric dimension (relevance, time, location, popularity, sentiment polarity or fuzzy metrics) which is different from the traditional way of visualizing with meta-data such as a time line, etc. The flexiblity to guide freely the exploration and browsing of topics within social media data

through an ad hoc semantic dimension is different from established browsing interfaces like a faceted search or a word cloud. The existing interfaces provide only topics and aspects that are within the search results context; therefore, user ability to browse into remote topic regions is not supported.
To the best of our knowledge, no previous work has explored this kind of topic map, especially in the context of helping users interact with poor search results and for social media exploration. Beomap has many potential benefits. First, it would be particularly useful when a user would like to examine the missing aspect or would like to explore relevant tweets from a particular perspective or angle. Further, Beomap addresses the problem of orientation in text data visualization since it is not obvious what directions in the topic map mean (e.g., what does moving in a particular direction mean in the topic space?). Beomap provides a solution to this directions interpretation problem when visualizing text data. Beomap enables a better topic understanding by enabling one to view a topic from multiple aspects or perspectives which can be expressed flexibly through a second ad hoc dimension query. Further, allowing the user to select a visualization metric also improves a topic space understanding.

3.1. *Various Metrics for Beomap Ad Hoc Dimension*

The visualization of a topic map is metric dependent. Hence, relatedness or closeness to the ad hoc query is calculated according to the user-selected metric. The calculation of the final score for a topic presented in the topic map is as follows: (1) Retrieve tweets matching a given query and calculate the metric score for each matching tweet; (2) Aggregate scores of mentioned keywords i.e., hashtags/noun phrases over the retrieved tweets; (3)Present keywords with the greatest levels of user-selected metric dimension in the topic map. In this work, we explored several metrics and analyzed in what situations these metrics might be useful for the end users. In the following, we present several metrics with a brief rationale:**Relevance** is a default metric in the Beomap system. The relevance score might be calculated with any retrieval ranking function. In this work, we utilize OKAPI BM25 retrieval function. Once relevance scores are calculated for the query matched tweets, the hashtags and noun phrases across tweets are aggregated such that final aggregated scores are obtained. The most relevant topics with respect to the main or second ad hoc query are visualized in the topic map. Obviously, the relevance based topic map allows users to explore and browse the topics that are relevant and related to the placed main or second ad hoc query. Further, it reveals how great the relevance relatedness is and what the relevance correlations are between individual topics. **Popularity** reflects a popularity of a topic keyword within the query matched tweets. **Time recency** calculates the recency score for topic keyword aggregating over tweets e.g., the more relevant tweets that are more recent attain a greater score. **Location** calculates a proximity with respect to the placed location-based queries. **Sentiment closeness** calculates for each tweet from the underlying collection a sentiment score. The sentiment scores are aggreagated according to a topic keyword and consequently visualized. This metric allows the user to explore positive or negative topics which can be either common or specific for both queries i.e., the main query as well as the ad hoc query. **Authority/Reputation** is a metric which would consider the tweet author's reputation within the social network for Beomap visualization. **Personalized** is a metric which adapts the visualization of the topic map with respect to user preferences as well as to the placed queries (when the system collects user preferences).In some scenarios, a **user customized metric** may be useful by combining some of the metrics described above.

3.2. *Tasks Supported by Beomap Interface*

The accurate definition of user supported tasks by Beomap interface enables one to define a basis for systematic comparison with other interfaces as well as for a further evaluation of the proposed interface. In this paper, we consider the following tasks as our targeted applications:

Missing Aspect Exploration: Find and browse relevant topics which are related to the main query as well as to the ad hoc missing aspect query. The second ad hoc query represents a missing aspect or a needed perspective when examining tweets that match the main query. For instance, users might be interested in this kind of task when the number of tweets matching the main query is large and users would like to examine only specific ones related to the missing aspect or requested perspective. Possible examples are: Which topics are relevant for the Russian Ukrainian conflict from the perspective of NATO, EU or Poland? Which topics are relevant to the placed query and are related as well to a particular geo-location?

Topics Space Understanding: Obtain topic space understanding and find relationships between interesting topics. Repeated examination of Beomap with respect to the main query and several second-dimension ad hoc queries combined with different meta-data metrics (repeating a missing aspect exploration task several times) will enable users to get an enhanced understanding of underlying topic space. Possible examples are: What topics are related and what are the relationships among them with respect to the main query e.g., Russian Ukrainian conflict?

The presented task types might be performed by users several times and combined in arbitrary order to accomplish the final goal i.e., finding relevant tweets that are not retrieved with the initial user standard keyword search.

3.3. *Social Media Prototype System*

We implement a prototype system for social media exploration and analytics. We refer to this system as BeomapApp. The BeomapApp consists of several components (see Figure 1). The main component is Beomap, a two-dimensional topic map that depicts a spatial distribution of topics according to a user selected visualization metric. Topics are presented according to two dimensions: (1) a relevance with respect to the main query (vertical dimension); (2) a metric-based closeness to the ad hoc second query. The topic map follows a two-dimensional cartesian coordinate system, presenting only the first quadrant. Hence, topics more related to the main query are presented in the top part of the map and topics related to the second ad hoc query are presented at the right side of the map. The topic map allows users to distinguish clearly which topics are related to the main or a second ad hoc query and the topics that are common for both queries. The common topics are presented in the top right corner of the map. The calculation of an exact position in Beomap is based on standard information retrieval ranking function OKAPI BM25 which can be defined as:

$$S(tw, Q) = \sum_{q_i \in Q \cap tw} TF(q_i, tw) \cdot IDF(q_i) \tag{1}$$

where

$$TF(q_i, tw) = \frac{f(q_i, tw) \cdot (k_1 + 1)}{f(q_i, tw) + k_1 \cdot (1 - b + b \cdot \frac{|tw|}{avgtwl})}$$

$$IDF(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}$$

and $f(q_i, tw)$ is a q_i term frequency within a tweet tw, $|tw|$ is the length of a given tweet tw, $avgtwl$ is average length of a tweet within the corpus, N is a total number of tweets in the corpus and $n(q_i)$ is the number of tweets that contain the term q_i. We set the following values for parameters $k_1 = 1.2$ and $b = 0.75$. When visualizing Beomap, we combine a main query Q_{main} and an ad hoc query Q_{adhoc} with logical OR into a query Q. Once a relevance score is computed for each tweet with respect to Q, we calculate an exact position for a particular hashtag h for a vertical dimension (y-axis) defined as:

$$pos_{vertical}(h) = \sum_{tw \in TW_{Q_{main} \cap h}} S(tw, Q)$$

where $TW_{Q_{main} \cap h} = \{tw | S(tw, Q) > 0, h \in tw, q_{main_i} \in tw\}$ is a set of all tweets which contain a particular hashtag h and at least one term q_{main_i} from the main query Q_{main} and the relevance score $S(tw, Q)$ with respect to the query Q is greater than 0. The calculation for the horizontal dimension (x-axis) is calculated in a similar way except that a set $TW_{Q_{main} \cap h}$ is replaced with $TW_{Q_{adhoc} \cap h}$

$$pos_{horizontal}(h) = \sum_{tw \in TW_{Q_{adhoc} \cap h}} S(tw, Q)$$

Due to long-tail distribution of hashtags in Twitter stream [16], there is a need to avoid depicting many hashtags with low levels of attained aggregated relevance scores in the same topic map region to provide an enhanced readability of the map. Hence, we order hashtags in an ascending order with respect to aggregated relevance score for each dimension and the rank position within the ordered set expresses a final position in the map for the given dimension. This solution is not optimal but we obtained the best topic map readability in comparison to standard scaling functions. Topics are represented in the bubble chart, each topic keyword is represented with a disk where the label is a topic keyword and the radius corresponds to a scaled topic popularity within the collection. We utilize log scaling to visualize a disk with the scaled topic popularity to make a topic map more readable. This is due to a long tail distribution of hashtags in the collection i.e., very few popular hashtags and many hashtags with low frequency occurrence. To obtain a full understanding of the topic map, the tooltip component was developed because topic keywords from social media data are not always easily understandable by the users e.g., *#tcot, #nomistrals4putin*. The tooltip is displayed when the user hovers over the particular topic in the map and presents a few relevant tweets for a given topic. The other components of the visual interface are two content panels displayed in tab view. In the first panel, a user can explore the most relevant topics to the placed queries with the corresponding summary tweets. The second content panel presents the most relevant tweets sorted according to the user selected metric. A user can place the main query in the input field located above the topic map and confirm it with the *Explore* button. Initially, the topic map is not visible to a user but can be diplayed by clicking on the button *Show topic map*. The user can define an ad hoc query through the input field which is located below the lower right corner of the topic map with the watermarked hint to enter the ad hoc dimension query. The ad hoc dimension metric can be chosen from the dropdown list which is presented in the top right corner of the interface.

3.3.1. *Sentiment and location metrics*

We add sentiment based exploration within Beomap to enable: (1) a discovery of positive/negative terms w.r.t to main and second query or both dimensions; (2) a particular position within the map indicates a sentiment strength of term w.r.t both query dimensions; (3) an understanding of sentiment topics structure which scope is defined with main and second ad hoc queries.

We implement sentiment closeness metric in the following way:

- Annotate underlying tweets collection with positive and negative sentiment scores. We exploit a SentimentStrength tool [15] which performs particularly well on social media text across different domains. The tool leverages a list of affective terms and set of rules to detect a sentiment of a tweet. The approach outperforms supervised sentiment analysis methods when amount of human coded data is unsufficient and when detecting sentiment across wide spectrum of domains.

- We introduce two sentiment visualization metrics which indicates positive and negative sentiment of hashtags and noun phrases w.r.t the entered queries.

- The final score of a keyword is combination (multiplication) of tweet relevance and its sentiment polarity w.r.t. to the main or second ad hoc query. Please note, that both vertical and horizontal axis combine relevance together with sentiment polarity w.r.t. to the main or ad hoc queries.

We illustrate sentiment based exploration with the use case described in Section 5.2.

Further, we extend the original Beomap to visualize relevance together with location proximity metric w.r.t. to the main and ad-hoc query. The location proximity of the tweet w.r.t the point of interest is combined with relevance. To illustrate location aware exploration, we present a use case in Section 5.1.

3.3.2. *Noun phrases*

When summarizing user generated text on social media, several types of terms might be leveraged. The most convenient choice is the exploitation of **hashtags** for summarization purposes. Hashtags are often mentioned in the social media text to direct and group the user generated content into a particular discussion thread or trending topic. Therefore, hashtags provide good insights when summarizing social media information landscape. However, the coverage of hashtags within user generated content is low. For instance, on Twitter only about 24% of tweets contain at least one hashtag. To ensure that majority of relevant user generated content is retrievable from the topic map, other extraction methods should be exploited. In Section 4.1.3, we extract named entities for simulated offline evaluation of Beomaps. Although, recognized named entities are reported by users as more relevant than extracted unigrams [11], we consider a problem of named entity recognition in a large scale on social media data for different languages and domains as a challenging research problem. Another approach is keywords extraction where informative phrases are selected from the text. Phrase mining can be conducted either through application of natural language processing tools or with statistical approaches [8]. In this work, we leverage a Standford POS tagger with adjusted model for Twitter [7] to detect noun phrases. Similar to [6], we annotate terms (stop words were filtered out) from each tweet with POS tags and consequently extract noun phrases matching the following regular expression:

$$Noun - Phrase := adjectives * nouns+$$

which indicates to extract all noun phrases which can have zero or non-zero adjectives and one or more nouns. Beomap generated with noun phrases ensures greater coverage of the underlying tweets. In Section 4.6, we further investigate whather users prefer more Beomaps with noun phrases over hashtags.

3.4. *Technical Implementation*

The system has been developed with J2EE using Primefaces framework for the web interface and javascript JQPlot library for the topic map rendering. The backend application has been deployed to the Jboss application server 7 and the data were stored in Elasticsearch. The aggregation component of the Elasticsearch engine has been used for the calculations of metric dependent scores. The benefit of our implementation is easy deployment to the cloud infrastructure which might be needed for large data collections.

4. Evaluation

To validate contributions of Beomap as well as the social media exploration system BeomapApp, we perform an evaluation which consists of a simulated offline browsing evaluation of Beomap. Further, to validate the benefits of Beomap when integrated in a real application like BeomapApp, we perform a user study. We used the following hypotheses as guidelines when designing the evaluation of Beomap and BeomapApp. The hypotheses are presented below:

- $H1$: Beomap enables a more effective retrieval of additional relevant tweets (those that are relevant but not retrieved within the initial query top-k results) than in the baseline.
- $H2$: BeomapApp provides greater search and browsing satisfaction than the baseline.
- $H3$: BeomapApp is perceived by users as more useful and flexible than the baseline.
- $H4$: Using Beomap, participants will feel more familiar with the topics and contents of a collection than in the baseline.

4.1. *Offline Simulation*

To validate whether Beomap enables and facilitates navigation into relevant topics and consequent retrieval of relevant tweets which are difficult to retrieve with original query, we design the following offline evaluation. Similarly as in [17], we perform a simulated retrieval of additional relevant tweets where a user repeatably reformulates his/her queries. The aim is to validate whether Beomap facilitates more effective retrieval of additional relevant tweets. We retrieved available tweets with relevance judgments from TREC 2011 collection [14]. The relevance judgments sets were built using standard pooling technique. The tweets relevance was assessed with the three-points scale: (0: irrelevant, 1: relevant and 2: highly relevant). In this work, we consider both relevant and highly relevant tweets as equally relevant. The TREC2011 microblog corpus consists of fifty distinct queries and we denote each query as $Q_{original}$. We define a set of additional relevant tweets $R_{additional}$ as tweets that were rated relevant in TREC2011 but were not retrieved within the top k positions for the particular query $Q_{original}$. We aim to measure whether Beomap enhances retrieval of additional relevant tweets $R_{additional}$ in comparison to various baseline methods.

4.1.1. *Baseline Methods*

BL: This baseline method is based on the initial query $Q_{original}$. The precision at k is measured on top of tweets retrieved by $Q_{original}$ positioned from k to $2k$ rank with respect

to the additional relevant tweets $R_{additional}$. This simulation can be perceived as if the user would like to retrieve more additional relevant tweets after already viewing some relevant tweets i.e., first top k tweets.
BLMR: This baseline method orders topic keywords e.g., hashtags by the aggregated relevance with respect to the $Q_{original}$. The most related relevant topic in combination with the $Q_{original}$ is used as a new query. The precision at k is measured on top of first top k tweets retrieved by the new query. This imitates user behaviour when the original query is automatically extended with the relevant related topic i.e., a query suggestion provided by a search engine.
BLBest: This method orders topics by aggregated relevance with respect to the $Q_{original}$. The best related relevant topic is combined with the $Q_{original}$ into a new query. The precision at k is measured on top of first top k tweets retrieved by the new query. This reflects user behaviour when the user carefully explores the provided list of query suggestions and picks the best with respect to his/her information needs.

4.1.2. *Beomap Browsing Strategies*

We compare baseline methods with the following browsing strategies of Beomap.
BeomapMisAsp: Thismethod corresponds to the missing aspect task definition (see Section). The aim is to support the exploration of topics and corresponding tweets related to the missing aspect as well as the matching original query. In this simulated evaluation, we define a missing aspect as a term from the multiple terms $Q_{original}$ which occurs the least within top k tweets returned by $Q_{original}$. Consequently, the missing aspect is used as an input to the ad hoc dimension of Beomap. This browsing strategy assumes that a user examines topics presented in the topic map and picks the best topic. Finally, $Q_{original}$ is combined with the best node and precision at k is calculated with respect to $R_{additional}$ and the new query matching tweets. **BeomapMR:** This method retrieves the top most relevant topic keyword T with respect to the placed query $Q_{original}$. The relevant topic T is input for a second ad hoc dimension which leads to the generation of the topic map. A top relevant topic denoted as $T_{map}|T$ from the generated topic map is used for a query reformulation. The created query consists of $Q_{original}$, T and $T_{map}|T$. The precision at k is calculated with respect to $R_{additional}$ and the retrieved tweets. This replicates the similar user browsing strategy as **BLMR**; the only difference is further exploration of the topic map and selecting the most relevant topic with respect to the $Q_{original}$ as well as a second ad hoc query. The most relevant topic keyword from the topic map is the one with the highest average relevance score which incorporates relevance score with respect to the main query as well as to the ad hoc second query.
BeomapBest: This method is similar to **BLBest**. It assumes that the user examines all the related relevant topics and picks the best topic. The best topic is input for a second ad hoc dimension which leads to the generation of the topic map. Consequently, the best topic node is selected from the map i.e., a topic keyword which best suits user information need. The best topic node, the most related relevant topic (second ad hoc query) and $Q_{original}$ are combined into a new query. The precision at k is calculated with respect to $R_{additional}$ and the retrieved tweets for the new query. This browsing strategy is similar to **BLBest**. From the generated topic map, a user picks the best topic keyword.

4.1.3. *Data Preprocessing*

Topic maps generated from tweets' hashtags, when only a small twitter collection is con-

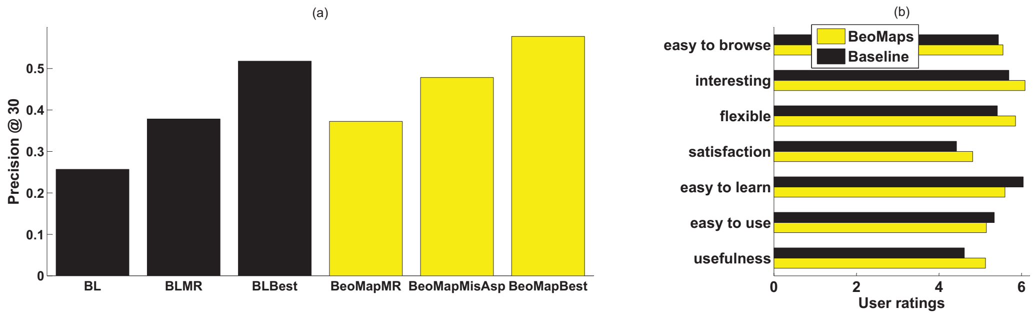

Fig. 2. Figure (a):Baseline methods are presented with black bars and yellow bars are corresponding to the browsing strategies of Beomap. We report means of measured precisions for individual queries from TREC2011 microblogging collection. Figure (b): Mean of user ratings for individual properties of both interfaces from the performed user study.

sidered, are sparse as only around 24% of tweets contain at least one hashtag. Thus, the generated topic map from hashtags, when an underlying tweets collection is small, prevent a complete retrieval of additional relevant tweets. To overcome this limitation, we have annotated tweets from TREC2011 corpus with recognized named entities using TextRazor tool. Each recognized entity has been converted into a hashtag i.e., Barack Obama transformed into #barackobama. In total, almost 77% of tweets contain at least one named entity which has been converted into a hashtag. Hence, 83% of tweets from TREC2011 corpus contains at least one hashtag or named entity. Almost, 90% of relevant tweets contain at least one hashtag or named entity.

4.1.4. *Results*

Obviously, the **BL** attains the lowest precision when retrieving additional relevant tweets. This supports a need for query reformulation and possible browsing of related relevant topics. The **BLMR** method attains higher precision than the **BL** as it extends the original query with the related relevant topic keyword which positively affects the retrieval of additional relevant tweets. The **BLBest** method examines a list of related relevant topics and assumes that the user would pick the best topic keyword and together with the original query form a new query. The browsing strategy **BeomapBest** attains higher precision as the best baseline method **BLBest**. Both methods **BeomapBest** and **BLBest** assume user optimal behaviour in selecting the best most relevant related topic which is used for the topic map creation in **BeomapBest**. Further, it assumes that the user picks the best topic keyword from the map. The evaluation reports the upper bound precision when (not) using topic maps. The precision improvements **BeomapBest** are significantly better than the **BLBest** (Wilcoxon signed rank test, $p < 0.00001$). Further, we report as well precision for missing aspect exploration method **BeomapMisAsp**, which achieves promising precision.

To illustrate the benefits of Beomap for the retrieval of additional relevant tweets, we examined the browsing logs of compared methods. For instance, the original query *White House spokesman replaced*, the **BLBest** method extended the original query with the *secretary* keyword. The Beomap methods attained better precision because of the topic keyword *c*arney when the second ad hoc dimension query was *presssecretary* for the **BeomapBest** method. This indicates the benefits of Beomap, a discovery of important and relevant topics with respect to the original query. Similarly, for the original query TSA airport screening, the **BeomapBest** method included *Seattle* keyword into a new query and consequently attained higher precision than the baseline. At first sight, it is not obvious that Seattle is related to the original query but that is because of *Seattle man acquitted in TSA airport case.*

4.1.5. *Normalization*

Obviously, the defined browsing strategies should be compared carefully because each strategy requires different amounts of user effort to perform browsing and final retrieval. Comparing upper bound strategies of **BLBest** and **BeomapBest** is fair because in both cases a user has to choose the best node from topic keywords sorted by relevance which matches his/her given information need. Additional user effort is required when exploring a topic map in **BeomapBest** and consequent selection of the best node. However, the user examination of visual topic map is trivial. Exploring a topic map provides the user with two additional benefits which are: (1) significantly improved precision when picking the best topic keyword from the map; (2) a better topics space understanding which might be beneficial for further browsing. **BeomapMisAsp** browsing strategy attains the best trade-off between user effort to generate and inspect a topic map and attained precision. Spotting a missing aspect of the original query is trivial and further exploration of the topic map leads to promising precision.

4.1.6. *Lessons*

The simulated evaluation of Beomap validates the benefits of the ad hoc topic map for enhanced retrieval of additional relevant tweets i.e., tweets that are not retrieved with the original query (validates hypothesis $H1$). Also, the simulation proved that Beomap interface is suitable for exploration of topics related to the missing aspect of the original query when the best trade-off between user effort and precision is attained. Further, we found that when the second ad hoc dimension is generated with respect to the prominent aspect of the query, there are no improvements in comparison to the baseline. However, an advantage of Beomap generated with the prominent query is facilitated topics space understanding which might be beneficial for further exploration.

4.2. ***User Study***

The purpose of the user study is three-fold: (1) to collect real users' feedback and opinions about Beomap; (2) to study the utility of the prototype system BeomapApp in comparison to a standard "Twitter like" list interface; (3) to understand what is the optimal intergration of Beomap into an application system. Participants in the study were asked to browse two different datasets; consequently we acquired their subjective assessments and opinions were acquired (similar to [2]). Further, we analyze log users' activities to explain the usage differences between two compared interfaces. We also measure standard information retrieval measures to compare both interfaces.

4.2.1. *Participants*

We recruited participants with Twitter accounts. The thirty-one study participants ranged in age from 18 to 38 years (average age 27.7). Fifteen were IT professionals, four were students (mathematics or economics background), and the other had careers in law, HR, design or the media. On average, they reported spending more than a half hour per week using Twitter.

Further, they all reported using other social media services like Facebook (all participants), Linkedin (9 participants), Google+ (6 participants), Instagram (3 participants).

4.2.2. *Compared Interfaces*

The performed user study follows within-subjects design. Beomap system was compared with the standard Twitter-like list interface. To ensure that acquired subjective assessments of the systems were not biased because of the coloring or layout differences of the systems, we implemented the baseline system. The baseline system is visually similar to the tweets content panel in Beomap [a]. We implemented the baseline system instead of using an external one because of the following: First, we ensured that participants will perform their searching and browsing activities on top of the same tweets corpus when doing evaluation tasks using both systems. Second, we assured that usability of both systems will be comparable and with minimized if any visual nuances. To further ensure that our systems are at a similar level of usability, we consulted the HCI expert about both interfaces. The suggested changes were incorporated before the user study was conducted.

4.2.3. *Procedure*

Each test participant completed a pre-questionnaire before performing the evaluation. After each performed task, participants were asked to complete the questionnaire to obtain a subjective evaluation of Beomap and baseline. We used the standard USE questionnaire [13] and extended it with a few additional questions about the following system properties: *flexible tool for exploration, browsing and analysis of Twitter data*,*interesting*, *familiar with the collection*, *easy to browse*. Users rated each aspect of the interface on a 7-point Likert scale i.e., 1 for strongly disagree and 7 for strongly agree. We used a 7-point scale to be consistent with the [13]. To minimize carryover effects in our within-subjects design, participants performed two tasks with the following conditions. Each task was performed on top of a different tweets collection. Further, the interfaces and task order were alternated and counterbalanced among participants. Participants were asked to perform the following two tasks

Task 1: After your graduation you will be looking for a job in industry. You want information to help you focus on your future job seeking. You know it pays to know the market. You would like to find information about employment patterns in industry and what qualifications employers are looking for in future employees. Further, you would like to improve your knowledge about the job seeking process including how to prepare your resume, how to prepare yourself for the interview and other important hints. The task selection has been inspired by simulated work task situations [3]. The collection consists of $2.5M$ tweets retrieved during the second half of September 2014. The participant is asked to perform the following subtasks: (1) Provide a list of five relevant links to the articles related to the employment trends, required qualifications needed for your job and similar topics which are relevant to you; (2) Provide a list of five relevant links about job seeking tips and hints which are interesting and relevant for you.

Task 2: Imagine you are about to write a report about military conflict in Ukraine. Before starting the writing you should retrieve relevant tweets and articles related to the topic. The task selection has been inspired by the simulated work task situation from [9]. The collection

[a] Screenshots as well as demo videos of both systems are available at `https://sourceforge.net/projects/mleginus/files/beomaps/`

consists of $130K$ tweets retrieved during the second half of September 2014.Users are asked to perform the following subtasks: (1) In which Ukrainian cities are there fights and military conflicts between Ukraine and Russian oriented rebels? (2) Which Ukrainian and Russian politicians and their opinions with respect to the military conflict are mentioned on Twitter?

The following subsections describe a qualitative analysis of user study as well as quantitative analysis derived from the usage logs. In the end, we are presenting learned lessons.

4.3. *Subjective Evaluation of Interfaces*

To analyze acquired user ratings for individual properties of compared systems, we utilized Wilcoxon Signed Rank test for paired samples (ratings are non-continuous and non-normal). BeomapApp is considered significantly more useful than the baseline interface ($p < 0.0001$). Participants found BeomapApp a more flexible tool for exploration, browsing and analysis of Twitter data than the baseline. Hence, the improved usefulness and perception of flexibility for exploration, browsing and analysis validate the hypothesis $H3$. Further, users were significantly more satisfied with BeomapApp ($p < 0.04$). Additionally, users rated Beomap more interesting and easy to browse (results are not statistically significant). Because users were significantly more satisfied with BeomapApp as well as they perceived it as more easy to browse and more flexible for exploration, browsing and analysis partially validate the hypothesis $H2$. Users considered both systems similar when becominga familiarized with the collection. A limitation of the current BeomapApp implementation is lower ratings for the easy to use aspect in comparison to the baseline system ($p < 0.02$). Further, participants assigned BeomapApp lower ratings for the easy to learn aspect ($p < 0.01$) suggesting that BeomapApp is more difficult to learn in comparison to the standard keyword search to which users are exposed on a daily basis. Upon analyzing participants' negative comments about BeomapApp, however, we found that most of the opinions were related to the design of BeomapApp and not Beomap topic map. For instance, a user reported that BeomapApp is not that intuitive e.g., *Not intuitive at first; UI is a bit cluttered*. This indicates the importance of concerned and careful intergration of Beomap into a real world application so that a user will not become confused and Beomap could be fully exploited. The lessons related to the Beomap integration into a real application that were learned from this user study are: (1) ensure that the input field for a second ad hoc dimension query can be intuitively found by the user; (2) ensure that a user easily recognizes how to change and select the ad hoc metric for Beomap. These findings will be used for a further BeomapApp improvement which is part of our future work.

The order in which users accessed the systems influenced the user ratings. In particular, when users first used BeomapApp, the subsequent ratings for the baseline system (used in the second task) were lower (similar finding as in [18]).

4.4. *Log Data*

The analyzed user activities indicate that when using BeomapApp users are more active in placing more queries, higher numbers of clicked links and hashtags from the tweets and of course several clicks on the topic keywords in the topic map. The detailed statistics are presented in Table 1: The presented results indicate that users placed more queries, explored more hashtags and obtained an overview about other possibly relevant hashtags by seeing them in the topic map. Hence, these results support the hypothesis $H4$ that a user become more familiar with the topics and content of the collection. Further, a high average number of seen hashtags from the map validates the benefit of the Beomap for better topics

Table 1. Average number of user activities for each considered interface.

User activity	BeomapApp	Baseline
# queries	15.39	13.68
# clicked links	3.857	4.6
# clicked hashtags from tweets	1.35	1.12
# seen hashtags from map	37.17	X
# clicked hashtags from map	4.61	X

structure understanding. We further analyzed placed queries to see whether the generated topic maps enhanced user ability to browse relevant related topics through query reformulations. For instance, when browsing tweets about conflict in Ukraine, Beomap helped users to discover keywords such as msf (*Doctors Without Borders/Médecins Sans Frontières*), russiainvadedukraine, natoforukraine,stoprussianaggression, savedonbasspeople, ukraineunderattack, ww3 . Similarly, for the job search task where users discovered topic keywords such as tipshintsjob, employmenttrends, jobhints. Often, users used Beomap for exploration of cities (missing aspect) which were relevant to their job search queries. Whereas when using the baseline system, users mostly formulated standard queries as they are used to when utilizing a standard search engine.

4.5. *Participants' Feedback on Beomap*

Users agreed that Beomap helped them to discover relevant topic keywords which enabled further browsing of other related relevant topic regions. The majority of users found Beomap useful *"It saves the time of a user by collecting the information in one place, it's also easy to use and interesting."*, *"Get a lot of info in a nutshell"* or *"fast data analysis"*. Users also appreciated the possibility to define a second ad hoc dimension e.g., *It aggregates tweets relevant to searched topic, and lets you sort the results by different criteria. The system also allows one to add a second topic to the search, and then creates a results map with tweet-to-topic relations. It simplifies the search for specific topics on twitter, and gives a graphic idea of the relation with other trending topics related to the search*. Several participants reported better topics structure understanding as well as an ability to explore relations between individual topics e.g., *"Quick overview of the topic within larger scope, correlation between topics and their relationship"*, *"It simplifies the search for specific topics on twitter, and gives a graphic idea of the relation with other trending topics related to the search"*. Users also found Beomap easy to browse e.g., *Easy to browse interrelated data* and appreciated overall topic map visualization e.g., *The graphical way to present the relevance of the subjects is great.*.

Negative comments of participants were mostly related to the system itself and not to Beomap. A few users perceived the system as not very intuitive (analysis of possible reasons is discussed in Section Subjective evaluation of interfaces) e.g., *needs time to get used to it*. Another user reported a limited number of tweets in the system as a limitation e.g., *It's quite a nice system, perhaps you can update some more topics from twitter*. Several users as well reported that BeomapApp responded slower sometimes e.g., *a bit too slow*.

4.6. *Noun phrases or hashtags*

We performed a crowrdsourced evaluation to find out whether hashtags or noun phrases are perceived as more relevant when exploring a particular topic/query on Twitter data. We have selected 8 different queries. For each query, a participant was asked to view a pair of (Beomaps) topic maps visualizing the underlying query with hashtags or noun phrases only. Further, each participant was asked to view a set of relevant tweets for the query as well as

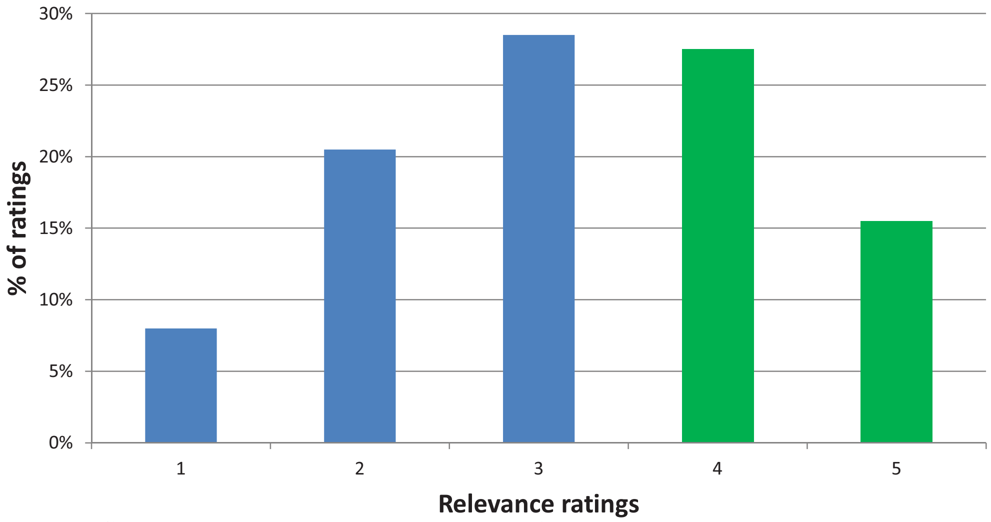

Fig. 3. Green bins correspond to the positive rating of users w.r.t. to Beomaps with noun phrases. Ratings 1 and 2 indicate users preference for Beomaps with hashtags and the rating 3 indicates that users prefer equally both topic maps.

a link to wikipedia page which further described the query topic. Participants were asked to rate which topic map is more relevant w.r.t. to the query on a Likert scale of 1 to 5 (Rating 1 indicates that topic map A is much more relevant than topic map B, Rating 3 represents that both topic maps are equally relevant w.r.t the query, Rating 5 indicates that topic map B is much more relevant than topic map A). We altered assignment of topic maps showing only hashtags or noun phrases to topic map A to avoid participants bias that a topic map A is always more relevant. Further, to get better understanding which phrases/hashtags are considered as (ir)relevant ones, participants were requested to list 5 most relevant terms and 5 most irrelevant terms from both depicted topic maps.

Participants found 6 out of 8 that Beomaps depicting only noun phrases are more relevant than Beomaps showing only hashtags (see Figure). Out of 200 distinct relevance assessments, 86 were positive towards topics maps with noun phrases. 57 ratings were neutral i.e., Beomaps showing either hashtags or noun phrases were equally relevant. Hashtag based Beomaps were considered as more relevant only 57 times.

To further investigate relevance of noun phrases and hashtags for topic maps construction, we analyzed participants answers on top 5 most (ir)relevant terms from both compared topic maps. For instance, when analyzing participant responses for US elections query, common relevant noun phrases are: *primary elections, presidential elections, senate republicans, iowa*. Conversely, participants indicated these terms as irrelevant: *money, time, people*. Participants indicated these hashtags as relevant: *#trump2016, #cruzcrew* and these as irrelevant: *#p2, #imwithher, #love, #media*. It is interesting to observe that some hashtags like: *#gop, #gopdebate, #tcot or #p2* are considered as relevant by some participants while others considers those as irrelevant. These hashtags abbreviations are indeed relevant to the US elections topic (to illustrate: *#gop* stands for Grand Old Party which refers to republicans, *#tcot* stands for top conservatives on twitter).

Beomaps with noun phrases were perceived much more relevant for queries like: *BMW vs Mercedes*, or *Samsung Galaxy vs Iphone 6s*. In these cases, noun phrases represented product models (e.g., *BMW X5, mercedes-benz c-class, samsung galaxy note*), features (e.g., *mercedes e-class interior, iphone charger*) or aspects (e.g., *horsepower, features, radio*) of individual brand or product whereas topic maps depicting only hashtags contained some terms which were not directly related to the query (e.g., *iphone, ios, vinnyapp, vinny*). Conversely, Beomaps with hashtags were perceived as more relevant when visualizing topics related to Super Bowl event. Users found hashtags like: *superbowlsunday, sb50, broncosdefense, gopanthers* relevant whereas some nouns were considers as irrelevant: e.g. *team, ball, game, time.*

The lessons learned from this user study are: (1) Popular hashtags might often be difficult to understand for users who are not familiar with the domain e.g., *#gop, #tcot*; (2) Beomaps with noun phrases are often perceived as more relevant mainly when exploring/comparing brands or products; (3) Hashtags based Beomaps when describing trending event are perceived as more relevant; (4) Although, noun phrases are often perceived as more relevant and at the same time obtain greater coverage of tweets than hashtags, ranking mechanisms should be leveraged to filter out general, common phrases; These findings should be used to devise an approach to intelligently combine noun phrases with hashtags according to user's needs (e.g., how familiar he/she is with the domain, a type of information retrieval task, a need for high coverage of the underlying corpus).

Limitations: Although, crowd-sourced evaluation provides insights about relevance of noun phrases and hashtags based Beomaps, a user study should be performed when users interactions would be analyzed to identify users preferences for hashtags and noun phrases based topic maps.

5. Use Cases

In this section we will exmplify two use cases of Beomap: *location aware exploration* and *sentiment based exploration.* In the use cases we are using data which was harvested form twitter in the period from 23rd January to 1st March 2016. The harvest contains 56 622 936 tweets. We only focused on tweets in english language and to ensure that we get only english text for noun phrases analysis we have set a bounding box over USA and also set the constraint for english language when crawling the Twitter data. For location aware use case we have further constrained the area to New York. The underlying tweets collection was geo tagged with the location within the bounding box for New York. The specific GPS location for individual neighborhoods was downloaded from Flickr Geo Location service.

5.1. *Location Aware Exploration*

Location based topic maps enable to explore, browse and analyze topics specific for the particular location as well as analysis of correlations between topics i.e., which topics are common for two locations, which topics are more specific for a particular location. The possible benefits of such location based topic maps can be used for market research, government and healthcare purposes as well as for exploration of social events in the locations of user interests.

Figure 4 depicts an example of a user interface where hashtags of discussion messages are shown in the two dimensions related to location proximity. User may choose for example one or two areas provided in the selection boxes down to the subareas and explore what are discussions and how they relate to different parts of the query. In the case of Figure 4, the Beomap shows a comparison of hastags related to hiring in Manhattan and Bronx. We can see that the discussions are different in both areas. While in Manhattan (hastags closer to the vertical axix), users are typically discussing hiring related to finance, marketing, banking, creative groups, starbucks, barista, it, and similar, in Bronx (hastags closer to the

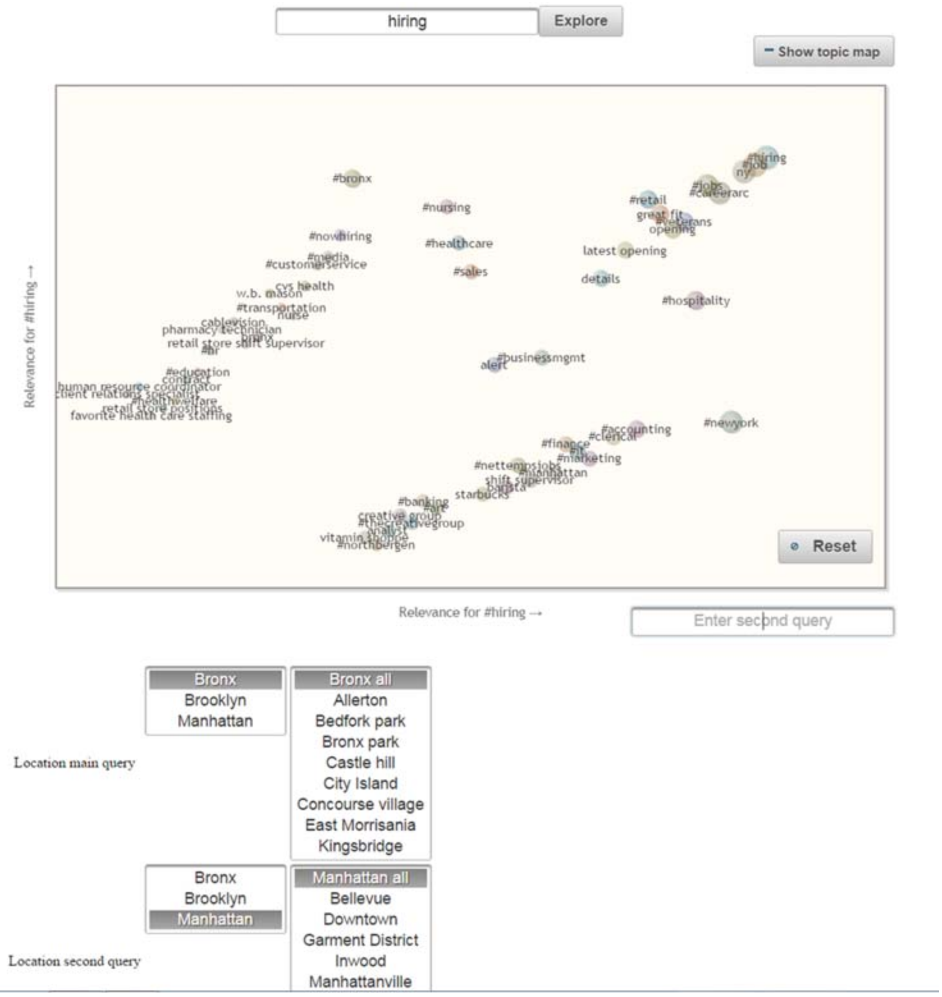

Fig. 4. Beomap for comparing hiring related hashtags in Manhattan and Bronx.

horizontal axis), users are discussing hiring related to health, transportation cablevision, retail and related. This is explainable by the population well being in the two different areas. We can also see that the hashtags in common in both areas related to hiring (hastags on in the right top corner of the map) are of a general nature not so much related to hiring, such as latest opening, veterans, but also general job hiring without mention of the hiring area.

Let us now look at how the Beomap changes when we would like to dig into more specific function such as hiring in management and whether there are some management related hirings in Bronx for example. Figure 5 depicts such a case. We can see that the dimmension for Manhattan and management does not change that much, but we can see that there are also some hirings for management related positions in Bronx but they are more related to retail store, such as retail store shift supervisor or action management developer advisor. But as can be seen these are only sparse examples for management in that area.

On the other hand, if we compare Brooklyn with Bronx on the Beomap 6, the differences

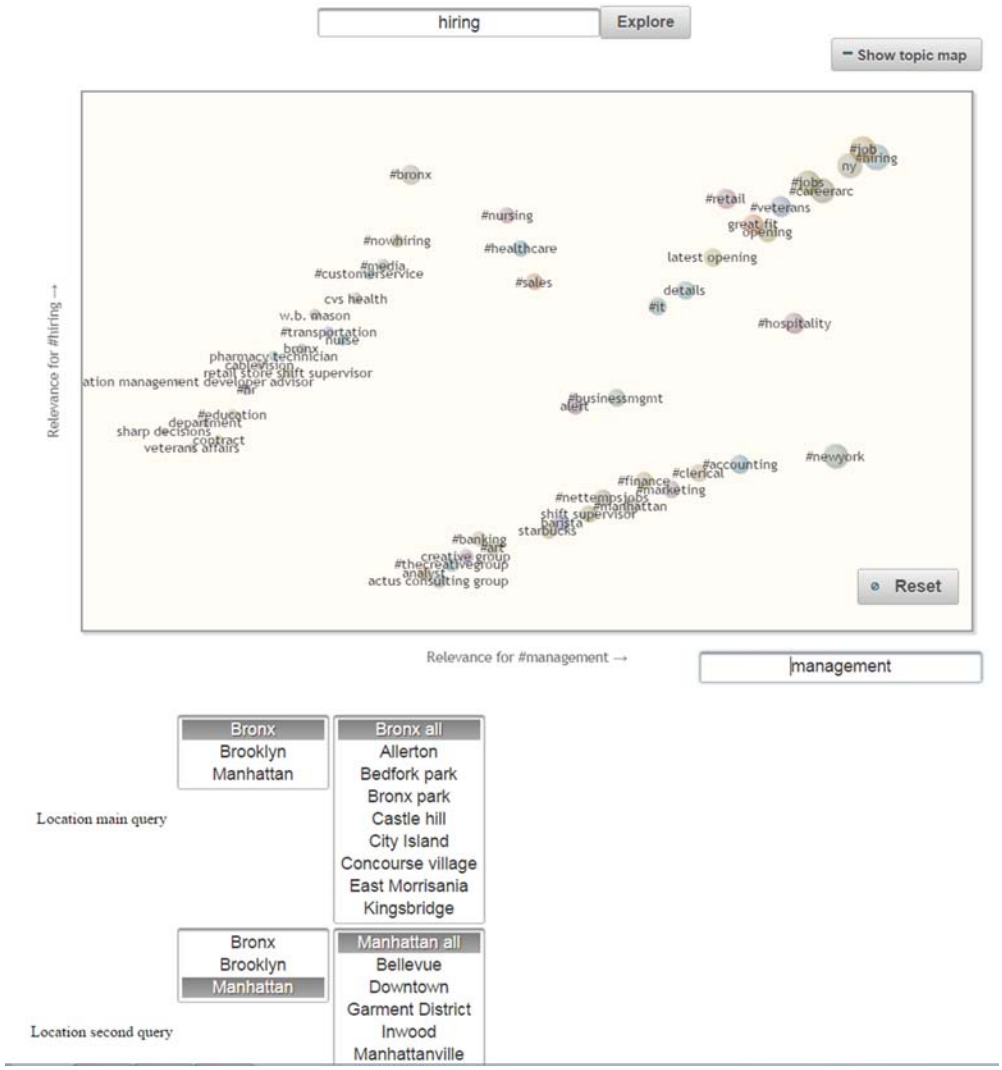

Fig. 5. Beomap for comparing hiring and management related hashtags in Manhattan and Bronx.

are not that big. In both area people are rather discussing management in relation to retail, though in Brooklyn, there are some mentions of banking as well. Also if we look in the area of right top corner of the map, there are more hastags in common than in the case of Brooklyn and Manhattan.

In general, very common topics for majority of neighborhoods in New York were *#job*, *#jobs*. Hence, we tried to analyze if there are some job patterns depending on a particular neighborhood and which are common in general. For instance, when comparing job related hashtags for downtown in Manhattan and Crown Hights in Brooklyn, common topics are *#clerical*, *#retail*, *#sales*, *#nursing*, specific topics for Manhattan are *#it*,*#investementbanking*, *#admin*, *#businessmgmt*, *#interpreter* while for Crown Hights in Brooklyn distinct topics

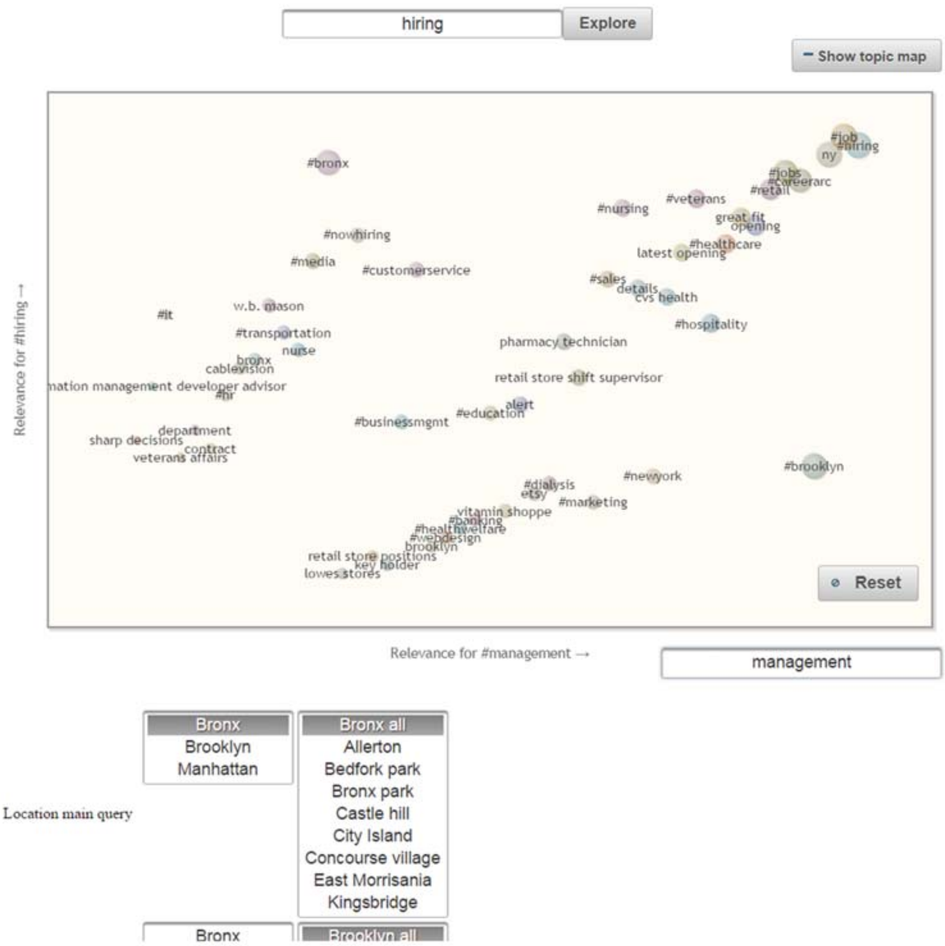

Fig. 6. Beomap for comparing hiring and management related hashtags in Brooklin and Bronx.

are *#transportation*,*#education*, *#veteranjob*, *#marketing*.

5.2. *Sentiment Based Exploration*

Sentiment based Beomap takes the sentiment into account when displaying noun phrases and hashtags in the map. This means that noun phrases and hashtags are located on the map relative not only according to relevance but also polarity. There are two polarities to select from: positive and negative. This means that the axis in the positive case represent the level of positiveness in the messages under certain phrase or hashtag and in negative case they represent the level of negativeness. It can also happen that the same hashtags can be shown in the same position when there is comparable number of negative and positive messages found under the same noun phrase or hashtag.

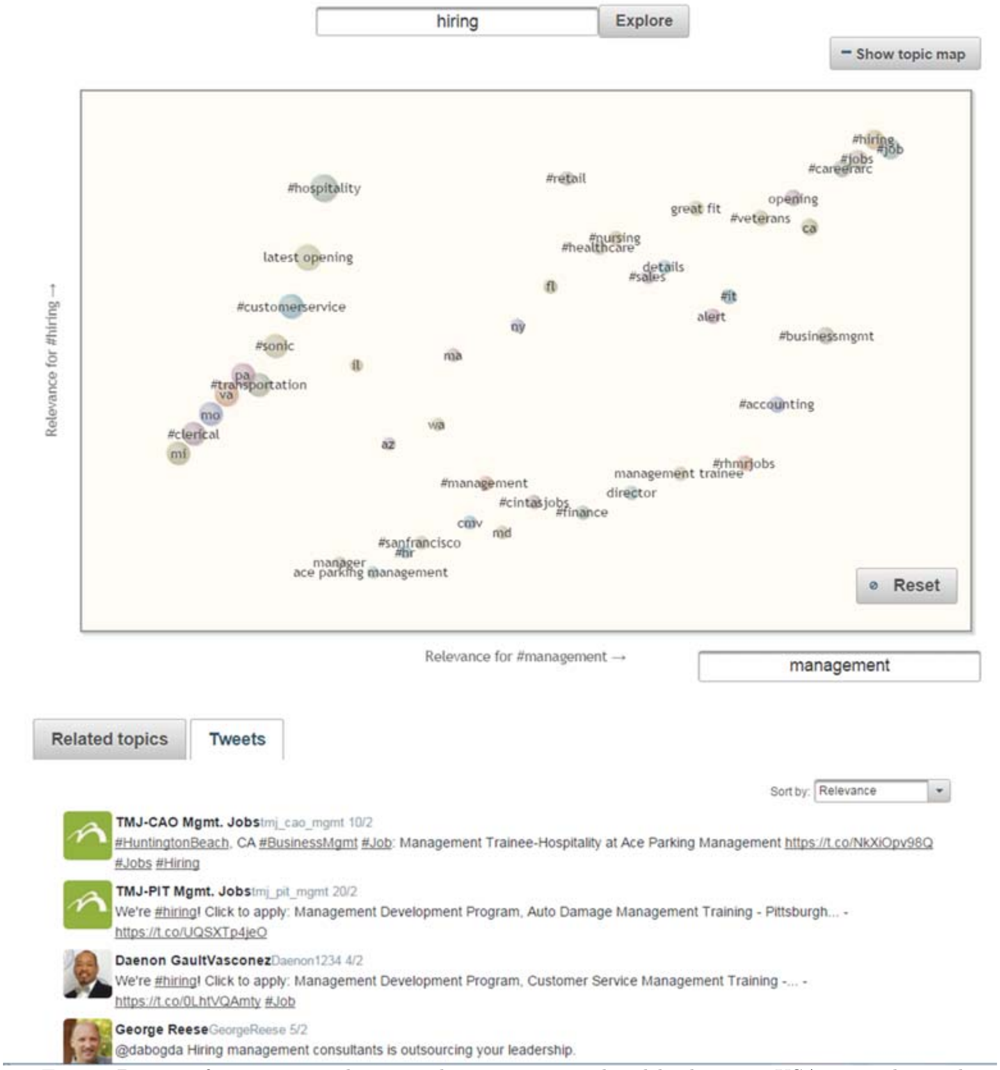

Fig. 7. Beomap for comparing hiring and management related hashtags in USA area when only relevance metric is used.

First, let's look at the Beomap at 7 where we have issued the same query as in previous location aware map and selected metric for relevance. We can see that in comparison to location based figures presented above, the distribution of phrases and hashtags on the user interface have slightly changed. We can also see that there are some different hashtags and some hashtags are also missing. We can see that there are more general hashtags or noun phrases refering to other states appearing (such as MA, NY, WA, FL, IL, and so on) in the map confirming that the data is from the whole USA. Further, there are other general tags appearing such as management or San Francisco. On the other hand, job specific hashtags related to concrete areas such as Bronx and Manhattan are missing.

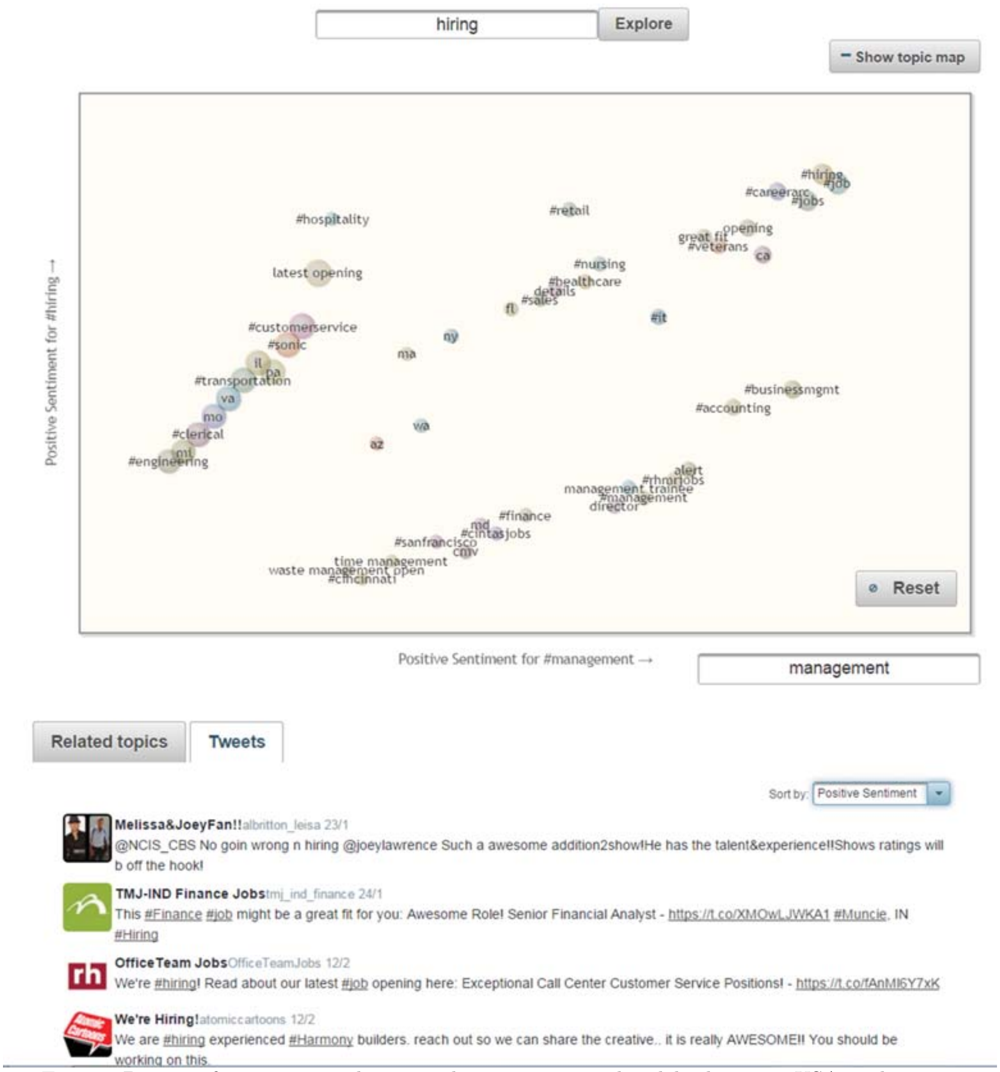

Fig. 8. Beomap for comparing hiring and management related hashtags in USA with positive sentiment metric on.

Now, we can also look at the Beomap which will add the positive sentiment bias to the relevance (figure 8). We can see that in this case the shape induced by the hashtags and noun phrases changed again. We can see for example that hashtags such as *#nursing* and *#healthcare* in the top right corner or *#management* in the bottom middle part have been alligned a little more towards hiring dimension with positive sentiment. On the other hand tags as *#il* moved a bit more towards right on the map. Additional hastags have appeared in the positive sentiment map such as *#engineering*, *#time management*, *#waste management* or *#finance*. In this was we can see, which of the discussions or topics contribute with some messages with positive polarity and look at them in more detail.

Figure 9 depicts the map with negative sentiment metric on. As can be seen again, in

Fig. 9. Beomap for comparing hiring and management related hashtags in USA with negative sentiment metric on.

comparison to figure 8 it adds some tags such as *pain management*, *anger management classes* or *a*nger management. We can also see that general tags representing some state in USA *ma* or *N*Y moved more towards hiring which means that discussions with negative sentiment are more related to general hiring than management. This way we can see how the negative polarity influences the map and also where the discussions with negative sentiment appears.

6. Discussions

Although we only explored the two-dimensional Beomap, the general idea of a topic map interface with ad hoc dimension and the possibility to visualize topics space with user-chosen metrics can be applied to many other systems. The generalization of this interface allows the user to exploit Beomap on top of different types of data e.g., social media data(Facebook, Flickr, Youtube, etc.), folksonomy data (Delicious, Bibsonomy, etc.), news articles or research publications. The current implementation of Beomap includes two dimensions for search interface but it is possible to extend the map into k-dimensions with k different ad hoc dimensions and user-defined metrics. Such a multi-dimensional topic map would enable users to examine their information need from multiple distinct missing aspects with different user-chosen metrics dimensions (e.g., putting sentiment polarity together with topical dimensions). It is also worth noting that the BeomapApp system has the potential to support many more applications than we explored. The performed evaluation of Beomap is based mostly on Relevance, Popularity or Recency metrics. However, the system can be easily extended with several different ad hoc metrics. Further, suitable machine learning algorithms might be exploited to introduce new ad hoc metrics for an enhanced visualization of the topic map. To facilitate a selection of a missing aspect, simple analytics could be developed to test different alternatives of queries and to provide a user with the quantity of items matching each alternative.

Limitations: Our evaluation also reveals two limitations of Beomap. The first is when underlying items are not densely annotated with topic keywords e.g., a few hashtags on Twitter, a limited number of user tags in folksonomy data, etc. This might prevent users from finding and retrieving documents which are relevant but are not annotated with a topic keyword. To minimize this, underlying items might be annotated with named entities or noun phrases or different machine learning based algorithms could be exploited for items annotation. The second limitation is that the controlled nature of the study prevents drawing more reliable conclusions when used by users over a longer period of time with real information needs.

7. Conclusions and Future Work

In this work, we proposed a novel interface (i.e., Beomap) for visualizing text information which allows a user to define ad hoc semantic dimensions with keyword queries. This feature not only makes the dimensions for visualizing topics in text more meaningful, but also naturally enables a user to explore the information space flexibly with arbitrary keyword queries. The topic map is two-dimensional with metric dependent visualization i.e., popularity, recency, proximity. The benefit of the topic map is the ability to steer browsing and exploration of underlying information space by the user through ad hoc queries into both topic map dimensions. Besides supporting ad hoc query-defined dimensions, the two-dimensional Beomap can also naturally support other metrics such as popularity,recency, proximity, etc., making it easy to do visual analytics. We evaluated the proposed Beomap in two ways. First, we used simulated user interaction with search results to compare the Beomap with multiple baselines representing state of the art interactive information exploration interfaces. The results show that Beomap improves the retrieval of additional tweets, the best trade-off between presicion versus user effort is attained when Beomap is generated with respect to the missing aspect of the query. Second, we conducted a user study to compare the BeomapApp system with a standard interface system for performing two information seeking tasks. The results show BeomapApp is perceived as more useful and that users were more satisfied with the system than the baseline. Further, they perceive the system as a flexible tool for exploration, browsing and analysis of Twitter data, and that is easy to browse, and interesting as well.

In the future, we plan to explore several directions. First, we will further evaluate the proposed Beomap with more users by deploying our prototype system on Amazon AWS (a grant is already approved for supporting this system) and making it available potentially for a large number of users Second, we will explore other social media such as Flickr, Instagram or folskonomy data, etc. Finally, the idea of allowing a user to define ad hoc semantic dimensions is not restricted to two dimensions.It would thus also be interesting to explore it with more than two dimensions, which can be particularly useful for visual text analytics.

1. J. C. Acevedo-Aviles, W. M. Campbell, D. C. Halbert, and K. Greenfield. Vizlinc: Integrating information extraction, search, graph analysis, and geo-location for the visual exploration of large data sets.
2. M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi. Eddi: interactive topic-based browsing of social status streams. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 303–312. ACM, 2010.
3. P. Borlund and S. Dreier. An investigation of the search behaviour associated with ingwersens three types of information needs. *Information Processing & Management*, 50(4):493–507, 2014.
4. M. Bron, J. Van Gorp, F. Nack, M. de Rijke, A. Vishneuski, and S. de Leeuw. A subjunctive exploratory search interface to support media studies researchers. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 425–434. ACM, 2012.
5. D. H. Chau, A. Kittur, J. I. Hong, and C. Faloutsos. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 167–176. ACM, 2011.
6. F. C. T. Chua, W. W. Cohen, J. Betteridge, and E.-P. Lim. Community-based classification of noun phrases in twitter. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1702–1706. ACM, 2012.
7. L. Derczynski, A. Ritter, S. Clark, and K. Bontcheva. Twitter part-of-speech tagging for all: Overcoming sparse and noisy data. In *RANLP*, pages 198–206, 2013.
8. A. El-Kishky, Y. Song, C. Wang, C. R. Voss, and J. Han. Scalable topical phrase mining from text corpora. *Proceedings of the VLDB Endowment*, 8(3):305–316, 2014.
9. D. Glowacka, T. Ruotsalo, K. Konuyshkova, S. Kaski, G. Jacucci, et al. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 117–128. ACM, 2013.
10. K. Lagus, T. Honkela, S. Kaski, and T. Kohonen. Self-organizing maps of document collections: A new approach to interactive exploration. In *KDD*, volume 96, pages 238–243, 1996.
11. M. Leginus, L. Derczynski, and P. Dolog. Enhanced information access to social streams through word clouds with entity grouping. In *Proceedings of the conference on Web Information Systems and Technologies (WEBIST)*, 2015.
12. M. Leginus, C. Zhai, and P. Dolog. Beomap: Ad hoc topic maps for enhanced exploration of social media data. In P. Cimiano, F. Frasincar, G. Houben, and D. Schwabe, editors, *Engineering the Web in the Big Data Era - 15th International Conference, ICWE 2015, Rotterdam, The Netherlands, June 23-26, 2015, Proceedings*, volume 9114 of *Lecture Notes in Computer Science*, pages 200–218. Springer, 2015.
13. A. M. Lund. Measuring usability with the use questionnaire. *Usability interface*, 8(2):3–6, 2001.
14. I. Ounis, C. Macdonald, J. Lin, and I. Soboroff. Overview of the trec-2011 microblog track. In *Proceeddings of the 20th Text REtrieval Conference (TREC 2011)*, 2011.
15. M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.
16. O. Tsur and A. Rappoport. What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 643–652. ACM, 2012.
17. X. Wang, B. Tan, A. Shakery, and C. Zhai. Beyond hyperlinks: organizing information footprints in search logs to support effective browsing. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 1237–1246. ACM, 2009.
18. K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing.

In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 401–408. ACM, 2003.