

Leveraging the Semantics of Tweets for Adaptive Faceted Search on Twitter

Fabian Abel¹, Ilknur Celik¹, Geert-Jan Houben¹, Patrick Siehndel²

¹ Web Information Systems, Delft University of Technology
{f.abel,i.celik,g.j.p.m.houben}@tudelft.nl

² L3S Research Center, Leibniz University Hannover
siehndel@l3s.de

Abstract. In the last few years, Twitter has become a powerful tool for publishing and discussing information. Yet, content exploration in Twitter requires substantial effort. Users often have to scan information streams by hand. In this paper, we approach this problem by means of faceted search. We propose strategies for inferring facets and facet values on Twitter by enriching the semantics of individual Twitter messages (tweets) and present different methods, including personalized and context-adaptive methods, for making faceted search on Twitter more effective. We conduct a large-scale evaluation of faceted search strategies, show significant improvements over keyword search and reveal significant benefits of those strategies that (i) further enrich the semantics of tweets by exploiting links posted in tweets, and that (ii) support users in selecting facet value pairs by adapting the faceted search interface to the specific needs and preferences of a user.

Key words: faceted search, twitter, semantic enrichment, adaptation

1 Introduction

The broad adoption and ever increasing popularity of Social Web have been reshaping the world we live in. Millions of people from all over the world use social media for sharing masses of (user-generated) content. This data, for example from social tagging or (micro-)blogging, is often unstructured and is not in compliance with the Semantic Web standards. Research efforts aiming at transforming social data into RDF data such as DBpedia [3], and services like revyu.com do exist along with other pioneer exceptions like Semantic Media Wiki³, semantic tagging [1], and semantic (micro-)blogging (SMOB [2]). However, the big players often do not adhere to Semantic Web principles. For instance, on Twitter, the most popular microblogging service on the Web, the content of Twitter messages (tweets) is not semantically described, which has a negative impact on search. Even though Twitter does allow for metadata⁴, this metadata is for describing the *context* of a tweeting activity; e.g. location of the user, Twitter client from

³ http://semantic-mediawiki.org/wiki/Semantic_MediaWiki

⁴ http://dev.twitter.com/pages/annotations_overview

which the user tweeted, date and time of the tweet and so on. Yet, there is still a lack of tools to automatically enrich the semantics of tweets and fill those meta-data fields with semantics in order to describe the *content* of a Twitter message. The lack of semantics and structure makes searching and browsing on Social Web applications like Twitter a really challenging task.

Although considerable amount of research has been directed towards Twitter recently, search on Twitter has not been studied extensively yet which motivates, for example, the TREC 2011 track on Microblogs that defines the first search tasks on Twitter⁵. In line with the TREC research objectives, we investigate ways to enhance search and content exploration in the microblogosphere by means of faceted search. In an open and enormous network like Twitter, users may get lost, become de-motivated and frustrated easily with the information overload. Hence, there is a need for an effective personalized searching option from the users' point of view that would assist them in following the optimal path through a series of facets to find the information they are looking for, while providing a structured environment for relevant content exploring. In this paper we propose and evaluate an adaptive faceted search framework for Twitter. We investigate how to extract facets from tweets, how to design appropriate faceted search strategies on Twitter and analyze the impact of the faceted search strategy building blocks on the search performance by means of an automated evaluation framework for faceted search. Our main contributions can be summarized as follows⁶.

Semantic Enrichment To allow for faceted search on Twitter, we present methods for enriching the semantics of tweets by extracting facets from tweets and related external Web resources that describe the content of tweets.

Adaptive Faceted Search Framework We introduce different building blocks that allow for various faceted search strategies for content exploration on Twitter and propose methods that adapt to the interests/context of a user.

Evaluation Framework We present an evaluation environment based on an established model for simulating users' click behavior to evaluate different strategies of our adaptive faceted search engine for Twitter. Given this, we prove the effectiveness of our methods on a large Twitter dataset of more than 30 million tweets. We reveal the benefits of faceted search over keyword search and investigate the impact of the different building blocks of our adaptive faceted search framework on the search performance.

2 Background

Twitter is the second most popular social media application which has experienced exponential growth over the last few years in terms of number of users and tweets published. A recent report shows that one billion tweets are published in

⁵ <http://sites.google.com/site/trecmicroblogtrack/>

⁶ Our adaptive faceted search framework, the code of our evaluation framework and the dataset are available via: <http://wis.ewi.tudelft.nl/iswc2011/>.

a week which corresponds to an average of 140 million tweets per day⁷. This astounding growth and popularity of the microblogging service have naturally been attracting significant amount of research from various perspectives and fields lately. In this section, we present the background and related work that help to understand the usage dynamics and semantics of Twitter messages and motivations for faceted search on Twitter.

2.1 Understanding Twitter Semantics

Tweets are distinctively short text messages of maximum 140 characters that do not explicitly feature facets, in addition to being too short to extract meaningful semantics from. Furthermore, the language and syntax of tweets are significantly different than other Web documents, since Twitter users tend to use abbreviations and short-form for words to save space, as well as colloquial expressions, which make it even harder to infer semantics from tweets.

Mining the semantics of tweets could lead to interesting applications. For instance, Twitris 2.0, a Semantic Web application, facilitates understanding perceptions from social media by capturing semantics with spatial, temporal, thematic dimensions, user intentions and sentiments, and networking behavior from Twitter [8]. Following a top-down approach, Stankovic et al. mapped tweets to conference talks and exploited metadata of the corresponding research papers to enrich the semantics of tweets in order to better understand the semantics of the tweets published in conferences [9]. We follow a similar approach to this where we try to leverage the semantics of tweets for enhancing search on Twitter. Therefore, instead of a restricted domain like scientific conferences, we try to enrich the tweets in general.

Studies on the social network of Twitter and information diffusion dynamics show that tweets are often news related. For instance, Kwak et al. showed that the majority of the trending topics and 85% of all the posted tweets in Twitter are related to news [4]. Sankaranarayanan et al. investigated the use of Twitter to build a news processing system from tweets, called TwitterStand, by capturing tweets that correspond to late breaking news [10]. Some researchers differentiated between news and casual information by investigating the credibility of news propagated through Twitter, while others studied the information propagation via re-tweeting during emergency events [11]. Building on such studies which revealed that Twitter is used more as a news media than a social network [4], and identified “information seekers” as a primary category of Twitter users [7], we try to map tweets to news articles on the Web over the same time period in order to enrich them and to allow for extracting more entities to generate richer facets for search.

Another distinct characteristic of the Twitter syntax is the use of hashtags. Hashtags are meant to be identifiers for related messages of the same topic. By including a hashtag in a message, users indicate to which conversations their

⁷ <http://blog.twitter.com/2011/03/numbers.html>

message is related to. Due to the unorganized and fragmented streams of information in Twitter, the use of hashtags has become the means of creating threads of conversations and gathering those serving for a particular interests. When used appropriately, searching on hashtags would return messages that belong to the same conversation. Huang et al. studied the use of hashtags and tagging behavior in Twitter in comparison to Delicious, where they found that hashtags are often just meaningful for a short period of time, and described tagging in Twitter as “conversational” [14].

2.2 Search on Twitter

Since Twitter has become an important source of information for late-breaking news, Twitter posts are already being exploited by major search engines such as Google and Bing. The simplicity of Twitter is one of its powers that has played an important role in its success. However this simplicity brings about negative effect when it comes to searching, browsing or mining the Twitterverse for various uses. Aggregating functions are limited to filtering tweets by users or hashtags, or restricting by keywords, organized by time and not by relevance [12]. Our work is motivated by the inaccuracies of the current keyword search option and the lack of semantics in tweets that hinders a better browsing experience.

Searching and browsing are indeed limited in Twitter. For example, one can search for tweets by a keyword or by a user in a timeline that would return the most recent posts. So, if a user wants to see the different tweets about a field of sports, and were to search for “sports” in Twitter, only the recent tweets that contain the word “sports” would be listed to the user. Many tweets that do not contain the search keyword, but are about different sport events, sport games and sport news in general, would be filtered out. This keyword search is not only imprecise, but is also missing out on a number of messages that do not contain the particular keyword. As tweets are unconventionally short and do not contain explicit meanings, searching microblogging platforms and making sense of the streams of messages passing through the system become even more challenging.

On the other hand, semantic search augments and improves traditional search results by using data from the Semantic Web. Guha et al. described two semantic search systems and outlined how the semantics of search terms can be used for improving search results [15]. We follow a similar approach to adding explicit semantics in order to improve search by extracting entities from tweets and linking external Web sources to tweets in order to enhance their semantics.

A systematic overview of search behavior on Twitter and what differentiates it from Web search was presented by [16]. Researchers investigated why people search Twitter and found out that people mainly search socially generated content to find temporally relevant information (e.g. breaking news, traffic jams etc.) and social information (e.g. opinion, general sentiment, information related to people of interest), as well as to “monitor” the associated results. It was also noted that Twitter search queries are shorter, more popular and less likely to evolve as part of a session, whereas Web queries change and develop during a session to “learn” more about a topic. We take the search behavior of the users

into account when we develop strategies for our faceted search framework, such as time sensitive or personalized rankings of facet values.

2.3 Faceted Search

Faceted search is becoming a popular method to allow users to interactively search and navigate complex information spaces. Faceted search systems help people find what they are looking for by allowing them to specify not just keywords related to their information needs, but also metadata which is used for query refinement. Hearst defined facets as “a set of meaningful labels organized in such a way as to reflect the concepts relevant to a domain” [18]. Koren et al. defined three common characteristics for faceted search interfaces; (i) facets and facet-values, (ii) previous search results, and (iii) the current query [17]. By choosing from the suggested facet-values, a user can interactively refine the query. Traditional faceted search interfaces allow users to search for items by specifying queries regarding different dimensions and properties of the items (facets) [19]. For example, online stores such as eBay⁸ or Amazon⁹ enable narrowing down their users’ search for products by specifying constraints regarding facets such as the price, the category or the producer of a product. In contrast, information on Twitter is rather unstructured and short, which does not explicitly feature facets. This puts constraints on the size and the number of keywords, as well as facets that can be used as search parameters without risking to filter out many relevant results.

As a solution, we enrich the semantics of tweets by extracting facets and assigning semantics to them, which allows for a rather semantic faceted search than a keyword search. For instance, given a tweet like “*Off to BNP Paribas at Indian Wells*”, entities such as “*BNP Paribas*” and “*Indian Wells*” are extracted and assigned to facet types such as “SportsEvents” and “Locations” respectively, which allows for searching in different dimensions (multiple facets) even though the words like “sport”, “event” or “location” are not included in the tweet (see Figure 1(a)).

2.4 Problem Formalization

On Twitter, facets describe the properties of a Twitter message. For example, persons who are mentioned in a tweet or events a tweet refers to. Oren et al. [19] formulate the problem of faceted search in RDF terminology. Given an RDF statement (*subject, predicate, object*), the faceted search engine interprets (i) the subject as the actual resource that should be returned by the engine, (ii) the predicate as the facet type and (iii) the object as the facet-value (restriction value). We follow this problem formulation proposed by Oren et al. [19] and interpret tweets as the actual resources (subjects) which the faceted search engine should return, entities that are mentioned in a tweet as facet value and the type of an entity as facet type.

⁸ <http://ebay.com/>

⁹ <http://amazon.com/>

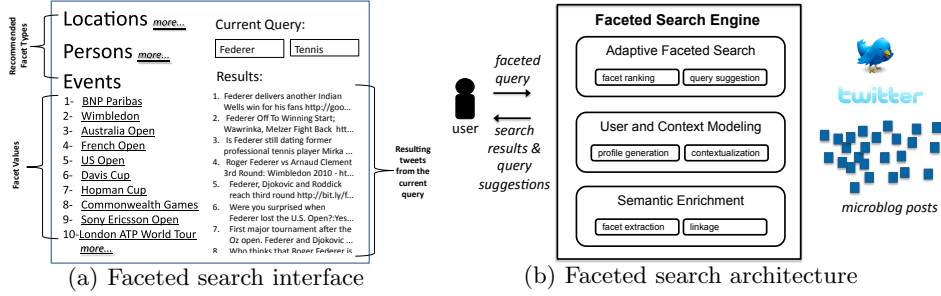


Fig. 1. Adaptive faceted search on Twitter: (a) example interface and (b) architecture of the faceted search engine.

Figure 1(a) illustrates how we envision the corresponding faceted search interface that allows users to formulate faceted queries. Given a list of facet values which are grouped around facet types such as locations, persons and events, users can select facet-value pairs such as $(URI_{event}, URI_{wimbledon})$ to refine their current query $((URI_{person}, URI_{federer}), (URI_{sportsgame}, URI_{tennis}))$. A faceted query thus may consist of several facet-value pairs. Only those tweets that match all facet-value constraints will be returned to the user. The ranking of the tweets that match a faceted query is a research problem of its own (cf. [16]). In this paper, we rank matching tweets according to their creation time, i.e. the older a tweet the lower its ranking. The core challenge of the faceted search interface is to support the facet-value selection as good as possible. Hence, the facet-value pairs that are presented in the faceted search interface (see left in Figure 1(a)) have to be ranked so that users can quickly narrow down the search result lists until they find the tweets they are interested in. Therefore, the *facet ranking problem* can be defined as follows.

Definition 1 (Facet Ranking Problem). *Given the current query F_{query} , which is a set of facet-value pairs (predicate, object) $\in F_{query}$, the hit list H of resources that match F_{query} , a set of candidate facet-value pairs (predicate, object) $\in F$ and a user u , who is searching for a resource r at time t via the faceted search interface, the core challenge of the faceted search engine is to rank the facet-value pairs F . Those pairs should appear at the top of the ranking that restrict the hit list H so that u can retrieve t with the least possible effort.*

3 Framework for Adaptive Faceted Search

The architecture of the engine that we propose for faceted search on Twitter is depicted in Figure 1(b) and features three main components. The *semantic enrichment* layer aims to extract facets from tweets and generates RDF statements that describe the semantic meaning of a Twitter message. In order to adapt the faceted search engine to the people who are using it, we propose *user modeling and context modeling* strategies that infer interests of the users in facets. Based on the semantically enriched tweets and the user profiles inferred by the user

Building Block	Description
Semantic Enrichment	Enriching tweets to extract FVPs for representing tweets: (1) tweet-based enrichment (2) tweet-based and link-based enrichment
User/Context Modeling	Strategies for generating profiles that represent (current) user demands in FVPs: (1) user modeling based on tweets published by a user [6] (2) context modeling based on user context when issuing a query (here: query time)
Adaptive Faceted Search	Strategies for adapting the faceted search interface to the user and context and for ranking FVPs in particular: (1) Occurrence Frequency: ranking based on frequency of a FVP in the tweets (2) Personalization: adapting the FVP ranking to a given user profile (3) Time Sensitivity: adapting the FVP ranking to temporal context (4) Diversification: strategy to increase variety among the top-ranked FVPs

Table 1. Building blocks of faceted search framework: strategies for extracting facet-value pairs (FVPs) from tweets, inferring user interest in FVPs and ranking FVPs.

modeling layer, the *adaptive faceted search* layer solves the actual facet ranking problem. It provides methods that adapt the facet-value pair ranking to the given context and user. Table 1 lists the components and different strategies of the three main building blocks. Below, we explain these building blocks in detail.

3.1 Semantic Enrichment

Twitter messages are short text messages that do not feature facets describing the content of the message. Twitter messages such as “Federer is great <http://bit.ly/2fRds1t>” can be represented in RDF using, for example, SIOC vocabulary¹⁰, the semantic meaning of such messages is however not explicitly defined:

```
<http://twitter.com/bob/statuses/48748435752333312>
  a <sioc:Post> ;
  dcterms:created "2011-07-08T15:52:51+00:00" ;
  sioc:content "Federer is great http://bit.ly/2fRds1t";
  sioc:has_creator <http://twitter.com/bob> ;
  sioc:links_to <http://bit.ly/2fRds1t> ;
```

While the above RDF representation specifies the tweet’s metadata such as the creator of the tweet or the creation time, it requires further enrichment so that the content of a tweet is semantically described as well. Representing the semantics of Twitter messages will allow for semantic search strategies such as faceted search (for casual users) or SPARQL queries (for advanced users and application developers). Our faceted search framework features two core

¹⁰ <http://rdfs.org/sioc/spec/>

strategies for extracting the semantics from tweets: (i) tweet-based enrichment where named entities are extracted from Twitter messages and (ii) tweet-based and link-based enrichment where tweets are further enriched with entities that are extracted from external Web resources that are referenced from the tweets. Therefore, our framework connects to three named entity recognition services: OpenCalais¹¹, DBpedia spotlight¹² and Alchemy¹³. Using our semantic enrichment infrastructure, we can represent the semantics of the above Twitter message:

```
<http://twitter.com/bob/statuses/4874843575233312>
  a <sioc:Post> ;
  ...
  sioc:has_topic <http://dbpedia.org/resource/Roger_Federer> ;
  sioc:has_topic <http://dbpedia.org/resource/Tennis> ;
  sioc:has_topic <http://dbpedia.org/resource/2009_Wimbledon_Championships> .
```

While the relation to *dbpedia:Roger_Federer* can be inferred by merely analyzing the tweet (tweet-based enrichment), inferring that the tweet refers to Federer’s achievements at the Wimbledon tournament 2009 is possible when following the link that is posted in the tweet (link-based enrichment). Our engine uses the identified entities as facet values and exploits the type of the entities to group facet values into facet types. In our evaluation, we process tweets by means of the OpenCalais API which allows us to infer 39 different facet types.

3.2 User and Context Modeling

The goal of the user modeling module is to create a user profile that represents the current demands of the user so that the faceted search interface can be adapted to the inferred profile. Therefore, we define a user profile as a list of weighted facet values (entities):

Definition 2 (User Profile). *The profile of a user $u \in U$ is a set of weighted entities where with respect to the given user u for an entity $e \in E$ its weight $w(u, e)$ is computed by a certain function w .*

$$P(u) = \{(e, w(u, e)) | e \in E, u \in U\}$$

Here, E and U denote the set of entities and users respectively.

In this paper, we apply a lightweight user modeling strategy that weights the entities according to their occurrence frequency in the complete history of tweets which have been published by the user u before she is performing the search activity. The time of a search activity is considered as context and in the subsequent section we introduce a strategy that exploits this feature to adapt the faceted search engine to the temporal context. For more detailed information on user and context modeling strategies that are part of our search framework, we refer the reader to [6].

¹¹ <http://opencalais.com/>

¹² <http://dbpedia.org/spotlight>

¹³ <http://alchemyapi.com/>

3.3 Adaptive Faceted Search

Given the strategies for enriching the semantic descriptions of tweets as well as user and context modeling strategies, the module for adaptive faceted search can operate on semantically rich Twitter items and profiles to solve the ranking task specified above (see Definition 1). Below, we present four ranking strategies that order the facet-value pairs to adapt the faceted search interface to the current context and user.

Occurrence Frequency A lightweight approach is to rank the facet-value pairs $(p, e) \in F$ based on their occurrence frequency in the current hit list H , the set of tweets that match the current query (cf. Definition 1):

$$rank_{frequency}((p, e), H) = |H_{(p,e)}| \quad (1)$$

$|H_{(p,e)}|$ is the number of (remaining) tweets that contain the facet-value pair (p, e) which can be applied to further filter the given hit list H . By ranking those facet values high that appear in most of the tweets, $rank_{frequency}$ minimizes the risk of ranking relevant facet values low. However, this might increase the effort a user has to invest to narrow down search results: by selecting facet values which occur in most of the remaining tweets the size of the hit list is reduced slowly.

Personalization The personalized facet ranking strategy adapts the facet ranking to a given user profile that is generated by the user modeling layer depicted in Figure 1(b). Given the set of facet-value pairs $(p, e) \in F$ (cf. Definition 1), the personalized facet ranking strategy utilizes the weight $w(u, e)$ in $P(u)$ (cf. Definition 2) to rank the facet-value pairs:

$$rank_{personalized}((p, e), P(u)) = \begin{cases} w(u, e) & \text{if } w(u, e) \in P(u) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Diversification The main idea of the diversification strategy is to produce facet rankings for which the highly ranked facet-values lead to diverse subsets of the current hit list H . For example, if a user is searching for news on “Egypt”, based on the frequency, the highly ranked facet-values would be entities such as “Cairo” or “Middle East”, because they appear in most of the resources in the hit list. However, these facet-values may refer to very similar items, i.e. issuing the query “Cairo” on top of “Egypt” will not filter out many more items. Hence, to drill down to a small result set as quickly as possible, it might be more appropriate to display facet value pairs which (i) are more selective and (ii) are diverse from the other facet-value pairs so that users with diverse information needs can be satisfied.

The diversification algorithm that we propose uses occurrence frequency as basis ranking strategy and then reorders the FVPs according to the number of items in the current hit list that (1) match the given FVP and (2) do not match the higher ranked FVPs (see Equation 3 and Equation 4).

$$rank_{diversify}((p, e), H) = rank_{frequency}((p, e), H) + d \cdot diversify((p, e), H) \quad (3)$$

$$diversify((p, e), H) = |H_{(p,e)} \setminus \cup_{i=1}^N H_{(p_i, e_i)}| \quad (4)$$

Here, $d \in \mathbb{R}$ allows for adjusting the influence of the diversification – in this paper we set $d = 1$. N is the number of higher ranked facet value pairs. All items in the hit list which contain higher ranked FVPs are not taken into account for the scoring of the remaining facet value pairs.

Time Sensitivity The time sensitive ranking strategy takes into account the current temporal context (query time) and the publishing time of the tweets that match a facet-value pair. The core idea is to rank those FVPs that recently occurred in tweets (*trending FVPs*) higher than FVPs that constantly are mentioned in tweets. To achieve this, we take the creation times of all tweets that match a facet-value pair (p, e) and calculate the average age of these tweets, i.e. the average distance to the actual query time. For each FVP, we therefore obtain a score that describes how recently the tweets are that match the FVP. The smaller the score – i.e. the younger the matching tweets – the higher the rank. In practice, we combine the time sensitive ranking score with one of the above ranking strategies such as occurrence frequency:

$$rank_{time}((p, e), H) = \frac{d}{\overline{avg}_{age}(H_{p,e})} \cdot rank_{frequency}((p, e), H) \quad (5)$$

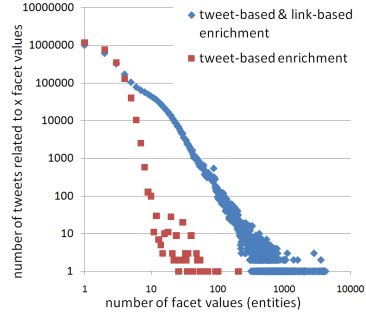
Here, $\overline{avg}_{age}(H_{p,e})$ is the average, normalized age of the tweets in H that match FVP (p, e) . Normalization is done by dividing $avg_{age}(H_{p,e})$ by the maximum average age associated with a FVP in H . The dampen factor $d \in \mathbb{R}$ allows to adjust the influence of the time sensitive score with respect to the ranking score $rank_{frequency}((p, e), H)$. In our experiments, we set $d = 1$ and test the time sensitive scoring method also with other ranking strategies such as $rank_{personalized}((p, e), H)$.

4 Analysis of Facet Extraction

In this section, we analyze the characteristics of a large Twitter corpus of more than 30 million Twitter messages, and investigate how the semantic enrichment of tweets impacts the facet-value pair extraction so that tweets are discoverable by means of faceted search. We collected those tweets by monitoring the Twitter activities of more than 20,000 Twitter users over a period of more than four months starting on November 15, 2010. We started the crawling process by monitoring popular Twitter accounts in the news domain such as the New York Times (*nytimes*) and CNN Breaking News (*cnnbrk*) and then extended the set of accounts in a snowball manner with users who replied or re-tweeted messages of Twitter users whom we followed already. Figure 2(b) shows the distribution of identified entities per tweet. The distribution shows that for most tweets only a very small number of related entities are identified. Nearly 50% of tweets contain only one entity. Around 92% of the tweets contain 3 or less entities. Moreover, Figure 2(b) shows the impact of the semantic enrichment for the tweets. While

Characteristics	Tweet-based enrichment	Tweet & Link-based enrichment
avg. num. of facet values per tweet	1.85	5.72
avg. num. of discoverable tweets	61161.23	75782.76
avg. num. of FVP-selects to filter results	1.95	2.25
avg. size of filtered result set	1685.320	189.48

(a) Impact of link-based enrichment



(b) Facet values per tweet

Fig. 2. Impact of link-based enrichment on (a) the characteristics of tweets and the faceted search settings and (b) the number of facet values per tweet.

the majority of the tweets still contain one, two or three different entities, a significant increase is observed for the number of tweets containing five or more different facet values with the aid of enrichment. In this scenario the number of tweets which are related to more than four facet-value pairs is around 14 times larger when using the semantic enrichment based on link exploitation.

Figure 2(a) overviews some of the characteristics of tweets and the faceted search scenarios for both tweet-based and tweet- and link-based enrichment. It reveals that the number of facet value pairs related to each tweet increases on average when using the semantic enrichment functionality which exploits links to external Web resources. While the tweets contain on average 1.85 facet values, the link-based enrichment strategy features 5.72 facet values per tweet, thus allowing the end-user to find a tweet via many more alternative search paths and faster by drilling down to a smaller set of resulting tweets. The numbers showing discoverable tweets, FVP-selects and the size of the result set are related to the simulated search scenario. Figure 2(a) shows that the link-based-enrichment increases the number of discoverable tweets significantly. This suggests more FVP-selects to drill down the result list. However our evaluations show that the slight increase of 15% more click actions result in a much smaller result set. When using the link exploration strategy, the size of the result set is 9 times smaller, this helps the user to find the tweet(s) of interest faster.

5 Evaluation of Faceted Search Strategies

Having analyzed the characteristics of the facet extraction, we now evaluate the performance of the faceted search strategies proposed in Section 3 and answer the following research questions.

1. How well does faceted search that is supported by the semantic enrichment perform in comparison to keyword search?
2. What strategy performs best in ranking facet-value pairs that allow users to find relevant tweets on Twitter?
3. How do the different building blocks of our faceted search framework (see Table 1) impact the performance?

5.1 Evaluation Methodology

Our evaluation methodology extends an approach introduced by Koren et al. [17] that simulates the clicking behavior of users in the context of faceted search interfaces.

The core evaluation setup consists of parameters describing the user interface itself and algorithms characterizing the simulated user behavior. In general, faceted search user interfaces share some common characteristics and feature at least two parts: an area displaying the facets and a part showing the search results (see Figure 1(a)). Based on such an interface, a user can perform different actions, where the goal is to find a relevant tweet. We consider a tweet as relevant for a user if it was re-tweeted by the user. In a faceted search interface, a user can perform different actions and we focus on the following types of actions: (1) selection of a facet-value pair to refine the query and drill down the search result list, (2) if no appropriate facet-value pair is shown to the user then she can ask for more facet-value pairs and (3) if the user cannot select further facet-value pairs then she has to scan the result set until she finds the relevant tweet. In our simulation, we assume that the user knows the tweet she is looking for and only selects facet-value pairs that match the target tweet.

We model the user’s facet-value pair selection behavior by means of a *first-match user* that selects the first matching facet-value pair. To evaluate the performance, we generated search settings by randomly selecting 1000 tweets that have been re-tweeted. Each search setting consists of (i) a target tweet (= the tweet that was re-tweeted), (ii) a user that is searching for the tweet (= the user who re-tweeted the tweet) and (iii) the timestamp of the search activity (= the time when the user re-tweeted the message). The set of candidate items was given by all those tweets which have been published within the last 24 hours before the search activity. On average, the number of candidate items is 61161.23 for the tweet-based enrichment strategy and 75782.76 for the tweet-based and link-based enrichment strategy (see Figure 2(a)) while there is only one single relevant tweet (target tweet) per search session.

For measuring the performance of our facet ranking strategies, we utilize *Success@k* which is the probability that a relevant facet-value pair, the user selects to narrow down the search result list, appears within the top k of the facet-value pair ranking. This metric is a direct indicator for the effort a user needs to spend using the search interface: the higher *Success@k*, the faster the user will find a relevant facet-value pair when scanning the facet-value pair ranking.

For evaluating the performance of faceted search in comparison with keyword search, we use hashtags as keyword queries and measure the performance by means of the mean reciprocal rank (*MRR*) which indicates at which position we find the target item in the search result ranking.

5.2 Results

Using the evaluation method presented above, we analyze the quality of the search strategies. Figure 3 overviews the results that allow us to answer the research questions raised at the beginning of this section.

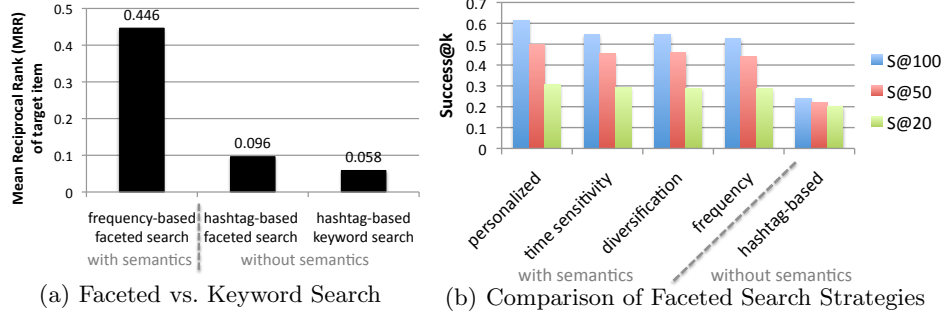


Fig. 3. Overview on results: (a) mean reciprocal rank (MRR) of target item in the search result ranking for faceted search and keyword search and (b) performance of the faceted search strategies for ranking FVPs.

Faceted Search vs. Keyword Search. Figure 3(a) shows that our approach to faceted search clearly outperforms faceted search based on hashtags as well as keyword search. Using tweet-based semantic enrichment for extracting FVPs and occurrence frequency as weighting scheme for ranking FVPs (*frequency-based faceted search*), we achieve an improvement regarding MRR of more than 360% (from 0.096 to 0.446) over *hashtag-based faceted search* where hashtags mentioned in the tweets are exploited as facets. Comparing the semantic faceted search strategy to *hashtag-based keyword search*, where a user issues a single hashtag as a query, shows actually an improvement regarding MRR of more than 660% (from 0.058 to 0.446). Furthermore, it is important to state that the results shown in Figure 3(a) are based on those 28% of the search settings for which the target tweet contains at least one hashtag¹⁴. For the remaining search settings, hashtag-based strategies fail which further proves that the semantic enrichment of our faceted search framework is highly beneficial and important for search on Twitter.

Comparison of Strategies for Ranking FVPs. Figure 3(b) gives an overview of the performance of the different facet ranking strategies measured by Success@100, Success@50 and Success@20. Again, we observe that the hashtag-based faceted search strategy, which exploits hashtags as FVPs and applies occurrence frequency as weighting scheme, is clearly outperformed by the semantic faceted search strategies provided by our framework. For example, the tweet-based semantic enrichment in combination with occurrence frequency as weighting scheme (*frequency*) improves over the *hashtag-based* baseline by 121.5, 99.5% and 42.4% regarding S@100, S@50 and S@20 respectively. For the faceted search strategies that make use of semantic enrichment, we observe that the *personalized* ranking strategy outperforms the other strategies for all metrics. When looking at the Success@100, the personalized strategy performs approximately 12% better than the other three strategies. Knowing the preferences of a user for certain topics, which are modeled via the FVPs, thus brings advantages for

¹⁴ Given the more than 30 million tweets of our dataset, we actually observe that just 19.82% of the tweets mention a hashtag.

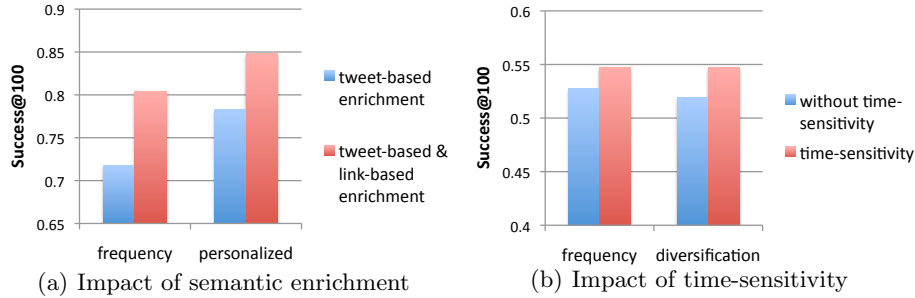


Fig. 4. Impact of (a) semantic enrichment strategy on the frequency-based and the personalized strategies for search settings where the target tweet contains a link and (b) impact of time-sensitive re-ordering on the performance of the frequency-based and the diversification-based strategies.

adapting the faceted search interface to the user who is searching for a tweet. Furthermore, knowing the user’s temporal context also improves the search performance slightly (see *time sensitivity* in Figure 3(b)). However, the differences between the ranking strategies based on time, diversification and frequency are very small which might be caused by the fact that both the time sensitive strategy and the diversification use occurrence frequency as basic weighting function. Hence, reducing the influence of the frequency-based scoring on these strategies could possibly lead to further improvements.

Impact of the different Building Blocks on Faceted Search. Figure 4 illustrates the impact of some of the building blocks of our framework on the faceted search performance. In Figure 4(a), we compare the performance of the frequency-based strategy and personalized strategy when doing (i) semantic enrichment solely on tweets (*tweet-based enrichment*) or (ii) semantic enrichment by analyzing both the content of the tweets and Web resources that are linked from the tweets (*tweet-based & link-based enrichment*). It shows that the link-based enrichment significantly improves the Success@100 for both strategies. Hence, while semantic enrichment by means of named entity recognition in tweets improves already the faceted search performance over faceted search based on hashtags, we achieve further improvements if we follow the links posted in Twitter messages to further describe the semantic meaning of a Twitter message. Furthermore, the improvement gained by personalization are consistent through the different enrichment strategies. For example, the personalized strategy improves over the frequency-based strategy by 8.3% when link-based enrichment is conducted.

Figure 4(b) shows how the Success@100 rates for the frequency and the diversification based rankings increase when the temporal context of the search activity is taken into account to adjust the ranking of the FVPs. One can see improvements for both strategies with 3.8% for the frequency-based strategy and 5.3% for the diversification strategy.

Synopsis. Given these observations, we now revisit our research questions raised at the beginning of this section. We showed that faceted search clearly outperforms hashtag-based keyword search. Using the our faceted search framework, we achieve a more than eight times higher MRR than keyword search. In response to which strategy performs the best for ranking facet-value pairs, we revealed that the personalization strategy – i.e. adapting the facet-value pair ranking to the interest profile of the user who is searching for a tweet – performs best for faceted search on Twitter. Furthermore, we showed that the different building blocks of our faceted search framework all have positive impact on the facet-value pair ranking in order to answer our second question. Semantic enrichment by means of exploiting both tweets and Web resources that are referenced from the tweets increases the number of tweets that are discoverable via faceted search. It also increases the number of facet-value pairs per tweet so that users have more alternatives in narrowing down the search result list. Moreover, we showed that time-sensitivity – i.e. adapting the facet-value pair ranking to the temporal context – improves the performance of the facet-value pair ranking so that users can find their intended tweets faster and with less effort.

6 Conclusions

In this paper, we tackled the problem of searching for relevant messages on Twitter. We introduced an adaptive faceted search framework that features semantic enrichment of tweets as a solution to this problem. Our framework allows adding semantics to tweets by extracting entities and enriching them with external resources in order to create facets (e.g. persons, locations, organizations etc.) and facet-values that describe the content of tweets. To support users in selecting facet-value pairs during their faceted search activities, we studied different strategies that adapt the ranking of facet-value pairs to the user and context (e.g. temporal context).

We presented an evaluation framework that allows for simulating users' search behavior and applied this simulator on a large Twitter dataset of more than 30 million tweets. Our evaluation proves the effectiveness of our strategies and reveals that our faceted search framework achieves tremendous improvements in comparison with hashtag-based keyword search. Moreover, we see that personalization and context-adaptation gain the best performance among the faceted search strategies. Our analysis of semantic enrichment strategies also showed that the exploitation of links that are posted in Twitter messages is beneficial for describing the semantic meaning of tweets and therefore improves the search performance as well.

Acknowledgements The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) in context of the ImREAL project¹⁵ and the SYNC3 project¹⁶.

¹⁵ <http://imreal-project.eu>

¹⁶ <http://sync3.eu>

References

1. Passant, A., Laublet, P.: Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In: Workshop on Linked Data on the Web, Beijing, China. (2008)
2. Passant, A., Hastrup, T., Bojars, U., Breslin, J.: Microblogging: A Semantic Web and Distributed Approach. In: Workshop on Scripting For the Semantic Web, Tenerife, Spain. Volume 368, CEUR-WS.org (2008)
3. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: ISWC, Busan, Korea, Springer (2007) 715–728
4. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: WWW, ACM (2010) 591–600
5. Bernstein, M., Kairam, S., Suh, B., Hong, L., Chi, E.H.: A torrent of tweets: managing information overload in online social streams. In: CHI Workshop on Microblogging: What and How Can We Learn From It? (2010)
6. Abel, F., Gao, Q., Houben, G.J., Tao, K.: Analyzing User Modeling on Twitter for Personalized News Recommendations. In: UMAP, Girona, Spain, Springer (2011) 1–12
7. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: Workshop on Web mining and social network analysis, ACM (2007) 56–65
8. Jadhav, A., Purohit, H., Kapanipathi, P., Ananthram, P., Ranabahu, A., Nguyen, V., Mendes, P.N., Smith, A.G., Cooney, M., Sheth, A.: Twitris 2.0: Semantically empowered system for understanding perceptions from social data. In: Semantic Web Challenge. (2010)
9. Stankovic, M., Rowe, M., Laublet, P.: Mapping Tweets to Conference Talks: A Goldmine for Semantics. In: Workshop on Social Data on the Web, Shanghai, China, Volume 664, CEUR-WS.org (2010)
10. Sankaranarayanan, J., Samet, H., Teitler, B.E., Lieberman, M.D., Sperling, J.: Twitterstand: news in tweets. In: SIGSPATIAL, ACM (2009) 42–51
11. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: WWW, ACM (2010) 851–860
12. Laniado, D., Mika, P.: Making sense of Twitter. In: ISWC, Shanghai, China. Volume 6496 of LNCS, Springer (2010) 470–485
13. Letierce, J., Passant, A., Breslin, J., Decker, S.: Understanding how Twitter is used to widely spread scientific messages. In: Web Science Conference. (2010)
14. Huang, J., Thornton, K.M., Efthimiadis, E.N.: Conversational Tagging in Twitter. In: Hypertext, ACM (2010) 173–178
15. Guha, R., Mccool, R., Miller, E.: Semantic search. In: WWW, ACM (2003) 700–709
16. Teevan, J., Ramage, D., Morris, M.R.: #twittersearch: a comparison of microblog search and web search. In: WSDM, ACM. (2011) 35–44
17. Koren, J., Zhang, Y., Liu, X.: Personalized interactive faceted search. In: WWW, ACM (2008) 477–486
18. Hearst, M.A.: Design recommendations for hierarchical faceted search interfaces. In: Workshop on Faceted Search co-located with SIGIR. (2006) 26–30
19. Oren, E., Delbru, R., Decker, S.: Extending faceted navigation for rdf data. In: ISWC, Springer (2006) 559–572