

# Ins and Outs of News: Twitter as a Real-Time News Analysis Service

**Arjumand Younus, M. Atif Qureshi**  
Department of Computer Science, KAIST,  
Daejeon, Korea  
[arjumandms@kaist.ac.kr](mailto:arjumandms@kaist.ac.kr), [atifms@kaist.ac.kr](mailto:atifms@kaist.ac.kr)

**Ahmad Nauman Ghazi**  
Blekinge Tekniska Högskola,  
Ronneby, Sweden  
[angh08@student.bth.se](mailto:angh08@student.bth.se)

**Samina Mumtaz, Muhammad Saeed**  
Department of Computer Science, Karachi  
University, Karachi, Pakistan  
[saeed@uok.edu.pk](mailto:saeed@uok.edu.pk), [saminamumtaz08@gmail.com](mailto:saminamumtaz08@gmail.com)

**Nasir Touheed, M. Shahid Qureshi**  
Institute of Business Administration,  
Karachi, Pakistan  
[ntouheed@iba.edu.pk](mailto:ntouheed@iba.edu.pk), [mqureshi@iba.edu.pk](mailto:mqureshi@iba.edu.pk)

## ABSTRACT

Both time and user popularity play a crucial role within the domain of news search - the fundamental problem lies in integrating these two onto a single platform for news retrieval systems. In this paper we propose techniques for enhancement of these systems by identifying the news topics that are popular over a certain period of time. The notion of “*popularity over time*” is studied with the help of the well-known real-time microblogging service “Twitter” – our method performs linguistic analysis of the news data published daily on news sites for extraction and detection of news topics that are in high demand using Twitter. We also present a prototype of our framework which detects popular news in real-time. The results obtained suggest the need of taking into account users’ interest for effective news services and strongly imply that harnessing of micro-blogging data for this purpose can lead to surprising outcomes.

## Author Keywords

Twittersphere, online news services, news retrieval systems, popularity over time

## ACM Classification Keywords

H.3.5 Online Information Services: Web-based services.

## 1 INTRODUCTION

The Social Web has introduced a major paradigm shift in the way the World Wide Web is envisioned and this has also had a significant effect on the users’ news consumption patterns –users share their chosen news (simply, a news of users’ interest or a popular news for users) through the social media platforms and as such there is an increased need for effective news services that take into consideration the users’ interest in the news [2, 5, 10, 24]. Furthermore, search engines are the primary means through which users access online news [15]. Hence in the domain of news

information retrieval there is a clear need for the use of approaches that take into account human factors; this is also the case for the news production and presentation process [24]. “*Popularity over time*” is a crucial human factor within the news retrieval and presentation domain.

Our notion of “*popularity over time*” with respect to news services encompasses two concepts: user *popularity* in the form of hot news topics and *time* which in the case of news services relates to freshness of news. Both these aspects play a very significant role in news information retrieval: popularity is a notion that has long been studied by information retrieval researchers [14] and recent focus of researchers in this domain is shifting towards temporal aspects of documents [1] – however what is being ignored in these efforts is a need to integrate these two important concepts which is a significant requirement for the special case of news information retrieval systems. This paper is a step in this direction and it proposes an approach for incorporation of both these concepts.

The goal of this paper is to demonstrate how information from a micro-blogging site can be exploited to study the evolution in news topics’ popularity. We specifically demonstrate how to use Twitter to automatically identify the hot news topics that Internet users are most concerned about – furthermore, we do this repetitively in real-time as and when news outlets produce more news. The paper also includes the description of our prototypical framework for identification and detection of popular news topics with time in which we apply natural language processing techniques on the news data and integrate our framework with the Twitter Search API. At the core of the framework is a visual interface [20] for effective navigation across the news data and Tweets’ data which in turn enables us to assess in real-time the popularity of a particular news article on the Social Web.

Additionally through experimental evaluations on real-time news data we provide empirical evidence supporting the following claims:

- Twitter acts as an effective news sensor for popular news topics and items.
- Twitter's social graph can give important insights into the news that are really meaningful to the readers.
- Twitter acts as a crucial information gathering point in the face of events that make breaking news.

These findings suggest rich ways in which next-generation news services can utilize users' interest in the news to improve the online news experience.

The rest of the paper is organized as follows. Section 2 discusses related work in this dimension. Section 3 presents a discussion of how Twitter is being used as an effective real-time news delivery service; it also includes results of a user survey we conducted to analyze users' news sharing patterns on Twitter along with a brief summary of existing news services and their limitations. Section 4 describes our methodology in detail along with an explanation of the underlying techniques. Section 5 presents the experimental results. Finally Section 6 concludes the paper with directions for future research.

## 2 RELATED WORK

There have been extensive studies and analysis on various user-generated content such as blogs, social networks and micro-blogging sites. This work has been carried out in various dimensions. In this section we provide a brief overview of works that are similar to ours along with a discussion of how our contribution differs from existing work.

Works on analysis of blogs pursue along the same lines as we do in this paper by analyzing the link between blogs and news articles thereby extracting and monitoring news trends from the blogosphere [4, 6, 23]. However with the emergence of social media and in particular micro-blogging platforms such as Twitter there has been an increasing trend of news sharing and searching [21] on these media with blogs now serving as a point where users publish their reactions to the news and social networking sites have replaced their role as a news sharing medium.

Due to this recent development in the Internet landscape within the past three years there has been an increasing amount of research on Twitter. Java et al. [9] performed a study on the topological and geographical characteristics of Twitter – their findings reveal that people on Twitter use it as a communication service and connect with people of similar interests. Hughes et al. [8] examine usage of Twitter in the face of crisis situations and their findings imply that under emergency events Twitter assumes the role of a significant information broadcasting medium. Sakaki et al. [17] also perform a similar study but their work differs in

that they use Twitter as a social sensor for emergency events such as earthquakes. Shamma et al. [19] analyze tweets in the face of live media events and conclude that tweets in the context of any particular media event give significant insights into the semantic structure and content of the event.

There is not much work on analysis of Twitter with respect to news data. The only work up to the best of our knowledge is by Sankaranarayanan et al. [18] – TwitterStand is basically a news processing system built on top of Twitter; it captures tweets corresponding to late breaking news and produces a distributed news wire service from these tweets. Our work differs from TwitterStand in that we use Twitter as an indicator for assessing news popularity; moreover our service combines news' data and tweets' data on a single platform. This is explained in further detail in Section 4.

## 3 OVERVIEW OF EXISTING NEWS SERVICES

In this section we provide an overview of the existing ways of news delivery – at the same time we discuss the role of Twitter as a real-time news service and we also look at questionnaire data of an online user survey we conducted as part of a preliminary study on Twitterers' perception of Twitter as a news service.

### 3.1 Traditional News Services

The oldest class of services that fall under the umbrella of traditional services are online versions of newspapers and television news with the average monthly reach of web newspapers among Internet households increasing from 27.4% in 2004 to 40.9% in 2008 [13], furthermore many television channels now make news clips available online. Another traditional news service constitutes search companies – according to the most recent Pew Research biennial news consumption survey [15] conducted in 2008, Yahoo (28%) and MSN (19%) are the most frequented websites among Web news users. Search engines also aggregate news from many sources at the same time categorizing them into news categories of World, Business, Sports etc. – these engines also present news clusters across multiple sources. Then there are also the services of email/mobile news alerts and RSS feeds that allow users to get news on demand.

An inherent limitation in all the above-mentioned news sources is that there is little or no consideration for the notion of “*news popularity over time*” – the traditional services do not take into account the local or user preferential buzz factor for the various news items/topics which is a very significant requirement in today's rich media.

### 3.2 Exploratory News Services

Recently researchers in information retrieval domain are shifting towards exploratory search paradigms for news services which allow users to explore and discover how entities such as people and locations associated with a news

query change over time. Some examples include Yahoo’s Time Explorer example [12] and Google News Timeline [7]; this search service incorporates a time-based clustering algorithm which also peers into the future based on predictions in news articles about the future events. The search results associated with a news query are arranged on a timeline and it allows users to interact with time and entities in a powerful way.

Although such an exploratory news search engine presents news in a very intuitive format, we believe it has the same problems as those of traditional news services because there is still a strong need to integrate the concept of popularity with time. Systems like Time Explorer are more focused on temporal aspects of news items which will eventually lead to the same problem as in traditional news services. Furthermore these systems are still in a phase of infancy and their use is not widespread till now. We advocate that such systems will be highly beneficial and effective if they incorporate both user popularity and temporal aspects of news documents.

### 3.3 Twitter as a Real-Time News Service

Twitter has emerged as a popular micro-blogging service – it is structured in such a manner that in addition to being a micro-blogging service it is also considered to be a social network [11]. Twitter users follow others or are followed – this relationship requires no reciprocation unlike other online social networking sites such as Facebook or MySpace. Being a follower implies that a user receives all the messages called tweets (where each tweet contains a maximum of 140 characters) from those the user follows. Other special features of Twitter include the ability to converse with other members of the Twittersphere using the @ symbol; RT stands for a retweet which refers to the practice of sharing the same content as that shared by another Twitterer – this is also another mode of replying on Twitter. Another special feature of tweets is that they sometimes contain the hash symbol “#” referred to as hashtags. By including a hashtag in a tweet, the user originating the tweet is suggesting that the word denoted by the hashtag makes for a good candidate as a search key for the tweet.

Twitter has become a very attractive news on-the-fly medium; in fact many a times using Twitter one gets information on newsworthy events even before the news organizations do [17, 18]. In fact recent research has demonstrated the use of Twitter as an effective information-finding medium especially with respect to temporally relevant information [21].

We conducted an online survey as a preliminary study on how users of Twitter perceive the role of Twitter as a news medium in which 51 active Twitter users participated and this helped us in laying the groundwork for the subsequent analysis. The questionnaire was designed on a funnel-shaped pattern to generate the flow of information from general to specific. The mode of the questionnaire was kept

in both dimensions i.e., open-ended and close-ended to maintain the criteria of required information gathering. Most respondents (84%) reported reading tweets one or more times per day. Forty seven (92%) of the respondents agreed that they used Twitter to update themselves about current issues and news and forty two (82%) respondents indicated a preference to share happenings around them on Twitter. Another important indicator used in this questionnaire was the marking on a scale of 1-5 (1 indicating lowest preference and 5 indicating highest) for the role of various sites as information-finding places: we included traditional media i.e., TV channels, newspapers, online news sites and search engines and the new media i.e., Twitter. The results of this marking yield significant insights into role of Twitter as a real-time news service and this is what motivates us for use of this service in studying news popularity over time: 29 (57%) of the respondents placed Twitter’s role on a higher or equal scale to that of search engines like Google and other news sites; seventeen respondents (33%) placed Twitter lower than other traditional media but the surprising aspect is that the difference in marking levels was only by a factor of 1 i.e., these users marked traditional media on level 5 and Twitter on level 4. This leads to a very significant observation that the role of Twitter as a real-time news service is increasing very rapidly. This serves as the basic motivation behind our approach of utilizing Twitter for a study of news evolution and assessment of user popularity regarding the various news items.

## 4 STUDYING NEWS EVOLUTION WITH TWITTER

In this section we provide a detailed description of our proposed framework for real-time detection of popular news topics. In Section 4.1 we begin with a discussion of the dimensions along which news retrieval systems align the news data and we mention the dimensions of our focus for the proposed framework. Section 4.2 describes the methodology in detail whereas Section 4.3 focuses on the user interface aspects of our prototype. Finally in Section 4.4 we discuss our ranking approach for determining the popular news topics from tweets containing news data; the ranking function is formulated on the basis of Twitter’s social graph.

### 4.1 Dimensions for a News Retrieval System

A news retrieval system has to take into consideration news data along three dimensions as Figure 1 shows.

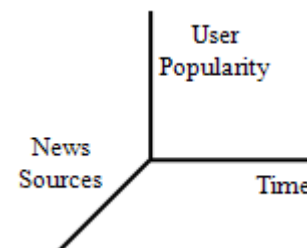


Figure 1: News Data Dimensions for a News Retrieval System

Time has long been an integral part of search engine ranking with most major search engines giving a ranking boost for recently published documents particularly in the news domain [12]. Furthermore, news ranking sites use the sharing of news items across social networks for assessing the popularity – algorithms like PageRank [14] are also used for popularity measurement of different news sites. Moreover as we mentioned in Section 3.1 news search engines also present news stories as clusters across multiple sources [24]. However, as we mentioned in Section 1 the fundamental problem lies in not integrating these three dimensions onto a single platform for news retrieval systems. In this paper we consider two of these dimensions simultaneously: user popularity and time.

## 4.2 Proposed Methodology for Assessing News Popularity

In this section we describe our proposed system in detail. Figure 2 shows the architecture of our system.

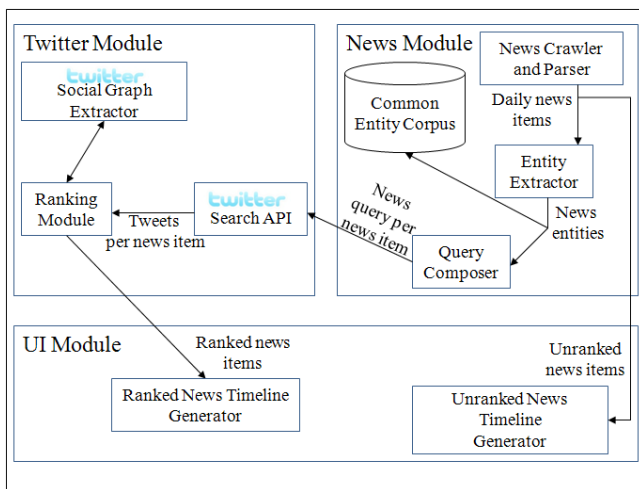


Figure 2: Overview of the System

The system is composed of three modules:

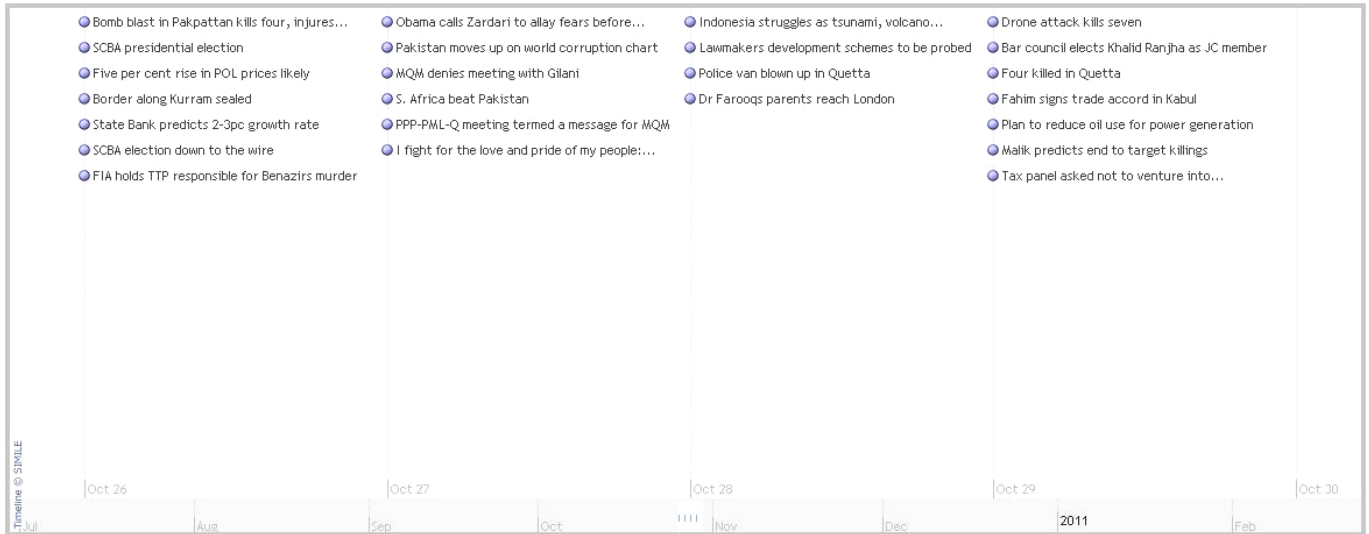
- **News Module:** as shown in the figure this module performs the daily crawl of news articles; the crawler [16] basically follows a refresh policy that is each day's newly published news articles are added to the archive. Currently the system provides a visual interface for displaying news articles of past four days only (as a prototype, believing that hot news items are generally effective over a period of maximum four days). Each news article is parsed into summary, title and news text. The entity extractor in the news module takes as input the parsed news data – it processes the news summary (provided by the news provider) part to extract the named entities of each news article using a named entity recognition approach [22]. These named entities are then matched across the common entity corpus on disk – this corpus is created from parsed

news data gathered from January to August, 2010 and the entities with high document frequencies during this interval are part of the common entity corpus; additional entities within this common entity corpus include those entities that are very common in news documents such as names of countries, cities, famous personalities, government officials etc. The entity extractor then sends entities tagged with information of which is a common one and which is not is sent to the query composer module which uses Boolean query predicates for the search task: those entities that are common are used with AND and those not common are used with OR, if all the entities for a particular news article are common entities then the news title is used for composition of the query.

- **Twitter Module:** This module is responsible for the fetching of the tweets relevant to the news articles. The query composer in the news module sends a composed query to the Twitter Search API: a query corresponding to each news article is sent. The API returns tweets that correspond to each news item; each returned tweet has associated metadata such as timestamp, screen name of user who tweeted it, whether it was in reply to someone etc. After this essential step in this module comes the task of ranking all the news articles – we use tweets extracted per news article for this task i.e., for each tweet relevant to a news article we extract the friends and followers of the user who tweeted it and apply our ranking function to rank each news article with respect to its popularity on Twitter. We explain this ranking function in further detail in Section 4.4.
- **UI Module:** this module comprises the user interface part of our system. It takes an input from both the News module and the Twitter module. The News module sends it an unranked list of news articles for each of the four days whereas the Twitter module sends it a ranked list of news articles for each of the four days. This module is further explained in Section 4.3.

Table 1 gives an algorithmic overview of our approach. The items marked in italics hold special meaning with respect to the proposed system.

Popularity by Twitter Analysis



DAWN.COM Original Order

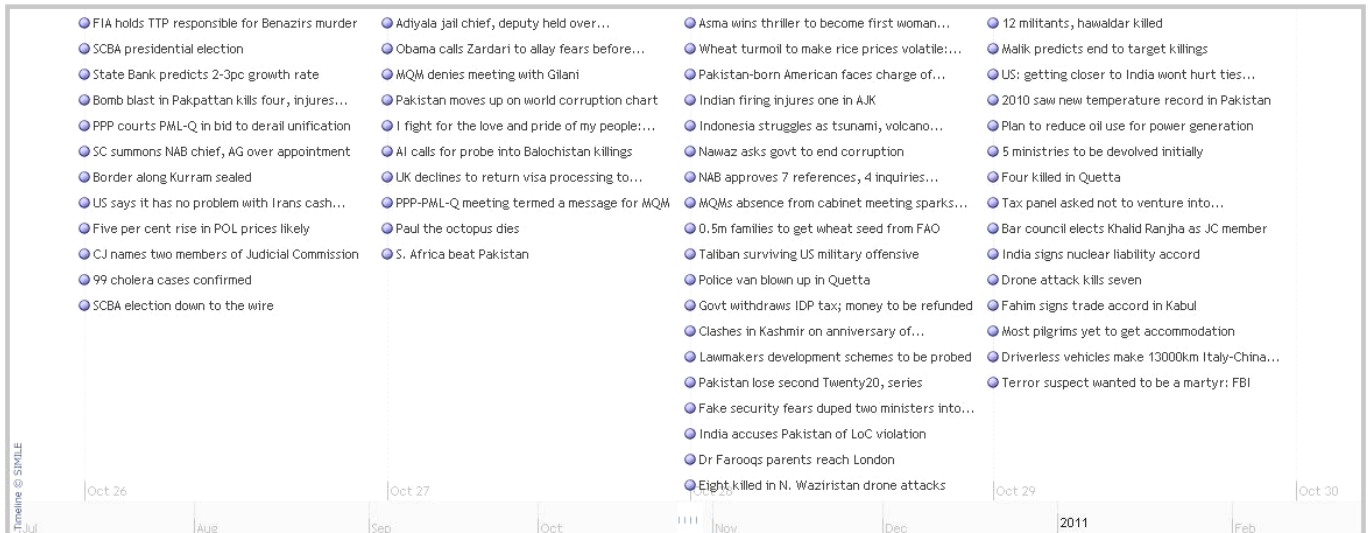


Figure 3: Snapshot of application timeline on 29<sup>th</sup> October, 2010

1. Crawl daily news data and send the *unranked news articles list* to the UI module of the system.
2. Extract *news title*, *news summary* and *news text* for each news article per day.
3. From the *news summary* extract *named entities* per news article through a named entity recognition approach.
4. Match each *named entity* across entities in the *common entity corpus* and tag each *named entity* per news article as common or uncommon.
5. Use Boolean query model to compose the *query* per news article:
  - a. Use AND predicate with each common entity and OR predicate with each uncommon entity. If all entities for a particular news article are common use *news title* for the query construction.
6. For all articles per day:
  - a. Send a request to Twitter Search API and extract *result tweets* per news article per day
  - b. For each *t* in *result tweet*:
    - i. Use *t*'s metadata to find following and follower statistics for each unique Twitterer who has tweeted about the news article
    - ii. Calculate rank of each article using the ranking function (explained in detail in Section 4.4)
7. Send the *ranked news articles list* to the UI module of the system

**Table 1: Algorithmic Flow of News Popularity Ranking using Twitter**

### 4.3 User Interface

The focus of our prototype system is to detect the popularity of a news topic/article over time and hence at the core of the user interface is a timeline. Figure 3 displays the timeline for 29<sup>th</sup> October, 2010.

As mentioned in Section 4.2 the system shows the timelined news assessment over a period of four days. The data of the news articles taken from the original news source is shown side by side with the popularity ranks calculated using our Twitter module. This system simultaneously addresses the temporal and popularity aspects within the news domain.

### 4.4 Ranking Approach

In this section we explain our ranking methodology for the ranking module within the Twitter module of our system – one easy approach could be to count the number of tweets corresponding to each news article that are retrieved by the Twitter Search API. We argue however that taking into account social network features of Twitter is a more effective means of ranking the news articles.

As explained in Section 3.3 there is a follower paradigm in Twitter – if user *a* chooses to follow user *b* then user *a* is a follower if we look at it from the viewpoint of user *b* whereas friend if we look at it from the viewpoint of user *a* – this friend is to followers relationship is what we make use of in our ranking methodology. With the growing popularity of Twitter as a news media as we explained in detail in Section 3.3, many news and media outlets have migrated to Twitter with these channels maintaining their channels' Twitter account for news dissemination and propagation. One potential drawback of our proposed

system explained in Section 4.2 is that it will not distinguish the tweets disseminated on news articles by these news outlets – hence we devise a mechanism to do this using the Twitter social graph. One obvious point to note in the social network of these news and media outlets is that they have a large number of followers but a negligible number of friends i.e., since they are news outlets many people follow them on Twitter but they follow none. Hence it makes sense to utilize the following to follower ratio for ranking of the news articles. Despite a lot of news providers tweeting a particular news item, common users would also tweet it if the news is really popular among users. Moreover, if it is broadcasted but users do not take interest in it then such a news item cannot be a popular news item (may be an over projected news). Equation 1 shows the formula that we use.

$$\sum_{\text{for all Twitterers Of News}} \frac{\text{followings}}{\text{followers}} \quad (1)$$

## 5 EXPERIMENTAL EVALUATIONS

This section presents the details of the experiments we conducted over a period of 10 days from 17<sup>th</sup> October to 26<sup>th</sup> October, 2010.

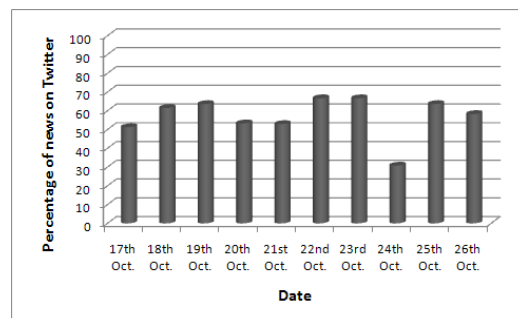
### 5.1 Experimental Dataset

The dataset comprises of the front pages news articles from Dawn.com [3], a famous Pakistani news paper's website which publishes news in the English language. We collect their articles on daily basis from 17<sup>th</sup> October to 26<sup>th</sup> October – we then perform the analysis using our system. For each day's news data the Twitter module is run at a gap of one day hence providing a fair comparison for all the days. The statistics collected for each run include the number of tweets per news article per day and the ranking scores obtained through equation 1 in Section 4.4.

### 5.2 Experimental Results

Three sets of experimental results are shown as verification for each claim made in Section 1.

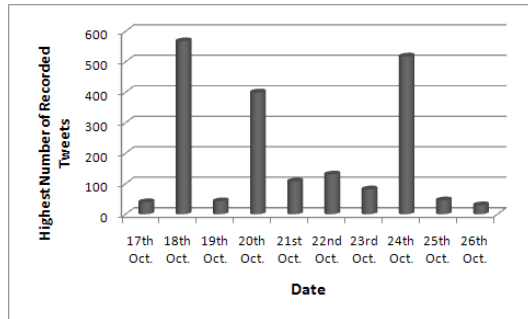
The first set of experimental results shows the percentage of news discovered in tweets during the period of our study. Figure 4 shows the results.



**Figure 4: Percentage of News in Tweets per Day**

As is obvious from the results in Figure 4 almost daily more than 50% of the top news in Pakistan have a mention on the except for just one day – there was a significant drop in the percentage of news in tweets on 24<sup>th</sup> October, 2010 and the reason for this was found to be due to the appearance of the news “*WikiLeaks makes fresh claim about Iraq deaths*” which shook the entire Twittersphere on an international level. Only up to a maximum of 67% of news were mentioned in tweets although these were news mentioned on the front page of the Dawn.com [3]. This proves our original claims that traditional news services tend to ignore user preferences for news consumption and also provide experimental evidence for the claim that Twitter acts as an effective news sensor for popular news topics and items.

The second set of experimental results shows the maximum number of tweets recorded on each day for the entire period of the experimental analysis. Figure 5 shows the results.



**Figure 5: Highest Number of Recorded Tweets per Day**

As is obvious from the results in Figure 5 there is a large fluctuation in the highest number of recorded tweets per day. This fluctuation is explained with the help of Table 2. The table shows the event against which the highest number of tweets were recorded for each day: the days on which there was a huge amount of tweets marked a significant international event that made breaking news i.e., 18<sup>th</sup> Oct., 20<sup>th</sup> Oct. and 24<sup>th</sup> Oct. – 24<sup>th</sup> Oct. was also the day in which the percentage of news in tweets as shown in Figure 4 dropped to 30.8%; this is reasonable because the news for which this occurred “*WikiLeaks makes fresh claim about Iraq deaths*” was also heavily tweeted by the Pakistani Twittersphere as it also affects them. An important point to note however is that the Dawn.com news website shows this news quite low in the front page which again verifies that traditional news sources have limitations due to not taking into account the notion of “*popularity over time.*” The second set of experimental results further provide experimental evidence for our claim that Twitter acts as a crucial information gathering point in the face of events that make breaking news.

DATE	EVENT
17 <sup>th</sup> Oct.	Karachi violence ( <i>local</i> )
18 <sup>th</sup> Oct.	Hopes fade for trapped Chinese miners ( <i>international</i> )
19 <sup>th</sup> Oct.	Lakki Marwat suicide attack attempt foiled ( <i>local</i> )
20 <sup>th</sup> Oct.	Rebels raid parliament in Grozny; 7 dead ( <i>international</i> )
21 <sup>st</sup> Oct.	Obama to visit next year ( <i>local</i> )
22 <sup>nd</sup> Oct.	Nuclear plant completes decade of good performance ( <i>local</i> )
23 <sup>rd</sup> Oct.	Ghazi shrine suicide bomber back home ( <i>local</i> )
24 <sup>th</sup> Oct.	WikiLeaks makes fresh claim about Iraq deaths ( <i>international</i> )
25 <sup>th</sup> Oct.	Court orders Iraqi parliament back to work ( <i>international</i> )
26 <sup>th</sup> Oct.	SCBA presidential election ( <i>local</i> )

**Table 2: Events Corresponding to Highest Tweets in Figure 5**

The third set of experimental results performed a comparison for proving the effectiveness of the ranking approach based on Twitter’s social graph and our proposed hypothesis that following is to follower ratio is extremely low for news outlets. We use Spearman’s footrule which is a widely used measure for evaluating similarity between two ordered lists, equation 2 shows the computation of the Spearman’s footrule for two ordered lists  $L_1$  and  $L_2$ ; in the formulation  $i$  is an element in these lists  $L_1(i)$  and  $L_2(i)$  respectively

$$Spearman(L_1, L_2) = \sum_i |L_1(i) - L_2(i)| \quad (2)$$

We ask a regular follower of news to give us a human-ranked list of the news items for the 4 days: 21<sup>st</sup> Oct-24<sup>th</sup> Oct and this ranked list is assumed to be perfect since it was produced by a regular Pakistani news reader. This measure is then used to evaluate the similarity between the ranked lists obtained using number of tweets (naive approach) and our formulation (equation 1) as a ranking function. Table 3 shows the values of Spearman footrule for both the ranking approaches.

DATE	NumTweets	Equation 1
21 <sup>st</sup> Oct.	4	1
22 <sup>nd</sup> Oct.	3	0
23 <sup>rd</sup> Oct.	5	1
24 <sup>th</sup> Oct.	4	1

**Table 3: Spearman Footrule Values for Both Ranking Approaches**

Table 3 clearly demonstrates that the dissimilarity between the human-produced list of news items and the algorithmic ranked list was higher when we use number of tweets (naive approach) as ranking criteria. On the other hand the our formulation (equation 1) performs almost similar to the human produced list and hence proves our experimental claim that Twitter's social graph can give important insights into the news that are really meaningful to the readers.

## 6 CONCLUSIONS

We have described a general framework for integrating the concepts of user popularity and time within the domain of news information retrieval. Our approach is built on top of a very popular micro-blogging service which also plays the role of a social network – furthermore, Twitter provides support for both the concepts that are crucial for news retrieval systems i.e., popularity and time and we proved its effectiveness through experiments performed on real-time news data. The work also includes a running prototype for real-time detection of hot news topics.

Future work in this area includes several dimensions: we plan to incorporate the third important dimension that news retrieval systems have to deal with i.e., the news sources. Moreover future directions also include refinement approaches for the common entity corpus using a learning approach which makes it adapt to the changing and evolving news patterns. We also intend to extend the prototype for Twitter's real-time message feed for different users by building it on top of the Twitter REST API.

## REFERENCES

- Alonso, O., Gertz, M., and Baeza-Yates, R. On the value of temporal information in information retrieval. In SIGIR Forum, Vol. 41, No. 2, pages 35-41, Dec. 2007.
- Anderson, C.R. and Horvitz, E. Web montage: a dynamic personalized start page. In *Proc. WWW 2002*, ACM Press (2002), 704-712.
- Dawn News Website <http://www.dawn.com>
- Fukuhara, T., Murayama, T., and Nishida, T. Analyzing concerns of people using Weblog articles and real world temporal data. . In *Proc. WWW 2005*, ACM Press (2005), 1-12.
- Gabrilovich, E., Dumais, S., and Horvitz, E. Newsjunkie: providing personalized newsfeeds via analysis of information novelty. In *Proc. WWW 2004*, ACM Press (2004), 482-490.
- Glance, N.S., Hurst, M., and Tomokiyo, T., BlogPulse: Automated trend discovery for weblogs. In *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, 2004.
- Google News Timeline <http://newstimeline.googlelabs.com>
- Hughes, A.L., and Palen, L. Twitter adoption and use in mass convergency and emergency events. In *Proceedings of the 6<sup>th</sup> International Conference on Information Systems for Crisis Response and Management*, 2009.
- Java, A., Song, X., Finin, T., and Tseng, B., Why we Twitter: understanding microblogging usage and communities. In *Proceedings of the 9<sup>th</sup> WebKDD and 1<sup>st</sup> SNA-KDD workshop on Web mining and social network analysis*, pages 56-65, New York, NY, USA, 2007.
- Kamba, T., Bharat, K., and Albers, M. The krakatoa chronicle: An interactive, personalized newspaper on the web. In *Proc. WWW 1995*, ACM Press (1995).
- Kwak, H., Lee, C., Park, H., and Moon, S. What is Twitter, a social network or a news media? In *Proc. WWW 2010*, ACM Press (2010), 591-600.
- Matthews, M., Tolchinsky, P., Blanco, R., Atserias, J., Mika, P., and Zaragoza, H. Searching through time in the New York Times. In *HCIR '10: Proceedings of the 4<sup>th</sup> Workshop on Human Computer Interaction and Information Retrieval*, pages 41-44, New Brunswick, NJ, USA, August 2010.
- Nielson/Net Ratings, MegaPanel Data <http://www.naa.org/TrendsandNumbers/Newspaper-Websites.aspx>
- Page, L., Brin, S., Motwani, R. and Winograd, T. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project, 1998.
- Pew Research Center for the People and the Press, 2008. Key News Audiences Now Blend Online and Traditional Sources – Audience Segments in a Changing News Environment. *Pew Research Center Biennial News Consumption Survey*. <http://people-press.org/reports/444.pdf>
- Qureshi, M.A., Younus, A., and Rojas, F. Analyzing Web Crawler as Feed Forward Engine for Efficient Solution to Search Problem in the Minimum Amount of Time through a Distributed Framework. In *Proceedings of 1<sup>st</sup> International Conference on Information Science and Applications*, pages 1-8, August, 2010, IEEE Publications.



17. Sakaki, T., Okazaki, M., and Matsuo, Y. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proc. WWW 2010*, ACM Press (2010), 851-860.
18. Sankaranarayanan, J., Samet, H., Teitler, B.E., Lieberman, M.D. and Sperling, J. TwitterStand: news in tweets. In *Proceedings of the 17th ACM SIGSPATIAL international Conference on Advances in Geographic Information Systems*, pages 42-51, Seattle, Washington, USA, November 2009.
19. Shamma, D., Kennedy, L., and Churchill, E. Tweet the debates: Understanding community annotation of uncollected sources. In *Proceedings of the ACM International Conference on Multimedia*. ACM, 2009.
20. SIMILE, <http://www.simile-widgets.org/timeline/>
21. Teevan, J., Ramage, D., and Morris, M.R. #TwitterSearch: A comparison of microblog search and web search. In *Proceedings of the 4th international Conference on Web Search and Data Mining (WSDM'11)*, Hong Kong, China, February 2011.
22. Tjong Kim Sang, E.F., and De Meulder, F. Introduction to the CoNLL-2003 shared task: language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4 Human Language Technology Conference*. Association for Computational Linguistics, pages 142-147, Morristown, NJ.
23. Thelwall, M. No place for news in social network web sites? In *Online Information Review*, vol. 32, pages 726-744, 2008.
24. Vydiswaran, V.G.V, and Chandrasekar, R. Improving the online news experience. In *HCIR '10: Proceedings of the 4<sup>th</sup> Workshop on Human Computer Interaction and Information Retrieval*, pages 122-125, New Brunswick, NJ, USA, August 2010.