Mapping Temporal Horizons *

Analysis of Collective Future and Past related Attention in Microblogging

Adam Jatowt
Kyoto University
adam@dl.kuis.kyoto-u.ac.jp

Émilien Antoine
Kyoto Sangyo University
first.last@cc.kyoto-su.ac.jp

Yukiko Kawai
Kyoto Sangyo University
kawai@cc.kyoto-su.ac.jp

Toyokazu Akiyama
Kyoto Sangyo University
akiyama@cc.kyoto-su.ac.jp

ABSTRACT

Microblogging platforms such as *Twitter* have recently received much attention as great sources for *Live Web* sensing, for real time event detection or opinion analysis. Previous works usually assumed that the tweets mainly describe "what's happening now". However, a large portion of tweets actually refers to time frames within the past or the future. Such messages often reflect expectations or memories of social media users. In this work we investigate how microblogging users collectively speak about time. In particular, we analyze half a year long portion of Japanese and four months long collection of US tweets and we quantify *collective temporal attention* of users and other related temporal characteristics. This kind of knowledge is helpful in the context of growing interest for prediction and for detection of important events within social media. The exploratory analysis we perform is possible thanks to the development of visual analytics framework for robust overview and easy detection of various regularities in the past and future-oriented thinking of Twitter users. We believe that the visualizations we provide and the findings we outline can be also valuable for sociologists and computer scientists to test and refine their models about time in natural language.

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation**]: Miscellaneous

Keywords

Twitter, temporal analysis, visualization

*This work was supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications of Japan, and JSPS KAKENHI Grant Number 26280042.

1. INTRODUCTION

Memory of past actions or events as well as expectations or predictions about future occupy much of our thinking every day. Rather than constantly focusing on the present, human thoughts tend to stray over different time perspectives within near or more distant time. Memory is useful for analyzing the past and for understanding the present, while future expectations and future-focused thinking serve an important role of arranging our actions and preparing for what may come.

As both remembering and expecting represent significant amount of human cognitive activities, not surprisingly, also in social media, users tend to leave traces of thoughts and opinions related to the past and future. Such temporal references constitute wealth of information about what users are going to do or what they expect to happen, as well as which past events played important roles in their lives. Recently, there has been significant interest in building models for prediction of real-world outcomes using social media. For example, it has been shown that movie box-office revenues can be successfully forecasted using Twitter [3]. Other works focused on predicting future activities of users [28, 30]. Most of the proposed approaches relied on detecting periodical patterns or trends of user activities for their subsequent extrapolation to forecast the future. We advocate here another approach to directly analyze expectations expressed by users for judging their future actions. For example, someone tweeting about the meeting scheduled in the evening leaves a strong signal as for her or his future activity at the end of the day. We believe that the traces of future-related thoughts are valuable to be mined and analyzed as they can be used for recommendation, targeted advertising, predictive user assistance and group behavior forecasting or even for tracking user whereabouts in case of sudden, massive disasters.

Similarly, we consider past-related references to carry high significance for assessing self-reported importance of events and activities users had in the past, or, in general, to characterize remembering and reminiscing patterns. In social sciences the concept of *collective memory* as a shared remembrance of the past has been known for quite a long time. However, it was usually measured in small scale or through controlled studies. It has becomes possible now to investigate the memory decay on the collective of social media users.

To sum up, we consider studying the expectations and memories of social media users to be useful not only for practical reasons such as constructing future prediction systems but also for fostering social science and better understanding user behavior.

In this work, we advocate the concept of *memory and expectation sensing* as a complement to the well-known notion of *social sensing* in microblogging and social media. As we are aware that human memory and futuristic thinking are inherently complex matters we thus first attempt to uncover any observable regularities and patterns in data through exploratory analysis of the way in which social media users collectively refer to the future and the past. By this we hope to shed more light on the feasibility of implementing effective future prediction solutions through improved feature engineering and for understanding the importance of future and past events for users.

In particular, we focus on Twitter which is commonly used in the nascent field of *computational social science* [18]. We first collect tweets with any temporal references treating them as reflections of user thoughts on the future and the past. We make use of a half year long portion of Japanese tweets from mid-July 2013 to mid-January 2014; and a four month long collection of tweets in USA from mid-September 2013 to mid-January 2014. Then we map any detected temporal references to their corresponding time points. This allows us to determine the *reference time* (focus time) of tweets. Next we manipulate the two temporal signals: tweet timestamp and reference time (disambiguated time mention) of tweets. Through the visual analytics process we then make it possible to uncover several important findings.

Note that we choose two culturally different regions in order to perform comparative analysis of the way in which users refer to both past and future in Japan and in the USA. We find many commonalities in the shape of temporal attention but at the same time we also pinpoint important cultural differences, for instance US Twitter user attention seems to be slightly more oriented to the future.

To sum up, we make the following contributions in this paper. We are the first to analyze the way in which microblogging users refer to the past and to the future, and to propose the idea of *memory and expectation sensing* in social media. Our focus is on two distinct countries and on tweets written in different languages. For enabling visual analytics, we build an *interactive system* that collectively visualizes historical and future-oriented perspectives in microblogging. The system is available online [32, 33] for anyone interested in a quick overview or in detailed exploration of temporal aspects in shared messages. Lastly, we discuss the analysis results and outline several avenues for further research.

The reminder of this paper is structured as follows. In the next section we review the related work. Section 3 describes the data processing model, while Section 4 introduces our methodological approach for visualizing temporal patterns in tweets and overviews findings from this study. We formally describe the technical aspects in Section 5 and summarize our findings as well as their implications in the subsequent section. Finally, Section 7 concludes the paper and outlines our future work.

2. RELATED WORK

2.1 Social Networks and Media Analysis

Numerous research works have shown the usefulness of Twitter and, more generally, microblogging services for real-time information extraction and analysis in many domains, ranging from detecting natural disaster related information [26] via analyzing how the popularity of topics emerge and evolve on time [30] and space [2] to opinion detection and analysis of the crowd about events [27]. Some works also harnessed microblogging platforms to detect disease outbreaks [16].

Increasing numbers of users leave numerous traces of their daily life activities in a digital world [18], especially, in social networks where users like to share information on their personal life. Some works then exploit this new flow of personal data to predict upcoming activities that users may perform in the near future, e.g., by detecting sentences like "Ready for bodypump class party tonight!" [28].

2.2 Studies of Memory and Predictions

Our work also aims to foster the subset of socio-psychological sciences related to the study of memory and expectations. *Collective memory* (*social memory*) is a term introduced in social sciences by Halbwachs [10] to define the collective view of society on the past. Russell Jacoby [12] also coined the term *collective amnesia* for describing forcible repression of memories when whole groups or nations "forget" events from the past, typically, inconvenient or embarrassing ones. In parallel to *social memory* many studies investigated the characteristics of personal memory and the forgetting process following the pioneering works of Ebbinghaus [8, 9].

Both the social and individual-focused memory studies were limited to small samples of users and had subjective character. Although computer science approaches still lack in terms of analysis depth, they could already offer a significant complement to any manual investigations thanks to the effect of large data and high coverage. Already, some recent works started investigating collective memory [7] or estimating collective predictions [13] in news, and on the web [23, 24, 14]. For example, [14] showed the distribution of search engine hit counts for queries extended with future dates and listed terms that co-occur frequently with particular future years. Unlike those works, in our research, we focus on social media and we perform large scale analysis on the way in which microblogging platform users refer to time.

2.3 Temporal Information Retrieval

Memories and predictions rely on temporal analysis of text that models and leverages temporal expressions as well as their uncertainty [4, 13]. After prior detection and disambiguation, the temporal expressions can be included in the Temporal Information Retrieval (TIR) process [1, 4, 17, 15]. TIR has an objective of improving the effectiveness of information retrieval methods by exploiting temporal information in documents and queries and by extending the traditional notion of *topical relevance* with *temporal relevance* [1, 5]. In fact, quite a significant fraction of Web queries have been found to contain explicit temporal expressions such as dates or names of calendar events [22], while many other queries are known to have implicit temporal intent.

The importance of time comes from the fact that the value of information and its quality are intrinsically time-dependent. Information science researchers [20] consider timeliness or currency as one of the five key aspects that determine a document's quality; the others being relevance, accuracy, objectivity, and coverage. Studies in TIR have actually shown that the retrieval effectiveness of temporal queries can be significantly improved by modeling and taking into account publication timestamps and time mentions of documents [4, 17, 15]. Through this work we hope to provide better outlook on different ways in which users refer to diverse time periods and by this to offer clues for solving the challenge of designing effective TIR systems and for understanding society and humans.

2.4 Event Detection

Event detection is a major field of research in microblogging research. The reason is that Twitter is capable of transmitting information faster than traditional media channels, such as, news or official outbreak reports. The most relevant work in this regard is one on event detection from bag of words and on topic extraction or modeling for visualization [29]. More specialized works on live web event detection from Twitter use localization data [27, 26], or predicte hot or emerging topics [6].

3. DATA MODEL

This section describes our model to organize the data to make it usable for visualization. The following entities are of our interest: *users*: people sharing messages, *time-referred messages*: atomic interactions of users with others that contain temporal references including expectations and memories about subjects and *subjects*: objects mentioned in time-referred messages. Our field of study is a portion of Japanese and US Twitter data. Therefore in our context the users are Twitter registered users, the messages are tweets written in Japanese and English and the subjects are topics of tweets. A tweet is a short message limited to 140 characters used to communicate with others but also to react to certain events. We are particularly interested in time-referred tweet messages from which we can extract temporal expressions. The temporal expressions allow us to categorize tweets into those about future and those about past as well as to map them on timeline after prior disambiguation. We further describe the datasets in Section 4.1.

3.1 Data Structure

Each tweet is first represented as a tuple of raw data:

$$< \textit{user id, tweet content, timestamp} >$$

The *user id* is the Twitter identification number, the *tweet content* is the text of the tweet and the *timestamp* is the time when the tweet has been published on Twitter. Based on these attributes, the information related to time is extracted and is represented by two additional attributes:

$$< \textit{time mention, time diff} >$$

The *time mentions* are extracted and disambiguated from the *tweet content* by applying temporal entity-recognition techniques further detailed in Section 3.2. The *time diff* is then computed as the difference between the time mention and the timestamp. Intuitively, the quintuple of attributes should allow to answer simple but fundamental questions:

- What people tend to say at the same time ? (timestamp)
- What people tend to say about the same time ? (time mention)
- What people tend to say before or after the same time ? (time diff)

The goal of our data model is to allow a combination of attributes to be contrasted leading us to infer information about the global and per user behavior of microblogging users.

3.2 Time Mention Extraction

As mentioned above we consider only tweets with time mentions. Note that while there are ready temporal taggers for English, to the best of our knowledge, none is available for extracting and anchoring temporal expressions from Japanese. Therefore we use the Stanford CoreNLP tagger [19] for US data set and built our own tagging system for Japanese. We match tokens in the tweet content with a dictionary of the lexical expressions related to time in Japanese. We provide a set of common time expressions in Japanese accumulated from diverse resources such as the list of temporal expressions given by Japan Meteorological Agency [1] and others. The list we made includes different time-related phrases that refer either to near or distant time, both in future and in the past. We also asked 3 native speakers to ensure that no common temporal expressions have been missed and included their new suggestions. We also defined a set of regular expressions to capture mentions of hours, days, including names of weekdays, months, seasons, national holidays and years. Eventually, the numbers of tweets containing time mentions detected in both datasets are comparable, as shown in Section 4.1, indicting that our dedicated system for Japanese language seems to be more or less reliable.

We emphasize that differently to previous studies on analyzing or visualizing collective time references [13, 14], in this work, we use not only absolute temporal expressions (e.g., “January 2012” or “December 2, 2013”) but also we extract and disambiguate relative temporal references (e.g., “next week” or “three months ago”). Intuitively, the latter are quite common in spoken and colloquial language and should be included in the analysis.

Lastly, we note that our model assumes the entire tweet content to be related to temporal expression that appears in the tweet. This assumption is reasonable given the short length of tweets.

Time Mention Disambiguation.

Each temporal expression is mapped to an absolute time period value in the past or in the future. The process is straightforward for absolute time expressions. Relative time expressions are resolved based on tweet timestamps.

Time Granularity Shifts.

It is well-known that humans tend to perceive time using scale similar to a logarithmic one [11]. When the time distance measured from the speaking time point increases, the granularity of temporal expressions used becomes coarser. For example, when someone talks about far future (e.g., several months or years ahead) he or she usually refrains from pinpointing exact hours, minutes or even days. On the other hand, when referring to near future, such as to actions in the same day or few days later, humans tend to use finer granularity time-references (hours or even minutes). This have been also confirmed by the analysis of time expressions which were extracted from large collections of news articles and then related to the article timestamps [13]. In our model we thus apply the logarithmic scale

[1] `http://www.jma.go.jp/jma/kishou/know/yougo_hp/toki.html` (in Japanese)

of time using three main time granularity levels inherent to human speech: minutes, hours and days depending on the distance to the start time of the disambiguated temporal expressions measured from the tweet timestamp (time diff). Below we show a sample of the mapping used in our model emphasizing heterogeneity of time granularity and the phenomenon of *granularity shifts* in natural language:

"Now"	→	timestamp (granularity of minutes)
"Tomorrow"	→	one day after the timestamp (granularity of hours)
"Next month"	→	one month after the timestamp (granularity of days)

Mapping Time Expressions.

Usually time mentions in human language do not express a specific point in time but rather a period. For example, for the time mention: "I will fly to Okinawa next week" the model should reflect the fact that the next week is a seven days long period, while the action (flight) may take place on any of them. To do this we represent the time mention by a uniform probability distribution over its time duration. For example, "tomorrow" is mapped into probability distribution over hours in the next day, while "next month" is converted to the uniform distribution spanning each day of the following month and so forth. For the Japanese dataset, we provide a set of this kind of mappings for each temporal expression based on the Japan Meteorological Agency guidelines[2]. Figure 1 shows a sample of temporal expressions referring to different time frames of a day (hour granularity) and their mappings to the absolute time values. Note that the expressions have been translated from Japanese. For the US dataset we mainly use Stanford NLP tagger for anchoring temporal expressions with some additional manually added mappings.

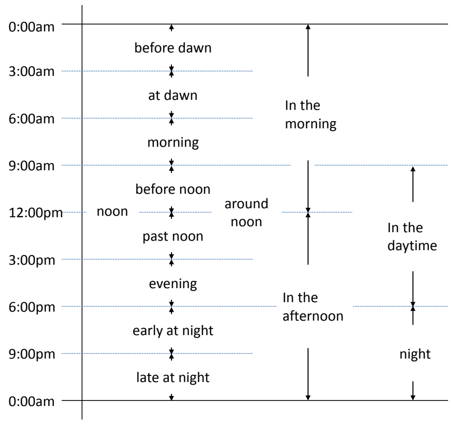


Figure 1: Sample of the mapping of time expressions referring to different times of a day (translated from Japanese).

To sum up, in this work, the weight coming from a single time-referenced tweet on a particular time point is determined using the uniform probability distribution over the period defined by the temporal expression in the tweet and according to the selected granularity level for this expression. Following, the weight of the time point coming from several temporal expressions that cover this time point is calculated as the sum of their corresponding probabilities. The technical details of this probabilistic mapping are formally described in Section 5.1. We note that some previous works [13] adopt a more complex approaches to map activities occurring within a period of time using either exponential, Gaussian or uniform distributions depending on used temporal modifiers. However deciding parameters of the probability distributions for different messages require deeper NLP processing such as analyzing the character of events and activities, which is outside the scope of this work. Such processing would be also very slow and require considerable computing resources for large datasets.

4. ANALYSIS

In this section we describe our visualization framework and detail findings we could obtain from the analysis. We defer the detailed description of the technical details until Section 5 to focus first on the conceptual way of portraying the data and on any observable patterns.

The following discussion will be mainly based on 2D plots called heat maps containing colored cells. A cell is the timespace on 2D time pane. The cell color represents the *intensity* with which tweets in the datasets refer to that cell. Below each graph, we display the color scale ranging from dark blue (the lowest value) to dark red (the highest value).

4.1 Distribution of Time References

First, we report the overall statistics of the data we use. The Japanese dataset has been build retrieving 31.6M tweets posted from Japan between July 21, 2013 and January 12, 2014; and the US dataset has been made by collecting 198M tweets between September 25, 2013 and January 17, 2014. Unfortunately, due to technical obstacles and the lack of resources the data crawling was disrupted at certain times. This explains the blank sections in Figures 3 and 4.

We use the language detection method[3], which is based on Naïve Bayesian filter and has nearly 99% precision, to select roughly 25M (millions) tweets written in Japanese in the Japanese dataset and 158M written in English from the US dataset. Among them we found 4M (16%) Japanese tweets and 30M (18%) English tweets that have temporal expressions. It means that a considerable fraction of messages contain some temporal clues. The slightly lower ratio of temporal expression usage in Japanese is more likely to be caused by our custom Japanese NLP parsing that may not be as efficient as the Stanford CoreNLP for English rather than by an actual difference within the datasets. The difference is reasonably small that indicates our custom Japanese time expression detection should work reasonably well.

We report now the overall strength of *temporal attention* of users towards the past, the present and the future. For this we group the extracted time references into those that relate to the three time scopes. The present is determined by several words denoting the concept of "now" in Japanese and English.

	Japan	US
# time-referenced tweets	3.96M	29.92M
about past	38%	26%
about present	22%	16%
about future	40%	58%

[2] http://www.jma.go.jp/jma/kishou/know/yougo_hp/saibun.html

[3] http://code.google.com/p/language-detection

These numbers indicate that the overall *temporal attention* of Japanese Twitter users is more or less uniformly distributed when it comes to the future and the past. However, according to our data the US Twitter user seem to be more oriented to future.

Next, we examined per user statistics. The ratio of users that at least once submitted time-referenced tweets is 50% for Japanese and 65% for US. Figure 2 shows the distribution of users according to the frequency with which they use time expressions. The horizontal axis is the percentage of tweets of a single user that contain any time expression and is grouped into 11 bins reflecting different usage rates of temporal expressions. The vertical axis denotes the percentage of users for each different usage rate of temporal expressions. We observe that about 40% of users in the Japanese and over 50% of users in the US datasets add time mentions in at least 10% or more of their tweets. This indicates that temporal expressions are relatively commonly used in Twitter and can be used as a signal for measuring temporal orientation of user messages.

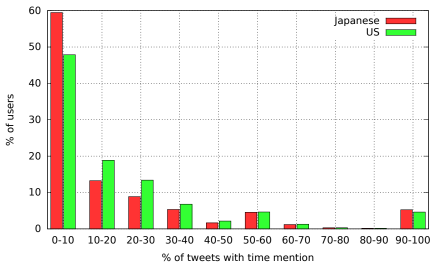


Figure 2: Percentage of users in Japanese and US datasets with given ratio of time mentions in their messages.

4.2 Scope of Temporal Attention

We expect the way in which users refer to the future or the past differs depending on whether the temporal perspective is near- or far-fetching.

In particular, we expect that forecasting and remembering attention should drop over time. By this we mean that the amount of predictions and remembrances about near future and near past should be greater than the ones associated with more distanced events. Previous studies [13, 14] conducted on news article and web page collections have demonstrated the drastic drop in the count of references that point to time periods far away from document timestamps.

To test this and other hypotheses we use the visualization shown in two graphs in Figure 3. The top graph (Fig. 3a) displays the Japanese dataset and the bottom one (Fig. 3b) is for the US dataset. The graphs show the difference between the time mentioned in tweets and their timestamps. The timestamp values of tweets published on a particular day are given on the abscissa and their agglomerated time differences are shown in ordinate in logarithmic scale. Each graph in Figure 3 is composed of two parts, the top one showing the map of future-related expressions and the bottom one portraying the expressions about the past. For example, when looking at the future-related parts of the graphs, the cells falling into the segment of 1 - 7 days represent aggregate tweets that contain time references pointing to the period from 1 to 7 days later when counting from their timestamps. Similarly, the same segment on the past-related graph part represents the aggregate of tweets with time references referring to the span from 1 to 7 days ago from tweet timestamps. Since the horizontal line denotes the tweet timestamps we should be able to observe any fluctuations in the *temporal attention* over the whole time frame of our dataset. Note that as discussed in Section 4.1 data is missing at certain time points (indicated in the graphs in Figure 3 as blank columns).

To observe any potential calendar effect we indicate by black horizontal lines the starting points of each week (dashed black lines) as well as the starting points of each month (solid black lines). Below the graph we also display the curve of the total number of tweets crawled (in blue) and the percentage of the tweets with time mentions (in orange) to compare the visualization on particular day with the total amount of data used to visualize that day. On the right-hand side of the graph, the blue, vertical curve shows the amount of tweets mapped to every *time diff* interval i.e. the sum of all the cell values on each lines.

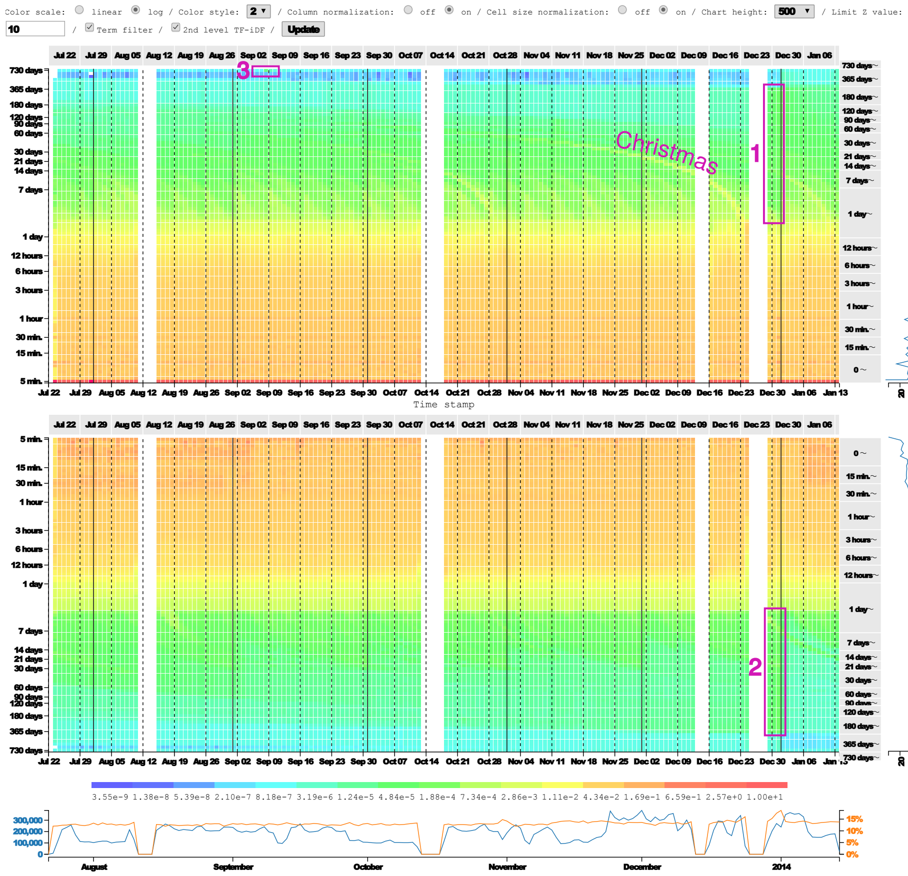
Lastly, we note that our system allows fine-grained investigation of tweets constituting any selected cell. Upon clicking on a given cell a new tab is opened in a browser with the list of top words, top temporal expressions and the table detailing all the attributes of tweets associated with the cell including their text. Similarly, selecting any gray cell on the right-hand side or on the top line will trigger pop-up window with ranked top terms for the cell. The computation of top words will be described in Section 5.4.

Looking at both the future and the past related heat map graphs in both the datasets we observe the pronounced color gradation with some abrupt changes in color drawing parallel to the abscissa which indicate varying intensity levels of future and past references. This shows the clear tendency in human speech to use landmark and idiomatic expressions such as “today”, “this afternoon”, “next year” or “next week” conditioned by Japanese and English languages. It appears that the strongest *temporal attention* of users seems to be mainly related to actions within the period up to the next 1 day in the future and down to 1 day in the past for both the languages (orange and yellow colors). Within these horizons, the activities up to the next hour and down to the past 10-15min seem to be even more often referenced, as especially visible in the US dataset. We observe significant decline in the amount of temporal attention further above or below the lines of 1 day as marked by green or light yellow color areas. In the US dataset we can also observe the effect of Fridays and Saturdays during which users seem to make slightly more plans up to 6-12h ahead (note the periodical, horizontal stripes in orange color) than on other days. In the Japanese dataset this is more blurred.

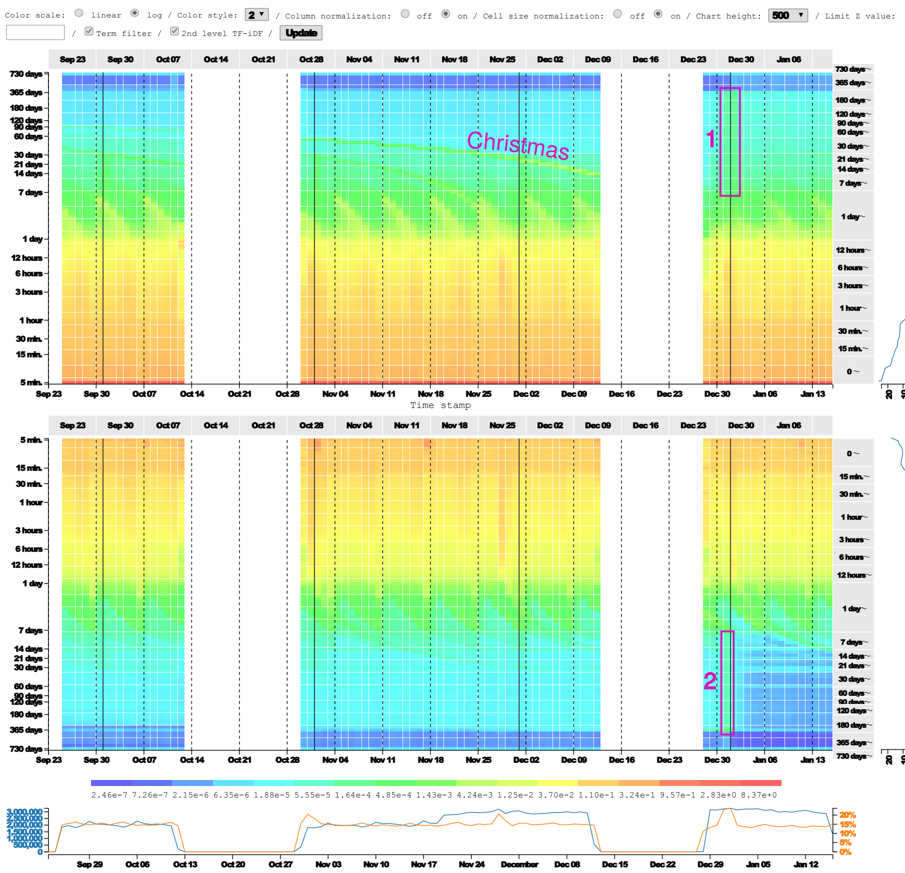
In general, the far past and distant future are characterized by fewer references what confirms our initial hypothesis that the attention decreases the farther one looks into the past and the future.

4.3 Similarity of Past and Future

Actually, the most striking pattern one can observe is the relatively high similarity in the scope of temporal attention related to the future and to the past. Thus at this point it is difficult to say that Twitter users are significantly more interested in the future than in the past or vice versa. This confirms similar observation noticed in large collection of



(a) Snapshot of Japanese dataset available online [34]



(b) Snapshot of US dataset available online [35]

Figure 3: Heat maps of relative mentions of future and past along time.

news articles [13] where the number of references to the near future were only slightly more common than the ones pointing to the near past. Of course, the symmetry is not perfect. There are certain differences noticeable such as curvy lines within the time frame further than several days from the timestamps (more visible in the US dataset). In general, the similarity between the past and future views is rather striking.

Next, we describe the periodical decreasing yellow or green curves in the future and the past-related parts of the graphs. After manual investigation of their top words, we found that they represent the tendency of people to use expressions referring to the end of a week. This is also confirmed by the fact that such lines last for a week when measured on abscissa and that they are vertically symmetrical based on dashed black lines indicating the change of the week. In both the datasets the periodical curves are more stronger in future (upper parts of the graphs) than in the past. When setting the different coloring scale (not shown in this paper) we managed also to observe similar situation for months. In addition, few curves are also visible that are longer than weekly or monthly curves. As we observed they are due to tweets referring to significant temporal landmarks within our dataset such as Christmas (see the "Christmas" label in both the graphs).

Interestingly, we can notice another intriguing symmetry in the future and past-related areas of the graphs centered around the start of the New Year (see areas indicated with labels 1 and 2). In the past-related part (area with label 2 in both the graphs) we interpret it as the process of *remembering of things that happened throughout the last year*. This phenomenon starts to surface several days (and is especially evident just one day) before the end of the year[4]. We can observe continuous green color of the cells on the column representing Dec 31, 2013. Note the rather drastic decrease in the past mentions referring to the time older than one day during the first days of 2014 what emphasizes that people tend to "forget" or, at least, to remember less the events and things of the previous year, right after the new year has come. When looking on the part of the graphs representing the future (label 1 in both the graphs) we conclude that the process of *making plans for the next year* starts also few days before the end of the year but is the strongest on the first day of the next year (Jan 1, 2014) and then lasts for several subsequent days. Same as in the past part of the graphs on Dec 31, 2013, the column representing the first day of the new year in the future area (see area with label 1) has the pattern of green cells spanning almost the whole year (cells on the column of that day and also several next days). It is likely due to users discussing plans of their future activities to be done in various time periods of the new year.

4.4 Variation of Patterns in Temporal Mentions over Time

The next observation is the apparent stability of temporal attention both for the future and for the past, if one neglects the effect of major calendar events such as the Christmas and the New Year's Eve. It means that the time horizons of users are essentially unchanging over the time period of our datasets.

[4]Unfortunately, due to data loss this pattern cannot be seen in its entirety.

When zooming in graphs, we were able to find some collective expectations of future events such as announcement of Tokyo city as a host of Olympics 2020 that was made public on Sunday of Sep 8, 2013 (area with label 3 in the Japanese graph). In the snapshot in Figure 4(a), the effect is not so obvious due to cell normalization effect, described latter in Section 5.3. To observe this effect the cell normalization can be deactivated in the options.

4.5 Topics of Expectations and Memories

We hypothesize that, on average, different expressions should be found when someone refers to events or actions in very near time (e.g., expecting breakfast or remembering lunch meeting at the end of the day) as opposed to happenings associated with more distant time (e.g., planning a vacation trip well ahead or reminiscing it some time later).

We examine then terms related to different temporal scopes. Our system allows seeing the representative words used in tweets that refer to given time frames. These appear in a pop-up window upon pointing mouse pointer over gray buttons on the right-hand side of both future-related and past-related plots. The calculation of word scores is detailed in Section 5. Similarly, for exploration purposes it is possible to see representative words chosen from all the time-refereed tweets on a given week by selecting the gray buttons on the top (bottom) of the future-related (past-related) plots.

Table 1 shows the top representative words for tweets mentioning selected time frames both in the past and the future in the case of the Japanese dataset (we skip the US dataset due to space constraints). We can observe that there are few expressions about the weather conditions for the near future (6-12h ahead), while the far future is dominated by the information about IOC decision to hold 2020 Olympics in Japan. The near past (6-12h ago) is about common things people have just done (ate lunch, studied, returned home, etc.), while the distant past expressions associate with tweets containing recollections and reminiscences.

future over 730 days	Olympics, marriage, live, good luck, Olympics (kanji), Japan, hold, words, decide, imagination
future 6-12 hours	weather, good morning, announcement, forecast, sleep, lunch, information, caution, exam, school
past 6-12 hours	weather, lunch, arrive, practice, sleep, study, announcement, good job (well done), leisure, tired
past over 730 days	nostalgic, change, recall, in those days, I, old, young, remember, know, period

Table 1: Top words in tweets related to near/distant future/past

Lastly, in Table 2 (future) and Table 3 (past) we show top words from aggregated predictions and memories for two different weeks, one in summer (Aug 19-25) and one at the change of the year (Dec 30-Jan 5). We see that for example, in the summer period, the tweets about future often concern weather predictions. There are also words indicating plans for summer holidays or for watching sports matches.

For the past related words in the summer time we observe memories of Obon vacation period (Aug 13-16) and some fireworks events. There are also some words about the earthquakes that happened at that time.

Looking at the top words for the predictions and memories at the time of the year change we observe many customary expressions for greeting the new year, remembering the old year and visiting shrines to celebrate the new year and to thank for the good things in the old year.

Aug 19-25	high temperature, °C, caution, thunders, information, sports match, summer vacation, hot, beach, rain
Dec 30-Jan 5	start of the new year, celebrate, (the new year) begins, first visit of the year to a shrine (typical custom in Japan), new year, welcome the new year and forget the old, shrine, kind regards, new year (kanji), welcome

Table 2: Top words in future related tweets in Aug 19-25 and Dec 30-Jan 5 in the Japanese dataset.

Aug 19-25	obon (holiday period in Japan on Aug 12-16), scary, fireworks, km, happen, deep, earthquake, hot, beach, breaking news
Dec 30-Jan 5	greetings, thank you (for everything this past year), farewell to the old year, new year, old year, (the new year) begins, the last year, congratulations, various things (of this year), year-crossing

Table 3: Top words in past related tweets in Aug 19-25 and Dec 30-Jan 5 in the Japanese dataset.

4.6 Comparison of Time Mentions and Timestamps

We investigate here to which days tweets written on a particular day refer to. Intuitively, the majority of tweets written on a given day should refer to the actions during that day. We test this hypothesis using the visualization shown in the graphs in Figure 4. Figure 4a refers to the Japanese and Figure 4b to the US dataset. The horizontal axis in the graphs denotes tweet timestamps, while the horizontal one denotes tweet mentions, both represented in days. Note that these matrix-like graphs have been computed based on all time expressions in our datasets, hence, not only on time expressions of daily granularity. We perform a Jaccard-style normalization, further described in Section 5.2 that is mix of column-wise and row-wise normalization, allowing us to avoid the effect of a non-uniform distribution of the tweet *timestamps* and *time mentions*.

When looking at the graphs, indeed, we observe the strong impact of timestamp on the time mentions in tweets in the form of pronounced diagonal (red and orange colors) of a width of about 3 cells. This means that many tweets refer to the same day as their timestamp or just 1 day before or after it. In addition, we observe light blue periodical rectangles which fit inside the week cells. A week cell is bounded by dashed line. Keeping in mind that all the cells above the diagonal represent tweets about future and below the diagonal tweets about the past, we conclude that these periodical rectangles portray expectations or memories within the same week as the tweet timestamp. This again confirms the usage of *temporal landmarks* in natural language. Users tend to relate to the future and the past in a way which is bounded by the end and the start of the current week. Also as seen in the graphs there seems to be more attention put on the future than on the past by the users.

Lastly, the three light blue vertical lines in the Japanese and the US datasets correspond to the long awaited major calendar events: Obon holiday (Aug 13-16), Halloween, Christmas and Halloween, Thanksgiving, Christmas, respectively.

Below the matrix graphs of Figure 4 we also display for reference the cumulative "popularity" of days mentioned in tweets, and on the right-hand side we show the frequency of tweets.

5. TECHNICAL ASPECTS

The system has been implemented in SCALA programming language on the server side and using D3 graphical library for enabling the web interface. Kuromoji [31] Japanese morphological analyzer was used for text processing tasks and Stanford CoreNLP [19] was used for text parsing and for detecting time expressions in English.

The following of this section presents the formal definitions of the main aspects of the visualizations we provide. Formally, let T be the set of all the tweets in the dataset, $t \in T$ be a tweet characterized by a bag of words W_t. Visualizations are heat maps discretized in cells, let $CELLS$ be the set of all the cells in a heat map and $C \in CELLS$ be one cell in the heat map. Finally, $C_{i,j}$ is a cell in position i, j in a heat map. Note that all the following measures depend on a given visualization (heat map) with *row* denoting the number of rows and *col* denoting the number of columns.

5.1 Probability Mass Function

According to time mentions in a tweet t the cell $C_{i,j}$ is associated with a probability $P(t|C_{i,j})$ such that $t \in C_{i,j}$. When the time expression is a point in time, $P(t|C_{i,j}) \to \{0, 1\}$. However when the time expression is a period of time, $P(t|C_{i,j}) \to \{weight \in \mathbb{R} : 0 \le x \le 1\}$, following a uniform distribution of the weights of the tweets over the time period discretized according to the granularity described in Section 3.2. Intuitively, $P(t|C_{i,j})$ is understood as the probability that a tweet is pointing to the cell. Hence the sum of the weight of a tweet t in all the cells of a heat map, must be 1 ie. $\sum_{C \in CELLS} P(t|C) = 1$.

Let $I(C_{i,j}) = \sum P(t|C_{i,j}) : t \in T$ be the *intensity* I of a cell $C_{i,j}$. This *intensity* value controls the color displayed in Figure 3 and 4.

5.2 Normalization

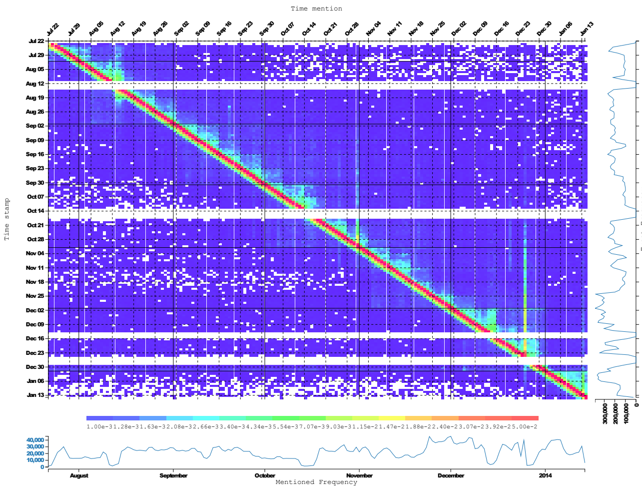
When looking at the blue curves at the bottom of Figure 3, one can remark that the total number of tweets in our dataset at each time is varying from a factor one to three (without considering blank periods). Therefore, without normalization the variation in time would be biased by the effect of peak time. The system provides three possible normalizations. The row-wise normalization allows comparing the cells on the same row: $\text{Score}(r, c) = \frac{I(C_{r,c})}{\sum_i I(C_{i,c})}$

The column-wise normalization allows comparing the cells on the same column: $\text{Score}(r, c) = \frac{I(C_{r,c})}{\sum_j I(C_{r,j})}$

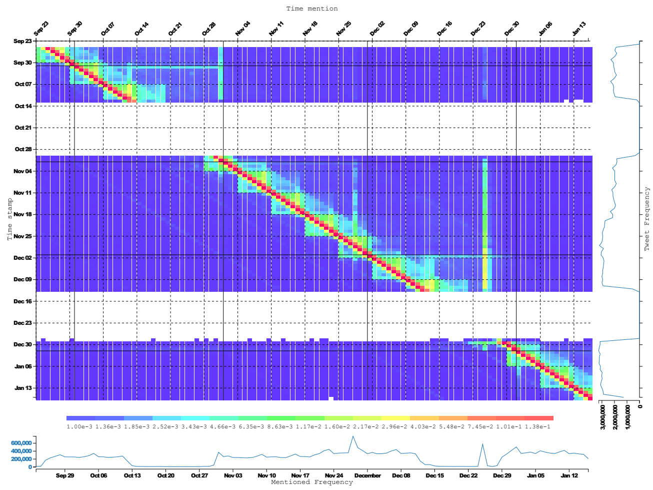
We also developed a Jaccard-style normalization to offer a way to "combine" both normalizations on the same visualization:

$$\text{Score}(r, c) = \frac{I(C_{r,c})}{\sum_i I(C_{i,c}) + \sum_j I(C_{r,j}) - I(C_{r,c})}$$

The Jaccard coefficient is a ratio to measure the similarity between two sets. It is computed as the ratio between the



(a) Snapshot of Japanese dataset available online [36]



(b) Snapshot of US dataset available online [37]

Figure 4: Heat map of timestamp vs. time mention matrix

intersection and the union of the two sets. In this application, for a given cell $C_{r,c}$ the two sets are, the set of all the tweets in cell of row r, and the set of all the tweets in cell of column c. Hence, it actually measures how much the cell is representative for the two values r and c on the axis x and y of the heat map.

These normalizations are optional, and the user can choose which one to apply to spot different pattern in the visualizations.

5.3 Normalization of Logarithmic Scale

Here we address the problem of time visualization on a logarithmic scale. To illustrate the problem let's consider an extreme case where all the tweets at a given timestamp ts have the time expression "this month". In such case we would expect that the column ts would have cells spanning from beginning of this month until the end displayed in one uniform color and all the rest of cells in white. However because the ordinate axis is in a logarithmic scale, cells farther from the origin (more distant past or future) represent a larger time range than cells closer to the origin (near past or future). They will thus aggregate much more tweet's weights and, consequently, will be represented with a much stronger color, giving the false impression than the far future and past are stronger referred to than they actually are. This is because the weight of the tweets will be uniformly distributed over time covered by the expression "this month".

Let $I(C_{i,j}) = \sum P(t|C_{i,j}) : t \in T$ be the *intensity* I of a cell $C_{i,j}$. Intensity represents a raw measurement, as the sum of the probability of the tweets in the dataset to appear in a cell. Therefore intensity should not be used directly to represent colors. Instead we normalize it according to the scale of the axis. The color of a cell is quantified as the *intensity* of that cell divided by time range of that cell:

$$color(C_{i,j}) = \frac{I(C_{i,j})}{size(C_{i,j})}$$

This normalization is called cell normalization and can be deactivated in the visualization options online.

5.4 Ranking Words

For each visualization, the system provides a ranking of the words in an arbitrarily defined area on the heat map called a region. The goal is to keep a top-k list of the most characteristic words for the region. The weight of each word is based on a TF-iDF weighting scheme to represent how much the word is characteristic for the given region. Intuitively, TF-iDF assigns high scores to terms that appear often in tweets associated with a given cell, while appearing infrequently in other cells. The novelty in our approach is in the count of term frequency using probability and in the definition of document based on the two levels of granularity defined for the regions: *cell* and *segment*. They are described in the subsequent paragraphs.

Cell.

In order to find the most relevant words appearing in tweets for each cell we measure the word relevance using a modified TF-iDF score where a term is a word and a document is a set of tweets corresponding to a given cell. The score to rank a word w in a cell $C_{i,j}$ using the whole set of tweets T is:

$$\text{Score}(w, C_{i,j}, T) = \frac{\sum_{t \in T} P(t|C_{i,j}) : w \in W_t}{\sum_{t \in T} P(t|C_{i,j})} \times \log \frac{|CELLS|}{|C \in CELLS : \exists t \in C : w \in W_t|}$$

Segment.

In the visualization in Figure 3, the gray cells on the top or bottom line and on the right-hand side are regarded as virtual documents that represent *segments*. For instance, the top left gray cell in Figure 3 with label "Jul 22" gathers all the cells between column July 22 to 28 on every row. The *segments* are either row-wise or column-wise; they are sets of cells that span respectively, on all columns covering the width of the selected rows, or on all rows covering the width of the selected columns.

A second level TF-iDF score is computed to rank top words in *segments.* The content of a virtual document (segment) is the set of words W_{S_i} built from all the words of the cells in that *segment.*

Let *SEGMENTS* be the set of all the horizontal segments in a heat map and $S \in \textit{SEGMENTS}$ be one horizontal segment in the heat map that is s set of all the tweets appearing in the cells of that segment. Finally S_i is a horizontal segment in position i in a heat map. The weight of a word w in a segment is the sum of the weights of the tweets in which the word w appears.

$$\text{Score}(w, S_i, T) = \frac{\sum_{t \in T} P(t|C) : C \in S_i : w \in W_C}{\sum_{t \in T} P(t|C) : C \in S_i} \times \log \frac{|SEG|}{|S \in SEG : \exists t \in S : w \in W_t|}$$

6. DISCUSSION

We will first list the important observations we noticed using our system. Then we enumerate several directions for future research.

6.1 Key Findings

Amount of Temporal Expressions. We have found quite many temporal expressions in tweets. They can be used for anchoring tweet content on the timeline useful for many tasks including information retrieval, summarization, trend detection and so on.

Short Time Horizons and Supremacy of the Present. As expected, tweets exhibit clear temporal constraints. The attention to distant time periods as measured by the amount of extracted temporal references is significantly lower than the one to the near time. The present when more generally defined by yesterday, today and tomorrow, dominates the time-span of the temporal attention. This observation confirms conclusions from some of the related studies on news articles or on the open web [13, 14].

Similarity of Remembering and Expecting. We have noticed strong symmetry in the visual maps of temporal references to the past and to the future. The time horizons and the patterns of temporal attention are quite similar with the future taking however more attention of the users than the past.

Temporal Landmarks. As observed users tend to use temporal landmarks such as the starts and ends of weeks or months. The strongest landmark is the end and the start of the year.

Relative Stability of Temporal Horizons. We observed that the patterns of attention are more or less stable over time apart from few exceptions (e.g., the time periods just before and after the New Years Eve). That is, for different periods of collected data there is relatively little variation in the distribution of temporal attention and the lengths of temporal horizons.

Diversity of Discussed Topics over Time. We observe certain variations in words used when referring to diverse time distances. Different time segments are characterized by differing sets of activities. All day life activities dominate the present, as shown in Table 1 but some verbs and popular events are typical of far future and past.

Cultural Differences. Some cultural variations can be observed by comparing views of the two datasets. The most obvious ones related to different calendar events (e.g., Thanksgiving day in USA). Nevertheless, on average more similarities can be observed.

Lastly, we note that our study is limited to half a year long time period for the Japanese dataset and 4 months long period in case of the US dataset. It would be advantageous to compare the results over longer datasets such as ones spanning a year or even few years. We plan to work on these issues in the future.

6.2 Applications and Future Extensions

Future Prediction. As mentioned before, the expressed expectations can be utilized for predicting user activities to support any pro-active, reactive or adaptive systems. By studying the visualization graphs the feature engineering for classifiers can be improved for any activity prediction systems. For example, to predict user future mobility (e.g., holiday travel) we would expect the references to the travel to appear long before the departure date. Thus any evidence data hinting travel plans to be used for creating classifier features should be searched within certain time ahead the start of the travel with corresponding granularity. Also, the typical variations in temporal attention that we observe in the graphs of Figure 3 suggest particular division of time into units from where features can be extracted. Instead of uniform time division one could use the one adopted to the significant changes occuring in temporal attention as portrayed by our visualizations. Obviosuly, different time divisions will result in different classification features collected and should impact prediction accuracy.

Recommendation. Based on captured user memories or expectations future systems could better suggest particular activities, products, services or locations to visit.

Sentiment analysis. It is known that humans tend to retain positive memories more than negative ones (*rosy retrospection* [21]) and favor positive future views than negative ones (*valence effect* [25]). It is then interesting to perform sentiment analysis to confirm any potential sentiment biases.

Verifying expectations and predictions. It should be possible to evaluate the predictability of future expressions and credibility of future plans. For example, for a user tweeting about travelling to Hawaii in the next month, we might search for any tweets of this user with GPS location confirming the fact of visiting the islands one month later.

Social computing. As future work we will extend our system to accept any datasets for visualization. This should be useful for social scientists who lack efficient means for making sense of their data.

7. CONCLUSIONS

In this paper we studied the dynamics of expecting and remembering processes characteristic to social media users based on data collected over nearly half a year long timespan for the Japanese and 4 months for the US datasets. Our focus is a large population of microblogging users who continuously share online information about their daily lives and things that matter to them. We have demonstrated visual framework to portray particular ways in which users refer to the past and to the future. Our methodological approach allows observing the scope of temporal attention of users, temporal horizons of their perspectives, the typical expressions and topics associated with references to given time frames (e.g., near, distant past or near, distant future) and other related features. In the future we plan to work on the previously listed directions.

References

[1] Omar Alonso, Ricardo Baeza-yates, Jannik Strötgen, and Michael Gertz. "Temporal Information Retrieval: Challenges and Opportunities". In: *TempWeb'11*. 2011.

[2] Sebastien Ardon, Amitabha Bagchi, Anirban Mahanti, Amit Ruhela, et al. "Spatio-temporal and Events Based Analysis of Topic Popularity in Twitter". In: *CIKM*. 2013.

[3] S. Asur and B.A. Huberman. "Predicting the Future with Social Media". In: *WI-IAT*. Vol. 1. Aug. 2010.

[4] Klaus Berberich, Srikanta Bedathur, Omar Alonso, and Gerhard Weikum. "A Language Modeling Approach for Temporal Information Needs". In: *ECIR*. 2010.

[5] Ricardo Campos, Gaël Dias, Alípio M. Jorge, and Adam Jatowt. "Survey of Temporal Information Retrieval and Related Applications". In: *ACM Comput. Surv.* 47.2 (2014), 15:1–15:41.

[6] Yan Chen, Hadi Amiri, Zhoujun Li, and Tat-Seng Chua. "Emerging Topic Detection for Organizations from Microblogs". In: *SIGIR*. 2013.

[7] Au Yeung Ching-man and Adam Jatowt. "Studying How the Past is Remembered". In: *CIKM*. 2011.

[8] Hermann Ebbinghaus. "Uber das Gedchtnis. Untersuchungen zur experimentellen Psychologie." In: (1885).

[9] Hermann Ebbinghaus. *Memory: A Contribution to Experimental Psychology*. Columbia University. Teachers College. Educational reprints. no. 3. 1913.

[10] Maurice Halbwachs. *La mémoire collective*. Les Presses universitaires de France. 1950.

[11] Christoph Hoerl and Teresa McCormack. *Time and memory : issues in philosophy and psychology*. No.1. 2001.

[12] R. Jacoby. *Social Amnesia: A Critique of Contemporary Psychology*. 1997.

[13] Adam Jatowt and Ching-man Au Yeung. "Extracting collective expectations about the future from large text collections". In: *CIKM*. 2011.

[14] Adam Jatowt, Hideki Kawai, Kensuke Kanazawa, Katsumi Tanaka, et al. "Analyzing Collective View of Future, Time-referenced Events on the Web". In: *WWW*. 2010.

[15] Nattiya Kanhabua, Klaus Berberich, and Kjetil Nørvåg. "Learning to Select a Time-aware Retrieval Model". In: *SIGIR*. 2012.

[16] Nattiya Kanhabua and Wolfgang Nejdl. "Understanding the Diversity of Tweets in the Time of Outbreaks". In: *WWW*. 2013.

[17] Nattiya Kanhabua and Kjetil Nørvåg. "Determining Time of Queries for Re-ranking Search Results". In: *ECDL*. 2010.

[18] David Lazer, Alex (Sandy) Pentland, Lada Adamic, Sinan Aral, et al. "Life in the network". In: *Science* 323.5915 (Feb. 6, 2009), pp. 721–723.

[19] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, et al. "The Stanford CoreNLP Natural Language Processing Toolkit". In: *ACL*. 2014.

[20] M. J. Metzger. "Making sense of credibility on the Web". In: *JASIST* 58.13 (2007), pp. 2078–2091.

[21] Terence R. Mitchell, Leigh Thompson, Erika Peterson, and Randy Cronk. "Temporal Adjustments in the Evaluation of Events: The "Rosy View"". In: *Journal of Experimental Social Psychology* 33.4 (1997), pp. 421–448.

[22] Sérgio Nunes, Cristina Ribeiro, and Gabriel David. "Use of Temporal Expressions in Web Search". In: *ECIR*. 2008.

[23] Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. "Learning Causality for News Events Prediction". In: *WWW*. 2012.

[24] Kira Radinsky and Eric Horvitz. "Mining the Web to Predict Future Events". In: *WSDM*. 2013.

[25] David Rosenhan and Samuel Messick. "Affect and expectation". In: *Journal of Personality and Social Psychology* 3.1 (1966), pp. 38–44.

[26] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. "Earthquake shakes Twitter users". In: *WWW*. 2010.

[27] George Valkanas and Dimitrios Gunopulos. "How the Live Web Feels About Events". In: *CIKM*. 2013.

[28] W. Weerkamp and M. de Rijke. "Activity Prediction: A Twitter-based Exploration". In: *Proceedings of TAIA Workshop*. 2012.

[29] Jianshu Weng and Bu-Sung Lee. "Event Detection in Twitter." In: *ICWSM*. 2011.

[30] Jaewon Yang and Jure Leskovec. "Patterns of Temporal Variation in Online Media". In: *WSDM*. 2011.

Online Resources

[31] *Kuromoji*. Japanese morphological analyzer. URL: `http://www.atilika.org/`.

[32] *Portal to visualizations of the Japanese dataset*. URL: `http://www.cwb-net.jp/~yasuhara/kyoto-su/tweetDelta/JP4/chartList.html`.

[33] *Portal to visualizations of the USA dataset*. URL: `http://www.cwb-net.jp/~yasuhara/kyoto-su/tweetDelta/US4/chartList.html`.

[34] *Location difference over time for Japanese dataset*. URL: `http://www.cwb-net.jp/~yasuhara/kyoto-su/tweetDelta/JP4/TweetHeatMap_time.html`.

[35] *Location difference over time for US dataset*. URL: `http://www.cwb-net.jp/~yasuhara/kyoto-su/tweetDelta/US4/TweetHeatMap_time.html`.

[36] *Time transition matrix for Japanese dataset*. URL: `http://www.cwb-net.jp/~yasuhara/kyoto-su/tweetDelta/JP4/TweetHeatMap_dateMatrix.html`.

[37] *Time transition matrix for US dataset*. URL: `http://www.cwb-net.jp/~yasuhara/kyoto-su/tweetDelta/US4/TweetHeatMap_dateMatrix.html`.