

2011

# “What are they Thinking?” - Accessing Collective Intelligence in Twitter

Martin Böhringer

*Technische Universität Chemnitz, Germany, martin.boehringer@wirtschaft.tu-chemnitz.de*

Patrick Helmholz

*Technische Universität Braunschweig, Germany, p.helmholz@tu-bs.de*

Follow this and additional works at: <http://aisel.aisnet.org/bled2011>

---

## Recommended Citation

Böhringer, Martin and Helmholz, Patrick, "“What are they Thinking?” - Accessing Collective Intelligence in Twitter" (2011). *BLED 2011 Proceedings*. 49.

<http://aisel.aisnet.org/bled2011/49>

This material is brought to you by the BLED Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in BLED 2011 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

## **“What are they Thinking?” - Accessing Collective Intelligence in Twitter**

**Martin Böhringer**

Technische Universität Chemnitz, Germany  
martin.boehringer@wirtschaft.tu-chemnitz.de

**Patrick Helmholz**

Technische Universität Braunschweig, Germany  
p.helmholz@tu-bs.de

### **Abstract**

*In today's social networks like Twitter and Facebook, each day millions of status updates provide a huge source of information on current sentiments of their users. However, still it is unclear how to structure problems in a way that they can be answered based on social networking activities. Facilitating an exploratory prototype, we tested the direct retrieval of user opinions during the FIFA World Cup 2010 in Twitter. Our findings based on an initial research framework suggest that using social networks can serve as an access point to collective intelligence. However, the experiment also showed weaknesses of the used approach. Therefore we discuss an updated research model based on our results which provides a foundation for future works.*

**Keywords:** social Network, microblogging, twitter, collective intelligence.

## **1 Introduction**

Microblogging has shown to be a sustainable trend in information systems. While microblogging today is a standard usage pattern of the web's largest social networks like Facebook and Twitter it has also seen a strong adoption of enterprise software vendors as well as researchers. Faster than previous web 2.0 applications, microblogging found its way into organisations with multiple “enterprise microblogging” vendors providing their services to companies.

While the reason for this interest by organisations may be based on expected positive influence on factors like company culture and knowledge management (Riemer et al. 2010) the strong focus of the research community on microblogging can be explained by the openness of this media. Few internet information systems have provided the possibility to dive into the social behaviour of so many people. Therefore, based on a number of broad studies, today we have a sound body of knowledge about these social

networks. We especially know that systems like Twitter and Facebook provide an access to the collective intelligence of their users. This insight already has been confirmed in studies about areas like political events (Shamma et al. 2009), natural disasters (Vieweg 2010) and marketing (Wagner & Strohmaier 2010).

In this paper we are going to address this problem of gaining insights in form of thoughts and feelings from large social networks. Previous research has shown the possibility to get answers from these communities. However, we do not know how to ask the right questions yet. Our considerations combined with a practical experiment during FIFA World Cup 2010 aim at providing a framework on how problems can be structured in a way that status updates (i.e. microblogging) from social networks can answer it.

The following chapter provides an overview on status updates in social networks including a special view on Twitter. Next, we describe the considerations which led to the initial assumptions. They provide the foundation for the exploratory experiment during FIFA World Cup 2010 which is discussed in chapter 4. Chapter 5 includes a critical review of the experiment and its results. Finally, the following chapter discusses implications and provides an updated research model before the conclusions end the paper.

## **2 Collective Intelligence in Microblogging**

Microblogging has become an integral part of our understanding of social networks. Twitter and Facebook, which both use status updates as the kit between their users, belong to the largest social networks in the internet. Especially Twitter has created a standard for our understanding of status updates including its own vocabulary like the verb “tweet” for posting such an update. Other important functionalities of Twitter include the “retweet” (re-posting somebody else’s tweet), referring another user with the @-operator (“@<username>”) and the hashtags which mark certain meta-data inside the tweets (“Submitting a paper to the #bled conference.”). It is important to point out that most of these unique functionalities have been introduced by Twitter’s users before being implemented in the official user interface. This is a first proof for the dynamic and crowd intelligence which can be observed on this platform (Jeners & Mambrey 2010).

Previous research has shown that Twitter usage differs very much from user to user (Java et al. 2007). This includes such factors like posting frequency, posting structure (usage of hashtags, formal or informal language) and content (e.g. emotions, information, links, questions). An important indicator for these variables is the user’s social network which drives adoption and usage (Barnes & Böhringer 2009).

No matter what and how users tweet it mostly is about current topics. Shamma et al. (2009) showed that the debates during the US election campaigns had significant influence on the conversations on Twitter. Further, in analysing Twitter updates it has also been possible to get information on the debate’s topic. Therefore, Cheong & Lee (2010) call Twitter a “mirror” of current events of user’s interest. This opens several use cases for using this information. Especially marketing could benefit from the early and broad feedback of such a huge social network. According to a study by Jansen et al. (2009) up to 19 percent of tweets include brand names. Often the information about frequently used words is enough to create an accurate overview on currently important topics (Cheong & Lee 2009). Combined with location-based information (e.g. GPS

coordinates) this created new possibilities for information retrieval e.g. in case of emergency events or natural disasters (Vieweg et al. 2010).

A common problem for the analysing of tweets is its character as short text snippets with often informal language and abbreviations (see for example the discussion by Pak & Paroubek 2010 as well as Pandey & Iyer 2009). Another critical problem which has to be kept in mind for all analyses of social networks is its reliability. Authenticity of users is a serious problem which cannot be guaranteed.

### **3 Assumptions**

The previous chapter discussed recent research which has shown the general applicability of microblogging analyses. Based on this insight, our basic research question is: What kind of question can be answered in analysing microblogging networks? Most known examples base on word analysis, mostly on simply counting the most used words or trending topics in a certain time frame. Our interest is in more complex questions like “Who will win the football game tonight?” Our initial approach to answer such a question would be finding microblogging postings which address this question (e.g. tonight’s game) and interpret them with a problem-specific routine (e.g. which is trained for finding score predictions). This leads us to the following working assumptions:

A1: Due to the mass of microblogging postings, there will be a significant number of postings which include the answer for even special questions.

A2: Leveraging a problem-specific routine can find the answer to questions from tweets.

We used an exploratory experiment to test these working assumptions. The experiment and its results are described in the next chapters.

## **4 Exploratory Experiment**

### **4.1 Overview**

The expectations on the project “Ballgezwoitscher” (German for “ball tweets”) are based on the already described insight that large public events like the FIFA World Cup lead to a particularly strong Twitter activity. The idea behind the project was to filter tweets with game predictions for every single game of the world cup, to analyse and aggregate them and present them on the project page. This data and the insights drawn from the experiment itself should provide a sound foundation for assessing the assumptions.

On the corresponding webpage the number of received tweets, the five most frequently predicted results and the average score are posted. Additionally, a pie chart indicates the chances of victory and draw of the teams. After the game has been finished, the view will be complemented next to the result with a qualitative indication of accuracy (correct result, correct trend, wrong prediction). Figure 1 illustrates this using the example of the game between Argentina and Germany. As an additional service in the world cup’s group stage the mathematically rounded average score was used to set up the several groups’ current table based on the Twitter prognosis and to compare it with the table based on the actual results.

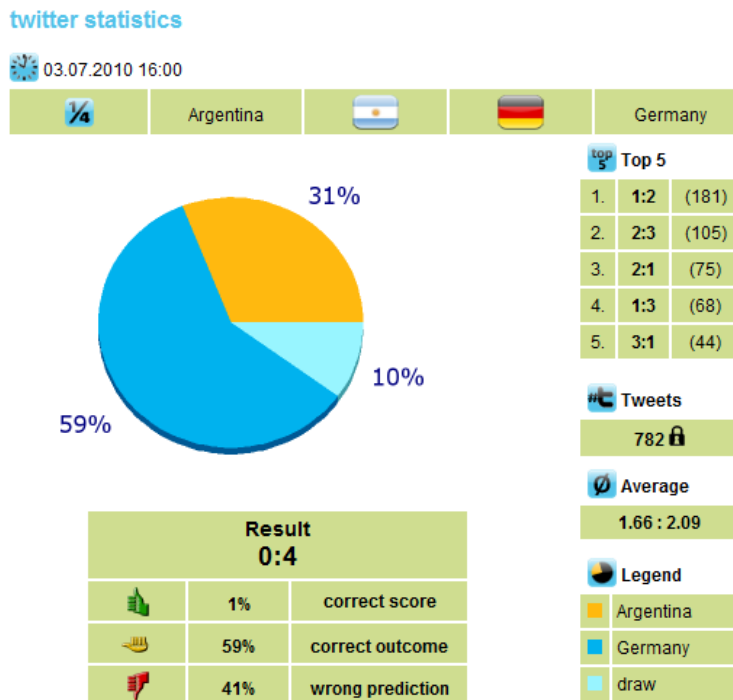


Figure 1: Presentation of the scores

The user's adoption of twitter for discussing the world cup was enormous. The increased use of Twitter even led to periodic breakdowns of the service. In retrospect published statistics by Twitter confirm this assumption and show that the FIFA World Cup is the event with the highest Twitter activity since the start of the service. Here, especially the final match is much emphasized. During the final match between two and three thousand tweets per second were recorded. People from 172 countries in 27 different languages tweeted about the match (Graves 2010).

## 4.2 Prototype's Concept

As technical foundation for retrieving the data the Twitter Search API was used for this project. The search queries are triggered by various hashtags and the search results are evaluated by game predictions. In addition to #wm2010 and similar spellings the tweets are filtered, stored in a database and analysed (for a list of used Tweets see Table 1).

The presentation of the aggregated data is realized by a web application with PHP and Java charts (see Figure 2). Beneath one of these hashtags two team names of the world cup participating teams (i.a. Argentina), the official abbreviations (here: ARG) or a "team nickname" (here: Gauchos) are needed in the Tweet as well as a result in form of two single-digit numbers separated by a special sign (e.g. "1:2"). Once the game kicks off, the filtering of other forecasts is locked, so the result is not affected by ongoing live events.

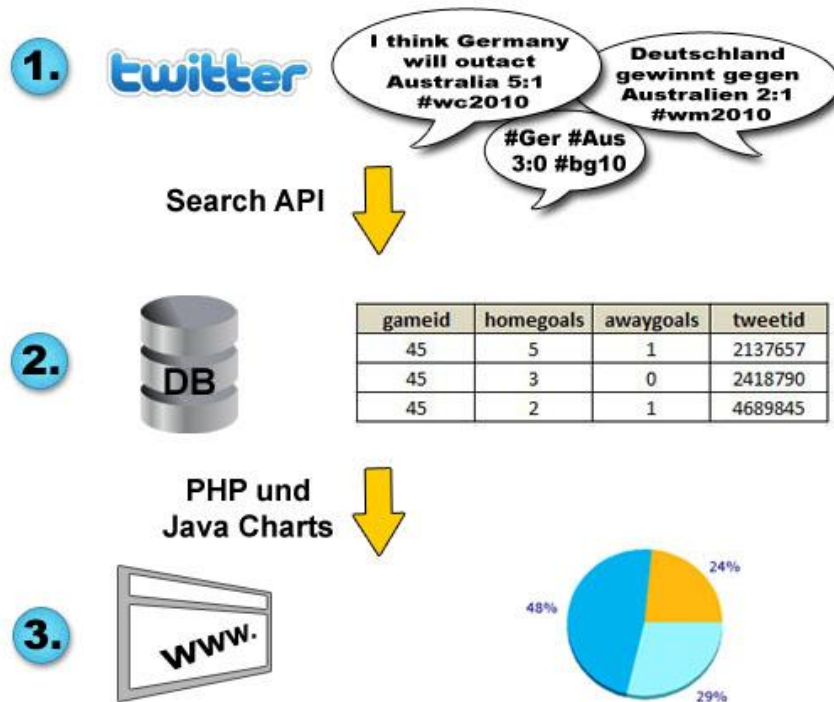


Figure 2: Page formats

### 4.3 Experiment

Ballgezwitcher was planned as an exploratory experiment, therefore in course of this project adaptations based on current insights have been allowed and welcome. Accordingly, an iterative process model was applied. After each round of the group stage, the obtained results have been evaluated and adjustments were made. This particularly affected the selection of the hashtags used for searching and the algorithm for finding game tips.

At the start of the project, a simple algorithm was used, which arranged a matching of predefined strings by Regular Expressions (regexps). A Regular Expression is a kind of language defining notation which provides a concise and flexible means for matching strings of text. It serves the description of sets and subsets of strings with the help of certain syntactic rules. Regexp thus represent a filter criteria for texts, where the respective regular expression is adapted with the text in form of a pattern (Hopcroft 1994). In using these regexps, the terms to filter the Twitter messages have been defined.

The used hashtags have been extended in particular to mention Twitter world cup tipping games. These tipping games called on to give tips via Twitter by using their hashtag and thereby participate in a special contest or lottery. Because of the structured form of these Tweets, the evaluation was easy to handle. The supported hashtags of these prediction games were #cuptweets (<http://www.cuptweets.com/>), #twitterwm (<http://www.twitterwm.de/>) and #sporawktweets (<http://www.sporza.be/>) and in addition game predictions in form of a reply @WBetcha (<http://wbetcha.com/>).

Furthermore, over the group phase in addition to the team names in German, English, French and Spanish team names have been included, which also took account of

linguistic features (e.g. the term "1/3" as a common equivalent for "1:3" in the Latin American region).

## 5 Results

### 5.1 Quality of retrieved opinions

During the worldcup we retrieved 24,512 tweets which our algorithm classified to contain tips. The most frequented game had 794 tweets (Netherlands vs. Brasilia), while the least detected number of tweets was 168 (Netherlands vs. Slovakia). The mean value was 383 tweets per game, which provided a good foundation for the forecast. As described by the previous chapter, we only retrieved tweets which were tagged with certain worldcup-related tags. The following table provides an overview on the frequencies of the used tags (note: as more than one hashtag may be used in a tweet the sum is over 100%):

**Table 1:** Hashtag frequency

Hashtag	Number of tweets	
	nominal	relative
all	24.513	100%
containing #wm10	7	0,03%
containing #wc10	15	0,06%
containing #bg10 (in reference to the experiment's German title „Ballgezwoitscher“)	114	0,47%
containing #wm2010	177	0,72%
containing @WBetcha	220	0,90%
containing #wc2010	429	1,76%
containing #sporawktweets	2.052	8,37%
containing #twitterwm	7.917	32,30%
containing #cuptweets	9.373	38,24%
containing #worldcup	10.641	43,41%

The correctness of our recognition algorithm has been checked manually in two groups. The tips of the group stage (18,823) were analysed based on a random sample of 500 tweets. Tips of the world cup's knockout stage (5,690) were analysed completely. We already mentioned the big share of tipping games in our sample set. Together they were responsible for nearly 80% of our tips. This had great influence on the general correctness of tip recognition as tipping game tweets were much more structured and have been written with computer-based interpretation in mind. Not surprisingly, the correctness of the algorithm was 100% for both groups for the tipping game tweets. The analysis of the free-form tweets showed a nonetheless good result of 15% (group stage) respectively 18% (knockout stage) falsely classified tips.

Finally, the degree of truth of our prognoses showed that football is a very unpredictable sport and Twitter's users cannot see in the future better than others. At least 48% of our tips had the right tendency (win/loose/standoff) which is above normal distribution. The correct result including the right number of goals for each team was found in 8% of the

tweets. These results are supported by the numbers of the similar project “Wisdom-of-Fans” (46%, 7%) which have been reported in Wisdom of Fans (2010).

## 5.2 Typical problems

In general there were three types of classification problems in the free-form tweet. (1) Mostly, a tip was connected to an incorrect game. (2) A game result has been classified to the correct game but the result was not meant to be a prediction. (3) The result was not recognized correctly (e.g. shifted goals).

The main reasons for these problems were:

- **Several facts in one tweet:**
  - (1) “chutes com ajuda do polvo: #esp 2x0 #ned , #ger 3x1 #uru #worldcup (Uruguai, cope o polvo :P); “ (2) „All #worldcup games so far are really low scoring 0-0 , 1-0. #Germany will change that soon enough against #Australia. Poor Soccerroos :(“
- **Historical tweets (wrong semantic interpretation):** „Denmark winning on penalties after a 2-2 draw in the 1992 Semifinals and the Netherlands winning 3-0 in the 2000 Group Stage. #WorldCup”
- **Review of a current game and preview of the next (wrong semantic interpretation):** (1) „Full-time #Ned 2-0 #Den. Good start from the Dutch - they look like they might be up for this #worldcup. Next up #Jap v #Cam!”; (2) „Ghana beats Serbia 1-0 on late penalty shot. Germany-Australia coming up <http://bit.ly/avIZ0J> #pittsburgh #worldcup”
- **Tweet’s structure makes it difficult to parse (here: Holland – Denmark – Netherlands):** „Holland vs Denmark, My predictions 3-0 in favor of The Netherlands #WC2010”
- **Meta-Tweets of aggregators:** „Over 2200 predictions in for the #ger v #aus game and just less then 2 hours to kick off. Most predicted scoreline 2-0 ger #worldcup”
- **Retweets (from bots):** (1) „RT @zwyback: achtelfinale - jetzt gehts gleich los @ZDFsport #esp 0:0 #por #worldcup #-wm2010” (durch wmbot); (2) „3:0 RT @rk030 RT @websenat Mein Tipp #eng2 #usa0 #twitterwm ~ Mein Tipp: #eng3 #usa1”

## 5.3 Assessment of the assumptions

Basically, the experiment supported both assumptions. It was possible to retrieve a clear tendency of game results from the Twittersphere. There have been several hundreds of postings containing such a prediction, which supports A1. Our pretty simple problem-specific algorithm was able to interpret most tweets correctly, which supports A2.

However, there are some limitations. First, we chose a very popular topic for the experiment. The world cup generated a huge amount of tweets. That being said, a few hundred postings are a quite modest result. For other problems like stock market prediction our research design could lead to a very low number of relevant tweets. Therefore support of A1 can be seen to be very weak.

The same limitation applies to A2 as football results are a very easy algorithmic problem. Our experiment does not test the assumptions for more complex problems which require a rich textual understanding of the Tweets. As our experiment showed



nonetheless problems in classifying Tweets (see 5.2) the applicability of algorithms to more complex tasks stills needs testing.

## **6 Discussion**

### **6.1 Experiment’s setting**

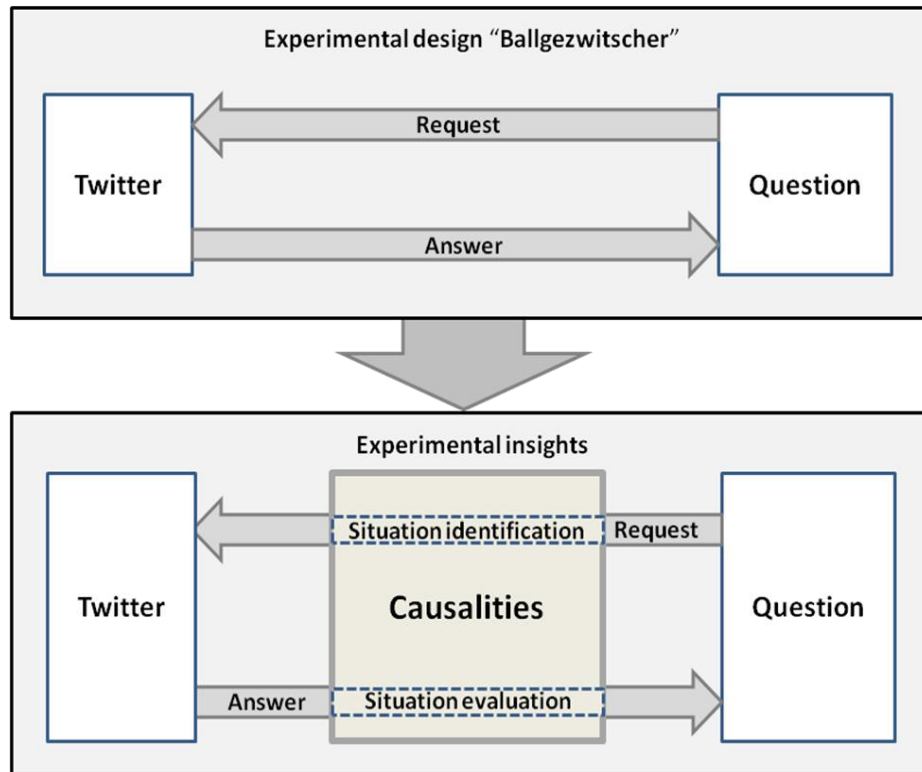
While the explanatory experiment gave us the opportunity to learn about collective intelligence in Twitter, it has some limitations. First, it did not include all available postings from Twitter (see 4.2). Retrieving and analysing the complete public timeline of Twitter would have been a huge technical effort. Planning the experiment we therefore decided to select only tweets with certain hashtags (see table 1) as (1) implementation was very easy using Twitter’s search API and (2) in filtering for world cup-related tweets we would have been able to avoid similar tweets containing game predictions of other sport events.

However, this conceptual decision brings a problem as it could have filtered the number of tweets too much. Boyd et al. (2010) found that only 5% of tweets contain hashtags. Given a normal distribution the experiment therefore excluded 95% of world cup-related postings. People not using hashtags could write their tweets in different ways then users including hashtags. Further, focusing on hashtags leads to a high value of structured tweets from bots or in our case tipping games. While this is an integral part of the Twittersphere and therefore a valuable input for our analysis, it could have distorted the results.

### **6.2 Assessing collective intelligence in Twitter**

As we discussed in chapter 5.3 the general assessment of our assumptions is positive. However, the encountered problems and especially the low number of usable tweets motivated us to rethink the working assumptions and to formulate an updated research model. The assumptions were built on the idea of directly asking Twitter “What will happen in future?”, e.g. “What will be the result of the next match?” Recently, several researchers found that Twitter is especially used to discuss on current events rather than future ones. Reinhardt et al. (2009) showed that for conference tweets, Shamma et al. (2009) for political debates and Vieweg et al. (2010) for natural hazards. This suggests that tweets are reactive in that they occur shortly after or while something happened (Shamma et al. 2009). This is supported by a study from Oulasvirta et al. (2009) who state that Twitter especially contains postings on „current activities and experiences“. The experiences of the Ballgezwoitscher experiment support this insight. We found much more tweets during the first minutes of a game than before its start. However, due to our initial research model they have not been counted.

These results suggests asking Twitter “What is happening now?” instead of “What will happen in future?”. To predict future events Twitter’s collective intelligence therefore has to be assessed using known causalities. In asking “Who will win the game?” such a possible causality would be “The team with the most tweets will win”. In counting such tweets before and during the game it would be possible to generate a prediction based on much more tweets.



**Figure 3:** Insights of the project

### 6.3 Use cases

Twitter offers access to a large amount of recent statements by its users. The analysis of this information is interesting for a number of use cases.

First, assessing Twitter's collective intelligence is an important use case for marketing. Possible scenarios are the evaluation of product announcements or general customer sentiments. Besides finding forecasts especially actual situations of concerns such as brand popularity and target group identification are typical use cases. Apart from these examples, Twitter is a valuable tool to avoid risks because of its speed. In the last time many negative corporate news found their starting point in Twitter, whereupon due to many re-tweets traditional media could become aware of a grievance. The permanent monitoring and evaluation of the company's own published Tweets therefore seem to be a good idea especially for companies in the retail business.

The collective intelligence of Twitter users combined with the often existing location based information could be helpful in many other cases. One use case is the possibility to conclude stock share values volatility based on the discussions on Twitter. Following this example, the early detection of trends is conceivable in various fields, for example, disaster detection and the tracking of flu epidemics (Hughes & Palen 2009).

A special use case applies in internal microblogging (enterprise microblogging). If an organization has an internal microblogging system, many of the previously discussed use cases are transferable. For example, the supplier "Socialcast" (<http://socialcast.com>) provides analysis on the key participants of the platform, trends in various departments and additional services labelled "Social Business Intelligence". Early recognition of

project risks or forecasts of cultural developments (e.g. on the basis of the "moods" in the micro-blogging system) can be adjusted applications in this context.

## **7 Conclusions**

How can we assess Twitter's collective intelligence? This was the initial research question of this paper. Based on a simple and naive set of assumptions we tested the question in the Ballgezwoischer experiment. The experiment itself can be considered to be a success with nearly 25.000 analysed tweets and a high correctness in the tweet analyses. However, based on our insights we redeveloped the working assumptions and provided a more complex research model for assessing Twitter's collective intelligence. This model provides a foundation for future projects both in research and practise.

## **References**

- Barnes, S. J., & Böhringer, M. (2009). Continuance Usage Intention in Microblogging Services: The Case of Twitter. In Proceedings of the 17th European Conference on Information Systems (ECIS), Verona, Italy, June 08-10.
- Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In HICSS-43. IEEE: Kauai, HI, January 6.
- Cheong, M., & Lee, V. (2009). Integrating Web-based Intelligence Retrieval and Decision-making from the Twitter Trends Knowledge Base. In SWSM (pp. 1-8).
- Cheong, M., & Lee, V. (2010). Twittering for Earth : A Study on the Impact of Microblogging Activism on Earth Hour 2009 in Australia. *Earth*, 114-123.
- Graves, Matt (2010). The 2010 World Cup: a Global Conversation, Twitter [Online], 15.07.2010. Retrieved from <http://blog.twitter.com/2010/07/2010-world-cup-global-conversation.html>.
- Hopcroft, J. E., Motwani, R., Ullman, J. D. (1994). Einführung in die Automatentheorie, formale Sprachen und Komplexitätstheorie. Addison Wesley, Bonn 1994, 28-36.
- Hughes, A. L., & Palen, L. (2009). Twitter Adoption and Use in Mass Convergence and Emergency Events. In Proceedings of the 6th International ISCRAM Conference, Gothenburg, Sweden, May 2009.
- Jansen, B., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11), 2169-2188.
- Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. In WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (pp. 56-65).
- Jeners, N., & Mambrey, P. (2010). Zur Appropriation sozio-technischer Systeme am Beispiel des Microblogging Systems Twitter. In Multikonferenz Wirtschaftsinformatik 2010 (pp. 717-728).
- Oulasvirta, A., Lehtonen, E., Kurvinen, E., & Raento, M. (2009). Making the ordinary visible in microblogs. *Personal and Ubiquitous Computing*, (January 2009).

- Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In International Conference on Language Resources and Evaluation (pp. 1320-1326).
- Pandey, V., & Iyer, C. (2009). Sentiment Analysis of Microblogs. Student Project. Retrieved from <http://www.stanford.edu/class/cs229/proj2009/PandeyIyer.pdf>.
- Reinhardt, W., Ebner, M. & Beham, G. (2009). How people are using Twitter during conferences. In Proceedings of 5th EduMedia conference (pp. 145-156).
- Riemer, K., Richter, A. & Böhringer, M. (2010). Enterprise Microblogging. In: Business & Information Systems Engineering, 2(6).
- Shamma, D., Kennedy, L., & Churchill, E. (2009). Tweet the debates: Understanding community annotation of uncollected sources. In Proceedings of the first SIGMM workshop on Social media (p. 3–10).
- Vieweg, S., Hughes, A., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In Proceedings of the 28th international conference on Human factors in computing systems (p. 1079–1088).
- Wagner, C., & Strohmaier, M. (2010). The Wisdom in Tweetonomies: Acquiring Latent Conceptual Structures from Social Awareness Streams. In Proceedings of the WWW2010, April 26-30, Raleigh, North Carolina.
- Wisdom of Fans – Press Release, 12.07.2010, <http://blog.wisdom-of-fans.com/initial-results-press-release/>