



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	On the need of integrating social media channels and open source software repositories
Author(s)	Iqbal, Aftab; Decker, Stefan
Publication Date	2015
Publication Information	Iqbal, A; Decker S (2015) On the need of integrating social media channels and open source software repositories IEEE International Conference on Information Reuse and Integration (IRI) San Francisco, CA,
Publisher	IEEE
Link to publisher's version	<a href="http://dx.doi.org/10.1109/IRI.2015.66">http://dx.doi.org/10.1109/IRI.2015.66</a>
Item record	<a href="http://hdl.handle.net/10379/5443">http://hdl.handle.net/10379/5443</a>
DOI	<a href="http://dx.doi.org/10.1109/IRI.2015.66">http://dx.doi.org/10.1109/IRI.2015.66</a>

Downloaded 2024-04-19T14:38:38Z

Some rights reserved. For more information, please see the item record link above.



# On the Need of Integrating Social Media Channels and Open Source Software Repositories

Aftab Iqbal

*Insight Centre for Data Analytics  
National University of Ireland, Galway (NUIG)  
IDA Business Park, Lower Dangan, Galway, Ireland  
Email: aftab.iqbal@insight-centre.org*

Stefan Decker

*Insight Centre for Data Analytics  
National University of Ireland, Galway (NUIG)  
IDA Business Park, Lower Dangan, Galway, Ireland  
Email: stefan.decker@insight-centre.org*

**Abstract**—The growing interest in the usage of social media channels have attracted the open source software community to adopt an identity in order to disseminate project-related information to a wider audience. We foresee the need to integrate social media channels and open source software repositories in order to get an integrated view on the software project not only from the software development perspective but also from social perspective. Therefore, in this paper we study the usage of Twitter by software developers through harvesting their project-related activities on Twitter. In particular, we present the most commonly used hashtags by software developers and further investigate if project-related hashtags are the most frequent and commonly used hashtags by software developers. Based on our findings, we argue that relevant information from social media channels should be integrated with the open source software repositories in order to provide a homogeneous view on a software project.

**Keywords**—Linked Data; Semantic Web; Social Media; Twitter; Software Repositories; Data Integration;

## I. INTRODUCTION AND MOTIVATION

The growing interest in the usage of online social media channels such as Facebook, Twitter, MySpace, LinkedIn etc., have attracted millions of users. Users join together on social media channels to share information about their activities and opinions on various topics. Topics usually range from daily routine activities to current events, media news etc. With the popularity of social media channels, we have seen emerging growth in the usage of these channels by the open source software communities in order to disseminate project-related information to a wider audience. We have noticed activities of software developers and users on these channels discussing or sharing their thoughts and opinions about a particular release or issues relevant to software projects.

Open source software communities are often found to adopt an identity on social media channels (e.g., Apache Solr/Lucene<sup>1</sup> on Twitter, *MySQL*<sup>2</sup> on Facebook) in order to disseminate<sup>3</sup> project-related information (release announcements, major bug fixes etc.) to a wider audience or gather

feedback/questions posted by the users. Software developers contributing to open source projects also exists on social media channels. Quite often, they discuss, debate or share experiences<sup>4</sup> with others relevant to a software project using project related hashtags (e.g., #apache, #maven, #hadoop etc.). Hence, the discussions covering open source projects are not limited to dedicated forums, blogs or mailing lists, there also exists huge amount of information on the social media channels. Therefore, it is worth mentioning that the information related to open source projects and software developers are distributed on the Web in heterogeneous data islands [1], i.e., social media channels and software repositories of a project as shown in Figure 1.

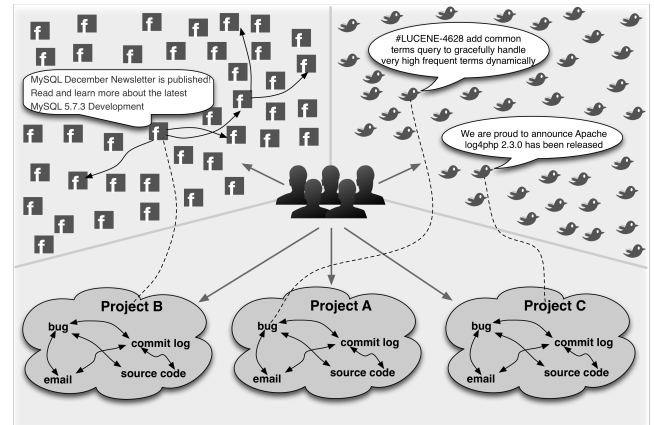


Figure 1. A conceptual diagram showing existence of software developers on social media channels and software projects.

Given the fact that the information relevant to software projects are shared by software developers and users on social media channels, one cannot ignore the importance of consuming this information. Therefore, there is a need to bridge the connection between software repositories and social media channels (i.e., interlinking project-related tweets/posts/hashtags, project IDs [2], software developer IDs [3] etc.). By enabling this connection, we will have an

<sup>1</sup><https://twitter.com/SolrLucene>

<sup>2</sup><https://www.facebook.com/mysql>

<sup>3</sup>for example, <https://twitter.com/olamy/status/305334578103582720>

<sup>4</sup>for example, <https://twitter.com/olamy/status/231031288734285824>

integrated view on the software project from the software development and social aspects that can be exploited to support certain use case scenarios:

- **Mining end-users response on the release or usage of a particular software project.** The variance and enormity of information that propagates through social media channels presents an opportunity to harness the data and build models on top to aggregate the opinions of the community related to a software project [4];
- **Investigating the popularity of a particular software project.** The popularity of a software project on social media channels [5] can be investigated by applying sentiment analysis on social messages [6], [7], [8];
- Analysis of the social behavior of software developers in different communication channels (i.e., social media channels and software project repositories) [9].

In the current scope of this paper, we do not take into account all Twitter users who are tweeting about some selected software projects but rather we focus our study to look into the tweets of only those Twitter users who are developers of those selected software projects. The reason behind this choice is to highlight the needs of integrating social media channels to software project repositories, which will be discussed later in detail through two different research questions.

Based on the underlying idea of integrating social media channels and software project repositories, the contribution of this work is manifolds. We have identified social media channels as a platform, which is used by the open source software communities to disseminate project-related information to a wider audience. Further, we highlight the needs of integrating software project repositories and social media channels in order to get an integrated view on the software project not only from the software development perspective but also from the social perspective. For integration purposes, we need to find out the activities of software developers on social media channels with respect to the software projects they are developing. Therefore, we investigate the usage of different hashtags by software developers on social media channels (Twitter, in particular) in order to discover if project-related hashtags are the most frequent and commonly used hashtags by software developers.

The rest of the paper is organized as follows: in Section II we discuss related work in the context of analyzing social media channels with respect to its usage by software developers. We investigate the usage of Twitter as a platform by software developers to disseminate project-related information in Section III. We report on our findings and further motivates the need of integrating social media channels to the software project repositories in Section IV. Finally, we conclude in Section V.

## II. RELATED WORK

Much research has been carried out on analyzing social media channels with different perspectives in mind; privacy issues [10], [11], influential users within a community [12], [13], [14], churns in online social communities [15], trending topic classifications [16], [17] etc. In the last few years, research has also started to emerge on analyzing the activities of software developers with respect to software projects on social media channels. Wang et al. [18] investigated how open source community uses social media channels through studying the usage of Twitter by software developers of Drupal open source content management system. In another study, Bougie et al. [19] found that software developers extensively leverages Twitter's capabilities for conversation and information sharing. Tian et al. [20] performed a preliminary study on what software developers microblogs about by analyzing the content of microblogs from Twitter and further categorized them. Similarly, Prasetyo et al. [21] automatically classified tweets regarding their relevance to software engineering. Leif et al. [22] discovered in their qualitative study that software developers use Twitter to keep themselves up-to-date with the fast changing development landscape, for learning and building relationships.

In particular, there are no research case studies available to date (to the best of our knowledge), which investigates if software developers use social media channels to talk about or promote software projects they are developing. The work presented in this paper differs from existing research in the sense that we investigate the most common and frequently used hashtags by software developers on social media channels (Twitter in particular) with respect to the software projects they are developing.

## III. ACTIVITY OF SOFTWARE DEVELOPERS ON SOCIAL MEDIA CHANNELS: PRELIMINARY STUDY

In this section, we investigate the usage of Twitter as a platform by software developers in order to find out if they promote or discuss software projects (on Twitter) that they are developing. This is required as a preliminary step towards integrating information coming through Twitter with the software repositories and to the software developers in particular. Therefore, we crawled tweets of software developers of some randomly selected *Apache* projects. The *Apache* projects and its associated software developers used in this study are shown in Table I.

For each *Apache* project under consideration, we considered only those software developers who have commit rights on the source control repository. For these software developers, we manually checked if they also exist on Twitter and using the Twitter account frequently. We found some software developers whose profile were set private by themselves and only authorized followers could view their tweets. We ignored such software developers in the Twitter data crawling and analysis phase. Table I shows

Apache Projects	Developers (SVN)	Developers (Twitter)
Apache Camel [23]	36	24
Apache Directory [24]	51	11
Apache Felix [25]	47	17
Apache Hadoop [26]	97	35
Apache Logging [27]	37	7
Apache Lucene [28]	51	18
Apache Maven [29]	40	10
Apache Mina [30]	28	9
Apache MyFaces [31]	82	16
Apache OfBiz [32]	25	11

Table I

SOFTWARE DEVELOPERS CONTRIBUTING TO APACHE PROJECTS AND ALSO USING TWITTER.

for each *Apache* project, the number of software developers who have commit rights to the source control repository and the software developers actually found on Twitter. Although, not all software developers developing the *Apache* projects (under consideration for this study) are found on Twitter but the results in Table I shows good evidence of the existence of software developers on Twitter, which will be sufficient to address some questions discussed later in this section. For each software developer, we obtained their tweets and profile information using Twitter API. Table II presents summary of our crawled Twitter data set by showing total number of software developers, total number of tweets and distinct hashtags used in the tweets by software developers.

Developers	Tweets	hashtags
158	186,709	20,395

Table II

TWITTER DATASET SUMMARY.

We crawled data from January 2009 till January 2013. The distribution of tweets per software developer is shown in Figure 2. The average number of tweets each software developer posted is 1181.70. The decline in the curve shows that certain software developers tweeted less on Twitter. One reason could be that certain software developers joined late on Twitter than other software developers. We queried our Twitter dataset in order to find out the oldest and recent joining date of software developers on Twitter. We found that the oldest joining date for a software developer was 2007-03-07 and the recent joining date for a software developer was 2012-04-11. The recent joining date of software developers may be the reason behind less tweets by the software developers on Twitter, however there may be cases where software developers exist on Twitter for a long time but have stopped tweeting shortly after joining Twitter or they may not be the active users of Twitter.

In the following, we will look deeply into couple of questions that will be discussed based on our crawled Twitter data set. These questions are designed specifically to investigate if software developers are discussing or promoting software projects on Twitter. This will provide us the basis to argue

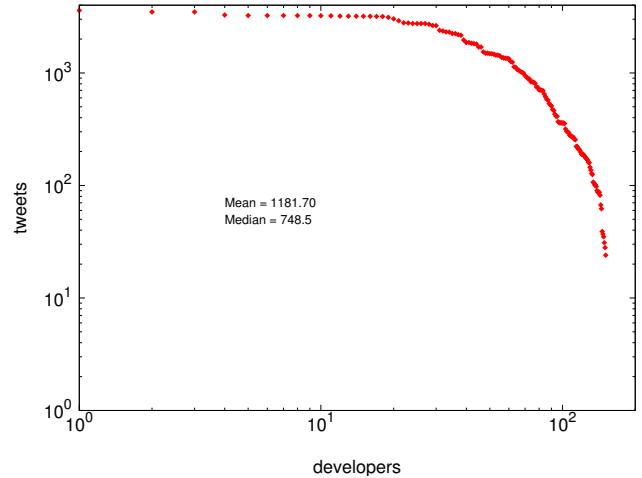


Figure 2. Distribution of the number of tweets sent per software developer of Apache projects.

the needs of interlinking software project related content, which is available on the social media channels to the information about software projects contained inside software repositories in order to address different use case scenarios, some of which are already highlighted in Section I.

**Q1:** *Are technology oriented hashtags most commonly used by software developers?*

The usage of hashtags are common among the users of Twitter. A hashtag is a character string preceded by a “#” sign. It is used to categorize the tweets. Hashtags are commonly used by users to specify category, topic or intended audience to a tweet. The usage of hashtags to mark tweets typically allow others to follow conversations regarding a particular topic. As mentioned by Evandro et al. in [33], it is used to not only add context and metadata to tweets but also for promotion and publicity purposes. Therefore, we are interested to investigate what kind of hashtags are mostly used by software developers and if they use project-related hashtags in particular, which will make it easier to integrate tweets relevant to a particular software project.

Based on our initial investigation, we found that software developers do use project-related hashtags to create communities of people (i.e., developers, contributors or users of a software project), who are interested in a particular software project. For example, “#hadoop” is a hashtag commonly used to associate tweets related to the *Apache Hadoop* project. Therefore, any one interested to know about what is under discussion on Twitter regarding *Apache Hadoop* project can search for tweets with a “#hadoop” hashtag. This shows that software project-related information do exist on social media channels and should be interlinked in order to provide a homogeneous overview on a particular software project.

We are typically interested to know about the most commonly used hashtags by software developers in our crawled Twitter data set. Therefore, we first retrieved a list of distinct hashtags. Then for each hashtag, we queried the number of distinct software developers who have used that particular hashtag in his/her tweets. We ignored the frequency of appearance/usage of a hashtag in this case and present only the top 15 most commonly used hashtags in Table III. The ranking of hashtags are done based on the number of software developers who have used that particular hashtag.

Hashtags	Developers
apache	147
java	139
opensource	106
maven	95
oracle	86
hadoop	80
fb	78
apachecon	74
git	73
android	71
devoxx	69
in	68
cloud	67
eclipse	62
scala	61

Table III  
MOST COMMONLY USED HASHTAGS BY SOFTWARE DEVELOPERS.

The result shows that tweeting about work related stuff is common among software developers. The usage of project-related hashtags is quite obvious because we have crawled data set of a specific community of users (i.e., *Apache* developers) from Twitter. The top 3 hashtags listed in Table III are obvious because of the following 3 reasons: (1) the projects under consideration are *Apache* projects; (2) the primary programming language used by the *Apache* projects under consideration is Java; and (3) *Apache* projects under consideration are open source. Moreover, most hashtags that are commonly used by software developers are technology-oriented (cf. Table III). Based on the results, we conclude that software developers do use hashtags to share work related information with others on Twitter. Therefore, we are required to integrate software project-related information, which are disseminated on social media channels to the respective software projects in order to enable a comprehensive overview on the software project from different perspectives (i.e., social and development perspectives).

**Q2:** *Do software developers tweet most frequently about software projects?*

We now investigate if software developers are tweeting specifically and most frequently about software projects they are developing. In other words, we are interested to know if software developers promotes or communicates about the software projects they are developing on Twitter using

project-related hashtags. As the main purpose behind this paper is to identify the needs of integrating social media channels (Twitter) to the open source software repositories, therefore, it is necessary to know how frequent software developers are communicating about software projects on social media channels. We believe that if software developers are discussing software projects frequently on social media channels then it will create communities of users around those software projects on the social media channels. With the establishment of such communities, we will have access to a large pool of information about those software projects from the social media channel.

Our approach towards computing most frequently hashtags are discussed in the following. Let's assume that we have a list of hashtags used by software developers,  $Hashtag = \{h_1, h_2, h_3, \dots, h_n\}$ . For each *Apache* project under consideration, we first retrieved a list of software developers,  $Dev = \{d_1, d_2, d_3, \dots, d_n\}$ . Then for each software developer, we extracted all occurrences of hashtags in his/her tweets,  $Hashtag_d = \{h_1, h_4, h_1, h_3, h_7, \dots, h_n\}$ ,  $Hashtag_d \subseteq Hashtag$ . Often, a software developer use same hashtag multiple times hence we summed up all occurrence of similar hashtags used by a software developer,  $Hashtag_{freq} = \{h_1(2), h_4(1), h_3(1), h_7(1), \dots, h_n(n)\}$  where  $Hashtag_{freq} \subseteq Hashtag_d \subseteq Hashtag$ . Later, we summed up all occurrences of same hashtags used by all software developers of an *Apache* project and ranked the results based on the most frequently used hashtags. We have selected only the top 2 most used hashtags for each *Apache* project under consideration and presents the results in Table IV.

$$\sum_{d \in Dev} Hashtag_{freq}$$

Apache Projects	Project Hashtag Used	1 <sup>st</sup> Most Used	2 <sup>nd</sup> Most Used
<i>Apache Camel</i>	✓	#camel	#apache
<i>Apache Directory</i>	✗	#apachecon	#apache
<i>Apache Felix</i>	✓	#osgi	#camel
<i>Apache Hadoop</i>	✓	#hadoop	#opendata
<i>Apache Logging</i>	✓	#rtw2012	#yam
<i>Apache Lucene</i>	✓	#lucene	#bbuzz
<i>Apache Maven</i>	✓	#maven	#apache
<i>Apache Mina</i>	✓	#fb	#netty
<i>Apache MyFaces</i>	✓	#myfaces	#primefaces
<i>Apache OfBiz</i>	✓	#rtw2012	#ofbiz

Table IV  
MOST FREQUENTLY USED HASHTAGS BY SOFTWARE DEVELOPERS OF APACHE PROJECTS.

In the results (cf. Table IV), we found that hashtag of *Apache OfBiz* project ranked 2<sup>nd</sup> but hashtags of *Apache Felix*, *Apache Logging* and *Apache Mina* projects ranked lower among the most frequently used hashtags by software

developers of these projects respectively. We found another interesting fact in the results of *Apache Felix* project where *Apache Camel* project’s hashtag appeared to be the 2<sup>nd</sup> most frequently used hashtag by the software developers of *Apache Felix* project. We investigated further to find the reasons behind it and discovered that 5 out of 17 software developers of *Apache Felix* project (cf. Table I) are also developers of *Apache Camel* project. This means that those 5 software developers are contributors to both *Apache* projects. Hence, we believe that those 5 software developers tweeted and communicated more about *Apache Camel* project than *Apache Felix* project on Twitter.

Based on the results shown in Table IV, we see that project-related hashtags are most frequently used by software developers of only 5 *Apache* projects out of 10 *Apache* projects considered for this study. Based on this observation, we can say that not all software developers tweet most frequently about software projects. However, this doesn’t mean that software developers haven’t used project-related hashtags for other *Apache* projects considered for this study. Therefore, we investigated if software developers have ever used project-related hashtag on Twitter and found that all *Apache* projects under consideration were mentioned by software developers on Twitter using their respective hashtags except the case of *Apache Directory* project. We found no usage of project-related hashtag by the software developers of *Apache Directory* project. In order to find out the reason behind it, we manually checked the general usage of *Apache Directory* hashtag (#ApacheDS) by the users on Twitter and found only 6 tweets (between January 2009 and January 2013), which contained that particular hashtag on Twitter. Further investigation revealed that none of the users who mentioned that specific hashtag was a core developer (i.e., the one who has source control commit rights) of *Apache Directory* project.

#### IV. DISCUSSION

In the previous section, we investigated the usage of Twitter as a platform by software developers to promote or communicate about the software projects they are developing. Based on the questions investigated, we discovered that software developers do communicate about software projects on Twitter using project-related hashtags. However, the frequency of using project-related hashtags vary among *Apache* projects under consideration. We believe that these findings can be made more concrete if applied to large number of Twitter data set relevant to open source software projects.

These findings are sufficient enough to argue the importance of integrating Twitter and software projects based on project-related hashtags or software developer IDs. However, we haven’t yet looked into how frequent software developers tweet about software projects. In particular, it will be interesting to know if software developers tweet about software

projects on daily, weekly or monthly basis.

In order to calculate the average number of days taken by a software developer to tweet about a software project (he/she is developing), we extracted all tweets mentioning a particular project-related hashtag such that,  $tweets = \{tweet_1, tweet_2, \dots, tweet_n\}$ . For each tweet, we extracted the timestamp value,  $date = \{date_1, date_2, date_3, \dots, date_n\}$  where  $date_1$  is the timestamp value of  $tweet_1$ ,  $date_2$  is the timestamp value of  $tweet_2$  and so on. Later, we computed the difference between two tweets containing project-related hashtag in terms of days, such that  $day = \{day_1, day_2, day_3, \dots, day_n\}$  where  $day_1 = date_2 - date_1$ ,  $day_2 = date_3 - date_2$  and so on. Finally, we computed the average number of days taken by software developers to post project-related tweet of *Apache* projects.

$$\sum \frac{day}{|day|}$$

Apache Projects	Tweets containing Hashtags	Average (#days)	Time
<i>Apache Camel</i>	475	2.68	
<i>Apache Directory</i>	0	0	
<i>Apache Felix</i>	22	62.27	
<i>Apache Hadoop</i>	1,578	0.89	
<i>Apache Logging</i>	20	44.30	
<i>Apache Lucene</i>	376	4.38	
<i>Apache Maven</i>	943	1.75	
<i>Apache Mina</i>	2	50	
<i>Apache MyFaces</i>	24	31.29	
<i>Apache OfBiz</i>	62	18.85	

Table V  
AVERAGE NUMBER OF DAYS TAKEN BY SOFTWARE DEVELOPERS TO POST A SOFTWARE PROJECT RELATED TWEET.

For each *Apache* project, we present the total number of tweets in which project-related hashtag was mentioned and the average number of days taken by a software developer to post about a software project using project-related hashtag in Table V. The time duration of each *Apache* project varies, which is due to the variation in the usage of project-related hashtags. We have ignored keyword-based mentioning of software projects (e.g., mentioning of “hadoop” in the twitter message text) while computing the average time, which may have reduced the average time of project-related hashtags usage in our study. Further, the results shows that software developers of only four *Apache* projects tweeted about the project in less than a week time using project-related hashtag. For other *Apache* projects, the average time taken by a software developer to post about a software project is quite high.

The results presented in Table V is based on a Twitter data set that contain tweets of only software developers of these *Apache* projects. We haven’t taken into account the usage of project-related hashtags by all users (i.e.,

users, bug reporters, patch contributors etc.) on Twitter, which we believe will provide more insights into the usage and promotion of a software project using project-related hashtags and may also contradict with the average time presented in Table V.

#### A. On The Need of Integrating Social Media Channels and Open Source Software Repositories

Our findings based on the questions reveal that there is relevant information about software projects exists on social media channels (i.e., Twitter). This information can be integrated (e.g., by interlinking project-related tweets/posts/hashtags, developer IDs [3], project IDs etc.) in order to get an integrated view on the software project. Once it is integrated, several use case scenarios can be addressed by exploiting the information. For example, we have shown elsewhere [9] that one could investigate the social behavior of software developers on a project mailing list and compare it with their social behavior on Twitter. This introduces a new dimension to the analysis of social dynamics of software developers by also taking into account social media channels. This would enable researchers to measure and compare the hierarchy and centralization of software developers in different communication channels (e.g., mailing lists, Twitter etc.) in contrast to previous studies where researchers have been using only mailing lists, bug repositories or discussion forums [34], [35]. Furthermore, integrating social messages/posts to project-related artifacts will open up new research challenges allowing to analyze the impact of end-users response on the success/failure of an open source project.

#### V. CONCLUSION

In this paper, we have identified social media channels as a platform, which is used by the open source software community to disseminate project-related information to a wider audience. Further, we highlighted the need to integrate project repositories and social media channels in order to get an integrated view on the software project. In order to support our proposition, we studied the usage of project-related hashtags by software developers on Twitter and found that software developers do use project-related hashtags to communicate about the software projects they are working on. Given that the ratio of usage of software project related hashtags vary for each *Apache* software project we considered in the study (cf. Table V) but this laid down the basis to integrate software projects with Twitter based on project-related hashtags. We further investigated if project-related hashtags are frequently used by software developers of *Apache* projects. We observed that not all software developers tweet most frequently about software projects but we found that software developers of *Apache* projects have used project-related hashtags in their tweets except the case of *Apache Directory* project. As our study consists of

limited number of software projects, it remains to be seen whether a large number of software developers demonstrate similar kind of behavior. Therefore, we can not say that all or majority of software developers frequently tweets about software projects using project-related hashtags.

While more research is needed to better understand these and other related questions, our study shows the existence of software project related information on Twitter. Therefore, this information can be integrated with software repositories hosting that particular software project in order to support certain use case scenarios that are mentioned in the beginning of this section.

#### ACKNOWLEDGMENT

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289.

#### REFERENCES

- [1] A. Iqbal and M. Hausenblas, "Integrating developer-related information across open source repositories," in *IEEE 13th International Conference on Information Reuse and Integration (IRI)*, 2012, 2012.
- [2] M. Conklin, "Project entity matching across floss repositories," in *3rd International Conference on Open Source Systems*, Springer. Limerick, IE: Springer, 2007, pp. 45–57.
- [3] A. Iqbal and M. Hausenblas, "Interlinking developer identities within and across open source projects: The linked data approach," *ISRN Software Engineering*, 2013.
- [4] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, ser. WI-IAT '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 492–499. [Online]. Available: <http://dx.doi.org/10.1109/WI-IAT.2010.63>
- [5] L. de Vries, S. Gensler, and P. S. Leeflang, "Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing," *Journal of Interactive Marketing*, vol. 26, no. 2, pp. 83 – 91, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1094996812000060>
- [6] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1-2, pp. 1–135, Jan. 2008. [Online]. Available: <http://dx.doi.org/10.1561/1500000011>
- [7] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in *Proceedings of the Workshop on Languages in Social Media*, ser. LSM '11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 30–38. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2021109.2021114>

- [8] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '04. New York, NY, USA: ACM, 2004, pp. 168–177. [Online]. Available: <http://doi.acm.org/10.1145/1014052.1014073>
- [9] A. Iqbal, M. Karnstedt, and M. Hausenblas, "Analyzing Social Behavior of Software Developers Across Different Communication Channels," in *25th International Conference on Software Engineering and Knowledge Engineering (SEKE 2013)*, Boston, USA, 2013.
- [10] L. Fang and K. LeFevre, "Privacy wizards for social networking sites," in *Proceedings of the 19th international conference on World wide web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 351–360. [Online]. Available: <http://doi.acm.org/10.1145/1772690.1772727>
- [11] D. Rosenblum, "What anyone can know: The privacy risks of social networking sites," *Security Privacy, IEEE*, vol. 5, no. 3, pp. 40–49, 2007.
- [12] M. Trusov, A. V. Bodapati, and R. E. Bucklin, "Determining influential users in internet social networks," *Journal of Marketing Research*, vol. XLVII, pp. 643–658, 2010.
- [13] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, "Identifying the influential bloggers in a community," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*, ser. WSDM '08. New York, NY, USA: ACM, 2008, pp. 207–218. [Online]. Available: <http://doi.acm.org/10.1145/1341531.1341559>
- [14] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*, ser. WSDM '10. New York, NY, USA: ACM, 2010, pp. 261–270. [Online]. Available: <http://doi.acm.org/10.1145/1718487.1718520>
- [15] M. Karnstedt, T. Hennessy, J. Chan, and C. Hayes, "Churn in social networks: A discussion boards case study," in *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, 2010, pp. 233–240.
- [16] H. Becker, M. Naaman, and L. Gravano, "Beyond trending topics: Real-world event identification on twitter," in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
- [17] K. Lee, D. Palsetia, R. Narayanan, M. M. A. Patwary, A. Agrawal, and A. Choudhary, "Twitter trending topic classification," in *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*, ser. ICDMW '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 251–258. [Online]. Available: <http://dx.doi.org/10.1109/ICDMW.2011.171>
- [18] X. Wang, I. Kuzmickaja, K.-J. Stol, P. Abrahamsson, and B. Fitzgerald, "Microblogging in open source software development: The case of drupal using twitter," *IEEE Software*, vol. 99, no. PrePrints, p. 1, 2013.
- [19] G. Bougie, J. Starke, M.-A. Storey, and D. M. German, "Towards understanding twitter use in software engineering: Preliminary findings, ongoing challenges and future questions," in *Proceedings of the 2Nd International Workshop on Web 2.0 for Software Engineering*, ser. Web2SE '11. New York, NY, USA: ACM, 2011, pp. 31–36. [Online]. Available: <http://doi.acm.org/10.1145/1984701.1984707>
- [20] Y. Tian, P. Achananuparp, I. Lubis, D. Lo, and E.-P. Lim, "What does software engineering community microblog about?" in *9th IEEE Working Conference on Mining Software Repositories (MSR)*, 2012, pp. 247–250.
- [21] P. Prasetyo, D. Lo, P. Achananuparp, Y. Tian, and E.-P. Lim, "Automatic classification of software related microblogs," in *28th IEEE International Conference on Software Maintenance (ICSM)*, 2012, pp. 596–599.
- [22] L. Singer, F. F. Filho, and M.-A. Storey, "Software Engineering at the Speed of Light: How Developers Stay Current Using Twitter," <http://leif.me/papers/Software-Developers-Twitter-DCS-350-IR.pdf>, University of Victoria, Technical Report, 2013.
- [23] <http://camel.apache.org/>.
- [24] <http://directory.apache.org/>.
- [25] <http://felix.apache.org/>.
- [26] <http://hadoop.apache.org/>.
- [27] <http://logging.apache.org/>.
- [28] <http://lucene.apache.org/>.
- [29] <http://maven.apache.org/>.
- [30] <http://mina.apache.org/>.
- [31] <http://myfaces.apache.org/>.
- [32] <http://ofbiz.apache.org/>.
- [33] E. Cunha, G. Magno, G. Comarela, V. Almeida, M. A. Gonçalves, and F. Benevenuto, "Analyzing the dynamic evolution of hashtags on twitter: a language-based approach," in *Proceedings of the Workshop on Languages in Social Media*, ser. LSM '11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 58–65. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2021109.2021117>
- [34] A. Wiggins, J. Howison, and K. Crowston, "Social dynamics of floss team communication across channels," in *Fourth International Conference on Open Source Software*, 2008.
- [35] K. Crowston, K. Crowston, and J. Howison, "Hierarchy and centralization in free and open source software team communications," *Knowledge Technology Policy*, vol. 18, pp. 65–85, 2005.