Improving Activity in Communities of Practice through Software Release Management

Kevin van Ingen
Utrecht University
k.l.vaningen@students.uu.nl

Job van Ommen
Utrecht University
j.vanommen@students.uu.nl

Slinger Jansen
Utrecht University
slinger.jansen@uu.nl

ABSTRACT

Keystone players, the company that occupies the crucial hubs, need to nurture its inhabitants to keep the ecosystem active. Many modern ecosystems have communities of practice where knowledge is transferred by collaborative problem solving, sharing of ideas, software components or configurations. In recent years a widely spread medium for communities of practice is the use of online discussion boards. This research proposes a method to analyze the relationship between software releases and activity in a community. This paper explains how software ecosystem Keystone companies can use software product release management to cultivate communities of practice using an illustrative case study. In this case study a comparison is made between two communities in the Android ecosystem. The results show peaks in community activity coinciding with the software releases. The release not only revitalizes the activity of the developers but the heterogeneous community in its entirety.

Categories and Subject Descriptors

K.1 [Computing Milieux]: The Computing Industry – markets, standards, suppliers

General Terms

Management, Measurement, Human Factors, Theory, Verification

Keywords

Software ecosystems, communities of practice, knowledge management.

1. INTRODUCTION

Software vendors have become networked and integrated with suppliers, resellers, and even customers. This phenomenon has been the enabler for the research field of software ecosystems. A software ecosystem is "a set of actors functioning as a unit and interacting with a shared market for software and services, together with the relationships among them" [5]. Within these ecosystems we can recognize different roles such as Keystone or Niche. The Keystone role within a software ecosystem belongs to the entity that behaves in such a way that its effects propagate through the entire system. In each case, the Keystones occupy richly connected hubs that provide the foundation for creating many niches, regulate connections among ecosystem members, and work to increase diversity and productivity [4]. The Niche role within a software ecosystem belongs to the entity that acts to develop or enhance specialized capabilities that differentiate it from other firms in the network, leveraging resources from the network while occupying only a narrow part of the network itself [4]. Among the ecosystem actors are developers who use the platform of the Keystone player to develop their own products. Because of their contribution to the ecosystem, developers can be considered Niche players. Actors in an ecosystem interact with each other by sharing information about software solutions, and by doing so become part of a community of practice [7]. These communities are not necessarily limited to developers. Ecosystem enthusiasts, customers, or even hardware vendors can partake in these communities of practice.

The goal of this study is to analyze the dynamics of an online community in relationship with software releases from a software ecosystem perspective. By examining the way Niche players are affected by actions of the Keystone this research contributes to the body of knowledge on software ecosystems. The results of this research will provide Keystone players in a software ecosystem with new tools to influence the community. Providing insight on the effects that different types of platform releases have on online communities of practice is a contribution to the body on knowledge of software product management. The main research question addressed in this survey is: how do platform releases impact activity in an online community of practice?

This research question is answered by these two sub questions:

1. Is there an effect on communities of practice during a product release?
2. Do the effects on a developers community correlate with the effects on a community with a broader audience?

The remainder of the paper is structured as follows. Section 2 describes the research method of this case study. Section 3 describes the theoretical background for software ecosystems, online communities of practice, and the relationship between communities and software ecosystems. Section 4 presents a description of the results of the case study and an analysis of the data. Section 5 provides the answers to the research question and sub questions. Section 6 discusses this research shortcomings and future research directions.

2. RESEARCH METHOD

This research identifies the relationship between software releases as an event and activity in an online community of practice. A case study of the Android ecosystem with statistical analysis will provide insight in recurring patterns.

MEDES'11, November 21-24, 2011, San Francisco, USA.

The research field of software ecosystems is a relatively young field of research. This research needs to be repeated and validated in other industry domains. The research method used for this research is explicit but can also be generalized for use in other domains.

The online community of practice subject to this research needs to offer information about the members and their activities in a significant time range so that we have several releases and release-intervals to measure. This research follows the following steps.

1. Conduct preliminary literature research to discover hypothesis, measurement, and case selection criteria.
2. Construct a list of events in the Android community that potentially influences community activity.
3. Select one or multiple case study subjects and conduct a preliminary assessment on the community dataset.
4. Choose largest possible case and construct a scraping tool to obtain community data.
5. Data gathering using the custom-built scraper and use Extract, transform and load (ETL) tooling to transform the cleaned dataset into a reporting tool.
6. Construct a timeline of community activity in the community.
7. Construct a timeline of Google trend analysis for the Android platform.
8. Transpose the list of Android related events on the timelines.
9. Visually analyze the relation of Android related events on community activity and Google search trends.
10. Explain the effect of Android related events on the community in the results.

The case study is validated through the use of multiple data sources for the analysis.

2.1 Case study criteria

The Android system is an open source software stack for mobile devices that includes an operating system with eight major releases and dozens of small releases in the past four years. This information can be obtained through the Android Developers website [1].

Data about software releases and community members and activity are analyzed in this research and need to be obtained. Criteria for a community are described below.

- The community has to be active from 2008 until now and have at least 10,000 post per year (about 30 per day) to identify trends.
- The community has to display meta-data about threads (date, replies, and views).
- The community has to display meta-data about members (post started, join date).
- The community has to have post and member data publicly available on the internet in a sort of structured form in order to be suitable to build a scraper for.
- The community preferably has an URL structure which enables an external party to do a preliminary data filtering on inactive users.

2.2 Data mining and analysis

The information from the selected community is obtained through a custom-built scraping tool. This tool is similar to a search engine's bots or spiders that crawl the web for new links. When given a base URL it filters out relevant information and URLs on a webpage using text-based pattern matching in the source code of the webpage. New URLs on their turn are checked for relevant information, and recursively, new URLs. Relevant information about posts is filtered out and stored in a temporal dataset. For analysis purposes ETL is used to structure information in a generalized format needed for this study. The data is saved in a database and is structured as described in Table 1.

Table 1. Data structure of mined Androidcommunity.com messages.

Field	MySQL type	Description
threadid	INT	Thread identifier
postid	INT	Post identifier
member	VarChar(255)	User who posted a contribution to a thread
date	Date	Date stamp of the thread

The data obtained from the communities will be analyzed and standardized into a comparable format. The standardization is done using an index scale for activity where for every time-interval there will be an index number relative to the maximum amount of activity which will be indexed at 100%. The indexing of activity is a necessary step to ensure the comparability of data from different sources.

For the data on platform releases we use the roadmap and version history provided by the platform owner [1]. Because in mobile platforms a software release for the end user is always preceded by a corresponding software development kit (SDK) release we use the release dates of the latter. The availability of the end user version differs with every carrier and model so the SDK release is more suitable for providing a single point in time which may be correlated with a change in the activity of communities of practice. Especially because the communities of practice contain developers who depend on the SDK for their role as Niche players, in this paper the release date of SDK versions are considered as the starting point of the actual platform release.

3. THEORETICAL BACKGROUND

In this chapter we use related literature to position this research. First, we describe research in the field of software ecosystems. Secondly, we describe research in the field of online communities of practice.

The role of a Keystone player within an ecosystem is to promote the overall health of the ecosystem [4]. By definition the platform supplier is a Keystone firm [3]. Because in this review we are looking at the effect on platform releases on the community of practice we hope to contribute to the understanding of how a Keystone player can revitalize the community and thus the ecosystem. For this research it is not relevant who plays the role of Keystone as long as it can be identified, so we will not get into the depths of exploring who exactly owns the Android platform and is thus the Keystone player. It is sufficient to recognize that a platform is released by a Keystone player and that the rest of the ecosystem may utilize this new release.

One of the key success factors for companies employing a Keystone strategy, like Microsoft and eBay, did was appealing to a community of people [4]. The value of a community is also

evident in the field of product innovation and development and has successfully been leveraged by companies, such as for example Ducati and Eli Lilly [6]. A community perspective in the software ecosystem includes external developers, domain experts and users. The role of this community is vital to the health of the software ecosystem and thus for the Keystone player whose health depends on that of the ecosystem [4]. In the next section we will elaborate on (online) communities and their dynamics.

Actors within the community may have different roles. Not every person in the community will be a "Niche player", a part will have more passive roles such as end-users who 'just' like to be up-to-date. Similar to Open Source Software (OSS) Communities the roles of community participants are not distinct: all users are potential developers and thus potential Niche players [8]. Any modification, improvement, and extension made to an OSS system also influence the OSS community. The dynamics of the OSS and the community associated with it can affect the level of participation a member has within the community. Because of this, evolution of the community's topic is important for keeping the community sustainable [8].

One of the success factors for a platform like Android is to constantly extend the set of features to remain attractive for the developer community. It is also a pitfall because functionality created by the developer community can be incorporated so that it becomes a commodity, thus alienating the developers of that particular functionality [5]. Especially on the internet the importance of a community of 3rd party application developers has become central to the strategy of many companies.

4. RESULTS

In this chapter we present the results from the case study of the Android community and the effects of Android releases on the activity in the community. This chapter starts with the descriptive results. Afterwards the research results are presented and explained.

At the time of this research the open source Android system seems to have multiple active communities. Different Android communities were selected and assessed based on research criteria. The community androidcommunity.com was found capable of supplying proper information for this research. The forum originated as a news and discussion platform from the creators from the news platform slashgear.com. The forum androidcommunity.com targets people that have interest in Android mobile hardware and software for tablets and phones. Its users are end users that have interest in news from manufacturers about new phones and operating system updates but there are also developers for the Android apps.

Google Groups is a second community and was considered as a good complementary source of information. It is a discussion platform primarily focused on Android developers. It is integrated in the Android development documentation because of its linkage to Google. It is important to keep in mind that the user population of this community is more homogeneous because it is targeted at developers.

Different communities will be affected by new product releases. A way to access information from the communities is by web search engine. Google being the leading search engine at this point in time offers statistics on search queries performed. The meta-data of search trends are used to validate findings in the communities.

This study covers the results of two Android communities. Both communities offer similar meta-data suitable for this research. The communities are both active since the beginning of Android development in 2008. Descriptive results of the sizes of both communities are depicted in Table 2.

Table 2. Descriptive statistics of the Google Groups community and Androidcommunity.com

Metric	Google Groups	Androidcommunity
Amount of members	49.245	22.369
Amount of posts	156.241	313.874
Average post/month	3.633	7.300

A list of platform releases was gathered using Android platform release notes. The platform release notes also classify the impact in of the release as a major or minor release. An overview of platform releases is depicted in Table 3. The active versions column provides a current overview of the different Android versions that have accessed the Android Market in a 14 day period [2].

Table 3. Android platform releases

Date	Platform version (SDK)	Classification	Active versions
2008-09	Android 1.0	Major	Not available
2009-02	Android 1.1	Major	Not available
2009-05	Android 1.5	Major	1.9%
2009-10	Android 1.6	Minor	2.5%
2009-11	Android 2.0	Major	Not available
2010-01	Android 2.1	Minor	21.2%
2010-06	Android 2.2	Major	64.6%
2010-12	Android 2.3	Major	9.2%
2011-02	Android 3.0	Major	0.3%

In this research the focus was primarily on activity expressed by the amount of posts in the community. For interpretation purposes the activity of both the communities was indexed on a percentage scale. The index numbers of both the communities are combined into a new index called the average community index. To ensure validity a correlation between the two datasets is constructed. The total dataset has a low correlation of 0,38. Close examination of the dataset revealed a significant decrease in the activity in Androidcommunity.com. A correlation of the two datasets for the years 2008, 2009, and 2010 results in a correlation of 0,76. The strong correlation is interesting considering the differences between the audiences of both communities. The correlation indicates a relatively similar response on Android SDK releases in the developer community and a community with a much broader interest.

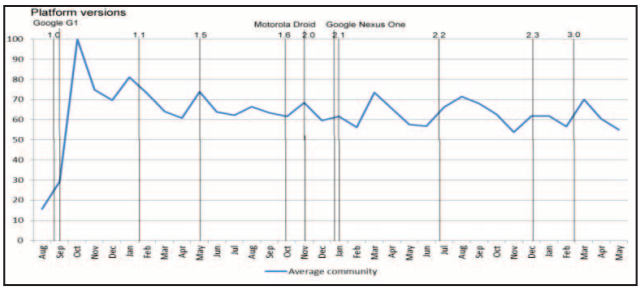

Figure 1. The indexed amount of activity for each month and the Android platform releases over the years 2008, 2009, 2010.

Visual analysis of the graph indicates that a relationship between releases and activity in the community is present. For most releases there is an increase in activity in the community. The details of the release can affect different communities in different ways. Some releases primarily affect application developers, others enable new hardware opportunities like tablets and multi-core processing, and others affect both. In order to illustrate this we also show the releases of a few notable phones. As the graphs show the releases of phones conincide with the release of important updates. The graph in Figure 1 primarly shows peaks that occur in both datasets. The classification of the impact of the release seems to have no significant effect on the activity.

A significant result in this graph is the difference of length of peaks. Peaks range from one or two months to 4-6 months. An observation is that high peaks (high amplitude) typically have a short duration.

The findings obtained by both communities are compared to Google Trend analysis. Google Trends provides data about popular search queries. In this research keywords are selected and a weekly search popularity ranging from august 2008 until may 2011 was combined in an average index. The keywords are selected in a way that they represent popular topics in both the Google Group community and androidcommunity.com. To ensure representativeness of the queries a correlation matrix was constructed. The correlation matrix described in Table 4 show higly correlated search trends.

Table 4. Correlation matrix of Google Trend analysis

	Android	Android SDK	Android Phone	Android Release
Android	1	-	-	-
Android SDK	0,97	1	-	-
Android Phone	0,95	0,96	1	-
Android Release	0,85	0,88	0,90	1

The popularity of these different search queries is plotted in a graph combined with platform releases depicted in Figure 2. The orange line in the graph is the average of the different queries combined. Visual analysis of the graph indicates that a relationship between releases and Google searches is present. For most releases there is an increase in searches in the Google search engine.

Both of the Android communities and the Google trend analysis indicate the existence of a relationship between activity in a community and platform releases. There is no visable relationship between the classification of the impact of the release on the response of the community. It is important to mention that peaks in activity or searchtrends can not be fully understood by new releases of the Android platform. There seem to be two distinctive

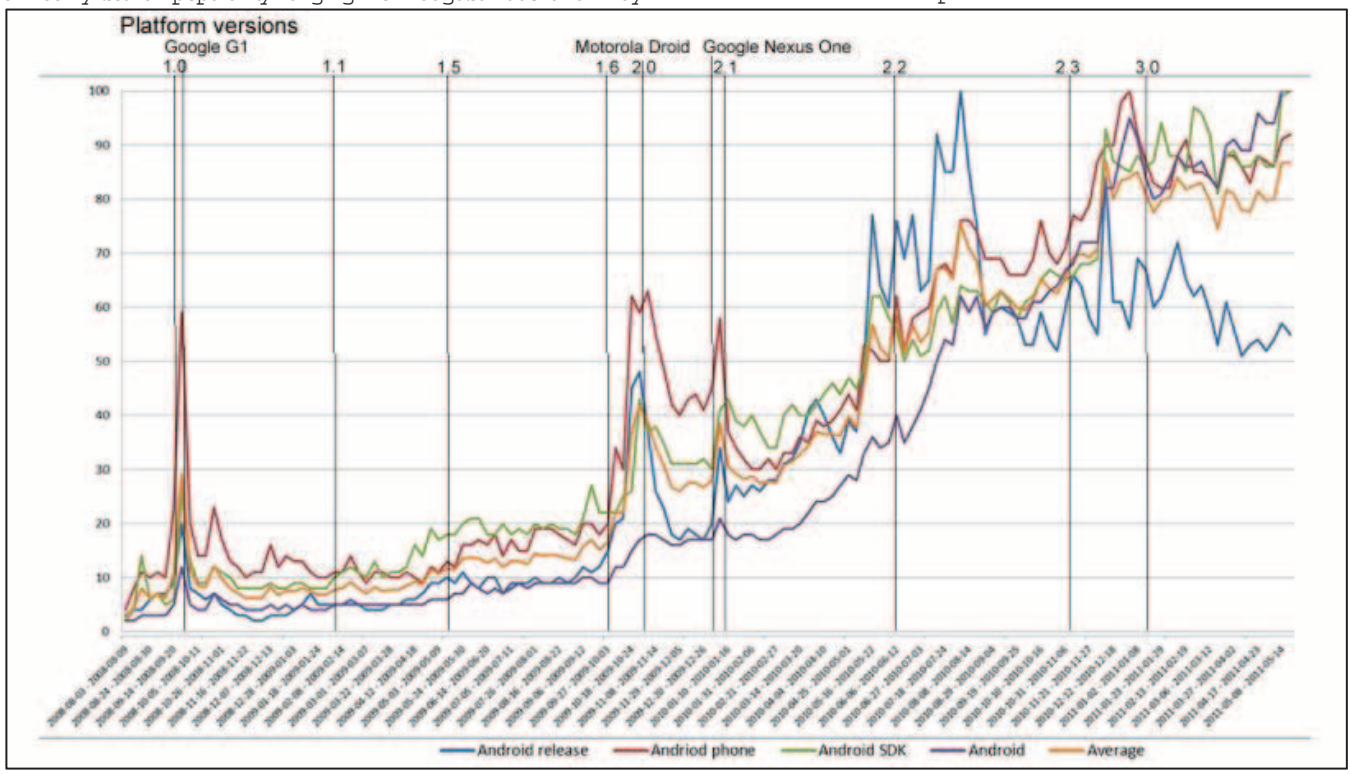

Figure. 2. Google trend indexes combined with Android platform releases

patterns in the activity peaks. There are short living bursts of activity with high peaks that end within one month and peaks with a smaller amplitude that have a longer duration of two or three months.

5. DISCUSSION AND FUTURE RESEARCH

Future research must investigate how this factor can be leveraged by the Keystone player. An important weakness of this research is to ensure that a certain event is correlated or even causes an effect in the online community of practice. Indeed, because before the actual platform release, the community is already aware of this pending event, thus activity may rise in anticipation. Even worse, activity may also rise due to rumors of a pending event. Other events, such as conferences or hardware releases, may also affect the community's activity and coincide with platform releases. This problem has also been visualized in the figures but has not been explored in this research. Such ecosystem events need to be researched so that we can establish a full set of measures affecting communities of practice within the ecosystem. In this study we were unable to differentiate between events on a detailed level. By knowing the details of an event it may be possible to perform an improved analysis which could explain the differences in corresponding activity. A study of a classification of the nature of product releases in the context of software ecosystems could provide a better explanation of community response. Such a classification system may also be used to formalize the meaning of notions such as major and minor release. Another weakness is the data we used to determine activity. In this research we use posts per timeframe to determine the community's activity. Forum statistics show other important factors like "users online" on a given date. Unfortunately we were unable to obtain such factors and thus the measure used for activity may not be as accurate as possible.

6. CONCLUSION

This paper addresses the research question: "How do platform releases impact activity in an online community of practice?" The results of this research show peaks coinciding or following platform releases in both the communities and in the Google trend analysis. It is therefore safe to say that platform releases cause higher activity in online communities of practice and increased interest for the entire ecosystem.

The results show a positive effect on the activity after each platform release. The overall effect on the communities in the case study is not enough to keep them from degrading. Of all the factors affecting the activity in communities of practice the topic evolution part – in this case a platform release - has now been established. This answers the first sub question: "Is there an effect on communities of practice during a product release?"

The activity in a community of practice with a homogenous population of developers corresponds to that of the community with a broader audience.

The correlation between these two communities proves platform releases have an effect on the entire ecosystem and not just on Niche players. In addition to the rise in activity in communities of practice, the rise in global interest for the ecosystem was clearly visualized by the Google search trends. This answers the second sub question: "do the effects on a developers community correlate with the effects on a community with a broader audience?"

A bi-product of this research is the creation of a research method which is explicitly mentioned in this paper and generalized to the level that it can be used in other research areas.

The results of this research show that Keystone players can utilize platform releases to improve the overall health of the ecosystem. It also shows that an important factor in online community research is the evolutionary aspect of the community's topic, a factor which has hitherto largely been ignored.

7. REFERENCES

1 Android SDK. Android Developers. [Online] Android, 2 June 2011. [Cited on: 14 June 2011.] http://developer.android.com/sdk/index.html.

2 Platform Versions. Android Developers. [Online] Android, 1 June 2011. [Cited on: 03 Juli 2011.] http://developer.android.com/resources/dashboard/platform-versions.html.

3 Bosch, Jan. From software product lines to software ecosystems. In Proceedings of the 13th International Conference on Software Product Lines (Pittsburgh 2009), Carnegie Mellon University Pittsburgh, 111-119.

4 Iansiti, M and Levien, R. The Keystone Advantage "What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability. Harvard Business School Press, Boston, 2004.

5 S. Jansen, A. Finkelstein, and S. Brinkkemper. Business network management as a survival strategy: A tale of two software ecosystems. (2009), In Proceedings of the First Workshop on Software Ecosystems., 187-190.

6 Sawhney, Mohanbir, Verona, Ginamario, and Prandelli, Emanuela. Collaborating to create: The internet as a platform for customer engagement in product innovation. Journal of interactive marketing (2005).

7 Wenger, E.C., & Snyder, W.M. Communities of practice: The organizational frontier. Harvard Business Review, 78(1), 139-145. Harvard Business School Press, Boston, 2000.

8 Ye, Yunewn and Kishida, Kouichi. Toward an Understanding of the Motivation of Open Source Software Developers. In Proceedings of 25th International Conference on Software Engineering (2003), 419-429.