Towards Improving the Quality of Mobile App Reviews

Mubasher Khalid
Dept. of Computer Science, COMSATS Institute of Information Technology, Sahiwal, 57000, Pakistan.
E-mail: mkhalid125@gmail.com

Muhammad Asif and Usman Shehzaib
Dept. of Computer Science, COMSATS Institute of Information Technology, Sahiwal, 57000, Pakistan.
E-mail: {drasif, usmanshehzaib}@ciitsahiwal.edu.pk

Abstract—Mobile app reviews are gaining importance as a crowd source to improve the quality of mobile apps. Mobile app review systems are providing a platform for users to share their experiences and to support in decision making for a certain app. Developers, on the other side, are utilizing the review system to get real-life user experience as a source of improving their apps. This paper has analyzed existing review system and proposed few recommendations for the current review system to improve the quality of app reviews. The proposed review system can help for collection and analysis of user reviews to make it more meaningful with less intensive data mining techniques. The proposed system can help the end users to get an overview of mobile apps. The recommendations in this paper are derived from the existing literature related to app reviews and will help to improve the current review systems for better app reviews from users as well as developers perspective.

Index Terms—Mobile Apps, App store, User Rating, Review System, Crowdsource.

I. INTRODUCTION

Over the past few years, the mobile ecosystem has grown at a remarkable rate. Apple triggers the revolution of mobile apps with the start of the App Store in 2008. According to Apple press info on January 8, 2015, App Store developers have gained all together $25 billion from the sale of apps and games [1]. Later, Google play store and Windows app store also joined the mobile app market. These App stores are providing a market place for mobile apps users and developers. These app stores are providing users a platform where they can choose different apps of their choices. Moreover, these platforms are providing a distribution channel for the developers to distribute and promote their apps. One of the most interesting aspects of the mobile ecosystem is the emphasis on user feedback. Most app stores allow the users to give public reviews to an app. Users can give a star rating to an app (all of which are aggregated and displayed at a version and app level basis), and provide a review-comment which describes their rating. If users have a bad experience with an app, they give it a low-rating. Thus star ratings of apps in the app markets (e.g., Google Play, iTunes App Store) give an insight into the user's opinion of the app.

Product reviews are already popular and playing a key role in driving the market. More than 60% of online customers read online reviews before buying a product, reviews being 12 times more trusted than the descriptions that are provided for the products by the sellers. In the past few years e-commerce has become extremely popular, and quantity of reviews that a product receives tends to grow exponentially. In the UK alone, almost half of the population (47%) has reviewed a product online, leading to the situation in which one single product may have thousands of reviews associated with it [16].

It is essential to recognize the key characteristics of mobile apps to fulfill the needs of the users [22]. App stores are also providing a way to the users to share their experience with apps in the form of app reviews. Users can provide feedback in the form of ratings and reviews about a certain app they used or downloaded. This type of feedback allows users to share their opinions and experiences about an app which can motivate or discourage other users to download an app. Users usually express their experience of an app in the form of star ratings and free text reviews that is public. These app distribution channels are providing a way for user-driven market where users can communicate their needs and developers can get valuable information in the form of user feedback.

The crowd source opinion in the form of mobile app reviews can play a vital role in driving the mobile app market. The analysis of mobile app reviews can reveal different aspects which can be useful for the users as well as overall mobile app market. In this regard, we have analyzed the literature related to mobile app reviews (See Table 1) and identified various aspects of the app reviews studied in by different researchers. Generally, reviews can be either positive or negative and can be helpful for driving mobile app market, app quality assurance, customer satisfaction and identification of innovative ideas. If app reviews are positive with better rating, the app may have more downloads. And if reviews are negative, then can help in bug identification and resolution. There is a third category of reviews which can

be neither negative nor positive reviews but may contain innovative ideas, user requirements and feature requests. In app reviews, users can communicate and describe their requirements depending on their real-life experience with the app. Negative reviews and low ratings can have significant effect on popularity of the app which in turn can affect the revenue generation for the developer. Therefore, there is a great need to understand the user reviews to flourish in the current high competitive mobile apps industry [2].

App stores are providing app distribution channels which are different from traditional software development and distribution mechanisms [3]. Developers can get a quick and productive feedback from users on time. This timely feedback has become crucial for developers to remove bugs, to make better user experience agility and to introduce new features [4].

User reviews about products are already being studied and utilized in an effective way. For example, cui et al. [5] have studied the effect of online product reviews and proposed a e-model WOM (Word of Mouth) for Amazon and argue that online reviews are stronger and have impact on launching new products and devising new e-marketing strategies.

Although users are providing a valuable feedback but users are not formal when they write a review about an app. The content of a user reviews usually consists of a mixture of different formal and slang expression with abbreviations and spelling fluctuations. Generally, user reviews are not written following the formal rules. So, the exact sentiment expression can't be extracted from the user reviews with some predefined patterns [6]. In current app review system, there are two major hurdles for app developers to obtain valuable information. First, the share of "informative" user reviews is comparatively less. Secondly, for some famous apps, the size of user reviews is so huge to do manual inspection on everyone. For example, a single app of Facebook on Google's Play Store gets 2000 new reviews per day, which makes it difficult to do manual examination [4]. Different data mining technique are in use to sort out user reviews. But this sorting technique required a lot of effort to classify data. And even after that you are not sure that reviews are properly classified or not.

The remainder of the paper is organized as follows: Section 2 describes the literature study done to identify the issues related to mobile app reviews in general and quality of app reviews in particular. In section 3, we have analyzed the current review system (Google play store) and described few limitations related to app reviews. Section 4 describes the recommendations to enhance the current app review system from user as well as developer's perspective. Section 5 concludes the research work and points out future work.

II. Literature Study

This section discusses literature related to mobile app reviews. We have focused mainly to identify the issues and limitation addressed in different studies related to mobile app reviews. Different app review studies have utilized users' feedback in a variety of ways. An overview of the purposes of these studies is presented in Table 1.

Understanding the relationship between the app market and user's experiences can help developers to better understand their app success and what attributes influence users' selection decisions [7]. In this way, developers can ensure the best possible end-user experience and can help to identify app attributes and opinions that are important to users. The size and average length of app reviews is a concern and have impact on the quality of reviews. Vasa et al. [2] have studied the effect of size of a user review and an average length. Authors also studied if star rating has affect on the length of a review and category of an app can influence the length of a review. Authors used their own script to parse and download the reviews. They restricted their focus to the top 400 famous apps only, based on their ranking at the time of compilation. They used the review size as a fundamental measure of appointment, in that it is an estimate of the concern the app has on the user.

Effective analytical tools for mobile app reviews have certain limitations and are not able to exploit user feedback properly. Chen et al. [4] identified problem that app developers spend considerable effort on collecting and exploiting user feedback to improve user satisfaction, but suffer from the absence of effective user review analytics tools. Authors grouped informative reviews using topic modeling, prioritizing informative reviews in ranking and via an intuitive visualization approach. Finkelstein et al. [8] also analyzed correlation between the price, popularity and ratings accorded to apps by their users. According to Hoon et al. [9], it is useful for developers to recognize what app users communicate in their reviews, which allow them to further gather user hope for quality. So here the main objective is what actual information, other than rating, developers can take out from these online reviews that may be of significance to them. Authors studied the most frequently word used for reviews and the vocabulary size used by the users of different languages.

Fu et al. [10] introduces techniques to mine reviews and discussed how these can give advantage different stake holders. End-users can get summary to decide the apps with the finest user-experience, without analyzing each and every comment. App developers can use these summaries to know why end-users like or dislike their apps, as well as evaluate apps critically, so that they can enhance their quality. Market app store providers such as Google Play can use these techniques to automatically identify awkward apps to have safe and quality products and can guide the market.

According to Oh, J, et al. [11], constructing a combined market place to facilitate the involvements of users with different apps is eventual aim. To deal with this issue, authors first perform a simple content analysis on app reviews from the developer's stance. They proposed algorithm which robotically sort out informative reviews presenting user concentration. The initial assessment conclude document the effectiveness of

their algorithm. Guzman et al. [12] introduced an approach to automatically extract fine-grained and coarse grained features from the text of app reviews. Second, authors present a method for aggregating the sentiments of many users for a feature by applying automated sentiment analysis techniques on app reviews. When a feature is present in the sentence, the sentiment score of the sentence is assigned to the feature. Finally, authors apply a topic modeling algorithm to the extracted features and their associated sentiment scores to create a more coarse-grained summary, which groups features that are mentioned in the same reviews.

Khalid et al. [13] qualitatively analyze app reviews to understand the user's perspective behind app ratings and identify the most frequent complaints of mobile apps. Lacob et al. [14] have extracted feature requests from user reviews based on linguistic rules. According to Pagano et al. [15], the AppStore shows that user publishes feature requests and other messages on improving the application. The study explores how and when users provide feedback, analyzing correlations between feedback metadata and its impact on the application popularity. Lacob et al. [16] analyzed the mobile app reviews to have insights of the most frequent issues which users report through app stores, and discussed how the rating and price of app can have influence on quantity and type of feedback. Khalid et al. [17] study provides developers vision into the issues which are reported by user in iOS apps. Their methodology was first they selected apps for which they wanted to analyze. Table 1 shows some of key terms used in user reviews in different studies.

Table 1. Compiled Chart of Key Terms Used in User Review Studies

Purpose	Literature											
	A	B	C	D	E	F	G	H	I	J	K	L
User Requirements						✓			✓	✓		
Pricing			✓									
Development	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rating			✓									
Interface						✓			✓	✓	✓	
User Satisfaction		✓			✓	✓					✓	
Quality	✓						✓			✓		✓
User Sentiments				✓			✓					
Review Quality	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

A:Vasa et al. [2], B: Chen et al. [4], C: Finkelstein et al. [8], D: Hoon et al. [9], E: Fu et al [10], F: Oh et al. [11] , G: Guzman et al.[12], H: Khalid et al. [13], I: Lacob et al.[14], J: Pagano et al. [15], K: Lacob et al. [16] , L:Khalid et al. [17]

III. ANALYSIS OF CURRENT REVIEW SYSTEM

Mobile app review systems are providing a way to the users to share their experiences with app in the form of reviews. Users who have downloaded the app can rate an app depending on his/her personal experience. In addition to this users have privilege to write a free text review. In this section, we have analyzed how the current review systems are provisioning users to share their experiences. Since app reviews can help users in decision making about an app but as number of reviews increases it becomes difficult for a user to spot weak points of an application reported by others. Although, ratings of an app is representing overall popularity of the app but in our opinion users should have more concrete information to buy or download an app.

Online apps distribution platform provides a market place for third-party app distributors. App stores have a public review system, which allows users to express their opinions concerning the apps they used. These reviews can persuade download decisions via polarized response which is 1 star to 5 stars and user review comments. High ratings and better comments can lead to improve ranking in app store by which app can be more visible. However, it is not clear what kind of information these app reviews contain and how n analysis of these reviews can inform developers about the requirements of the users. For analysis of app review system, we chose Google Play store for analyzing how current user review system is provisioning users to get quality app reviews and what are the limitations users are facing in this regard. At first, we have provided a brief introduction to the features provided by the review system later, we will discuss what the limitations are in the current review system and how different features can be useful for the users to provide quality reviews and decision making.

Fig.1 shows view of review system on smartphone. Google play store added new feature to sort review but it still requires some improvement. Equation (1) and Fig.2 shows that how rating of app is calculated.

$$((4270079*5=21350395) + (1159529*4= 4638116) + (517424*4= 1552272) + (229722*4= 459444) + (398280*4= 398280)) / 6575074=4.3$$

(1)

Fig.3 shows screenshot of options where user can give feedback about a certain review on app store that how he/she find out, whether review written by other user is helpful or not.
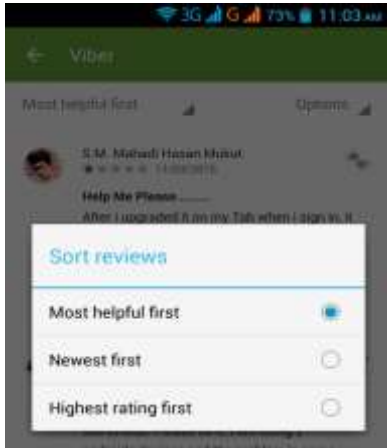

Fig.1. User Reviews Categories on Smartphone

Fig.2. Ratings of Viber on Google Play Store.

Fig.3. Detecting Spam and Helpfulness.

Reviews are listed as per criteria selected based on following options as shown in fig.4.

- First Option: As shown in Fig.4 a user has three options to sort the reviews 1) *Newest*- Where most recent reviews are posted first 2) *Rating*- Reviews will be listed according to rating value in descending order which means 5-Start rating review will come first and then so on and 3) *Helpfulness*-Users can declare a review as helpful or unhelpful and review can be sorted according to helpfulness as shown in fig.3.
- Second Option: *All Versions*: Reviews of all versions will be shown and *Latest Version*: Only reviews of latest version will be listed.

Following are few limitations identified in the current app review system:

- There is no option for reviews to be sorted according to different categories.
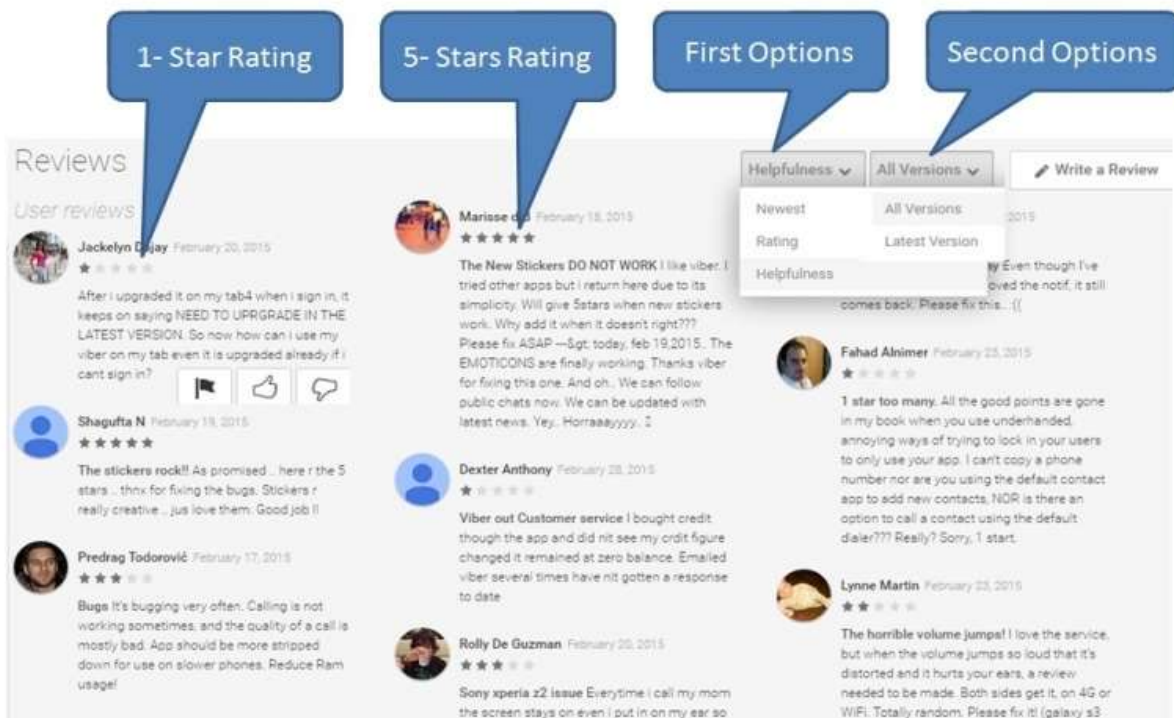

Fig.4. App Reviews on the Web

- Users can only view 5-stars ratings first and there are no options to view low star ratings.
- It is not visible how many number of users found a

review helpful or unhelpful.
- There is no privilege to sort the reviews according to unhelpfulness.

- If other users want to reply on the reviews given by other users there is no such option available. This kind of feature can trigger different users' opinions about the different functionalities of apps.

IV. RECOMMENDATIONS FOR APP REVIEW SYSTEM

In this section, we have discussed the limitation of the existing review system and suggested how the current review system can be improved to get quality app reviews. Depending on the limitation identified, we have proposed enhanced review system that can provision users to provide more informative reviews and get better decision making regarding an app. Fig.5 provide an overview of the proposed review system and brief description is as follows.

Fig.5. Recommendations of proposed review system

A. Category

As discussed in literature review section and summarized in Table 1, we proposed a list of categories. Although in existing system developer can use different data mining techniques to sort out reviews but user who wants to download or want to buy an app can't use these techniques to get more specific information but it requires lots of work to do and result may not be very accurate. To manually extract your reviews of interest is difficult due to the varying amount of user reviews. So If a person selects category before writing a review as in proposed system it will be very useful to know what actually user want to address in the review. We will be able to write and view app reviews as per our interest. For example if a user wants to buy an app it can easily sort out reviews of other users who are satisfied with this specific app. Ha et al. [18] manually analysed android user's reviews what exactly user want to verify if users are concerned about security and how much they speak about security and permissions in the comments. Only 1% of the comments contain references to application permissions. They found that majority of reviews are on quality of applications. Pagano et al. [15] argue that analysing data with strict syntactic entities can mislead developers e.g " *I lost all my phone contacts. Great, thank you*!". So this necessitate for a more quantitative feedback becasuse it is difficult to mine syntax and semantics of a review in true sense. But when a user is given an option to catagorise their own review it will show clear idea about the sentiments.

B. Star Rating

Although existing system provides star rating (5-Highest & 1-Lowest) with its count as well but when you want to view reviews it will show you 5-Star reviews first. In case of Viber app there are 4,270,079 reviews of 5-star rating only. So it means users have to spend much effort in finding negative aspects of an app. This system gives you freedom to view an apps pros and cons equally as you can go to any star rating category directly. By this feature one can be able to view reviews in ascending or descending order both equally and you get a real picture of apps performance. Linares et al. [19] demonstrate a correlation among the stability of APIs and the Android apps success. They also use user reviews for finding evidence of problems. They manually analyze the reviews of apps which contain low ratings. They look for reviews that contain some predefined error-related keywords. Their aim is to assess that the identified unsuccessful apps express the cause of the errors in their reviews. With our new features lowest star rating reviews can be easily extracted. It has been evident that the reviews that contain bug reports or other critical app issues have lowest average rating less than 2.0 are usually ignored by the stakeholders. So we are missing important aspect by ignoring this kind of reviews. Lowest average rating reviews can be more productive to improve the quality of apps as these reviews contain critical user experience that influence download decision.

C. Highest Likes/ Dislikes

Google play store allow users to classify reviews as helpful or unhelpful but it doesn't mean that if you like a review it is always useful for you as well, it may be

irrelevant for you but review is good so our suggestion is that option should be like and dislike. In our system total number of likes and dislikes will be available with the review and can be ranked with highest number of likes and dislikes. And this way you will get majority view of the users. Reviews are sorted according to helpfulness but you can't sort according to unhelpfulness. In new system dislike option will give pure feedback about a review and will discourage biased or false reviews. If someone complains about a non- existing bug which other user didn't find problem it can be disliked. User feedback content have impact on ranking of an app hence more visibility and increased number of downloads [15]. If a user wants to buy or download an app so he/she is able to decide it easily by just having a look on number of likes or dislikes. This will show more realistic statistics about app. Also with the help of number of likes and dislikes we can judge about the capability and authenticity and evaluate either his opinion is biased or not. So users will be more careful while writing reviews hence more productive reviews are expected. Users can be further motivated to give more quality feedback by giving incentive for innovative ideas or suggestions. Hyrynsalmi et al. [20] analysed the relevance of user ratings in Android operating system's application marketplace. They used large datasets of applications gathered from the Google Play. They found that a high average rating correlates positively with improvements in sales, and a high variance of user ratings together with a high average rating correlate positively with a high number of installations. User can be motivated to give more quality feedback by giving incentive for innovative ideas or suggestion. These productive feedbacks can be identified by highest number of likes.

D. Reply Feature

We suggest a new feature to reply a comment by users. For example a user is complaining about some bug or any other problem that can be immediately replied by other user on behalf of developers that if a person by mistake complaining about an app or missing feature. In this way we can have feedback on reviews by which we can get more quality reviews. Pagano et al. [15] discussed that current app stores are lacking in replying feature of bidirectional communication. Google developer console gives facility to app developer to reply to review publicly but this facility is not being used in most of reviews. This avoids developer to contacting specific user and to give solution for their problem. Harris [21] has surveyed that submitted negative feedback for the duration of the shopping season on different platforms. The survey results indicate that 68% of users were satisfied when they got reply of negative reviews, and 18% of these persons became loyal customers to the company, and made more purchases. Users are becoming active and are actively participating in creating a society which allows the fusion of the most diverse efforts and knowledge to produce qualitative changes in all areas of human activity [23]. We believe that by introducing reply feature in app review system can provide a similar kind of effect on users' satisfaction.

V. Conclusion

Mobile app reviews are being utilized as a crowdsource for the developers and the potential users. In this study, we have analyzed the existing literature related to mobile app reviews. Mobile app reviews are being utilized in a variety of ways. The major concern of revealed in app review studies is the quality of app reviews. To address this, we have analyzed the existing review system and identified how the current review system is provisioning users for providing real-life experience. We have identifies few shortcomings in the current system and suggested few recommendation that can improve the way the users do decision making and write reviews. By adding few more features to the existing review system can improve the quality of user feedback. The recommendations presented for the existing review system can improve the collection and analysis of user reviews can and make it more meaningful for the end users. We believe that the recommendations can help to improve the quality of app reviews from users as well as from developer's perspective. This study has a limitation that we have analyzed only one review system as a case study. This study can be further expanded to analyze other app review systems. As a future work, the enhanced review system can be built and an empirical study can be performed to analyze the effectiveness of the proposed review system.

References

[1] Christine Monaghan, T.N., App Store Rings in 2015 with New Records. 2015, Apple Press Info.

[2] Vasa, R., et al. A preliminary analysis of mobile app user reviews. In Proceedings of the 24th Australian Computer-Human Interaction Conference. 2012. ACM. 10.1145/2414536.2414577.

[3] Finkelstein, A., et al., Mining App Stores: Extracting Technical, Business and Customer Rating Information for Analysis and Prediction. RN, 2013. 13: p. 21.

[4] Chen, N., et al. AR-miner: mining informative reviews for developers from mobile app marketplace. In ICSE. 2014.

[5] Cui, G., H.-k. Lui, and X. Guo. Online reviews as a driver of new product sales. in Management of e-Commerce and e-Government (ICMeCG), 2010 Fourth International Conference on. 2010. IEEE.

[6] Islam, M.R. Numeric rating of Apps on Google Play Store by sentiment analysis on user reviews. in Electrical Engineering and Information & Communication Technology (ICEEICT), 2014 International Conference on. 2014. IEEE.

[7] Chiu, C., et al., app review analytics of free games listed on Google play.

[8] Finkelstein, A., et al., App Store Analysis: Mining App Stores for Relationships between Customer, Business and Technical Characteristics. RN, 2014. 14: p. 10.

[9] Hoon, L., et al. A preliminary analysis of vocabulary in mobile app user reviews. in Proceedings of the 24th Australian Computer-Human Interaction Conference. 2012. ACM.

[10] Fu, B., et al. Why people hate your app: Making sense of

user feedback in a mobile app store. in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013. ACM.

[11] Oh, J., et al. Facilitating developer-user interactions with mobile app review digests. in CHI'13 Extended Abstracts on Human Factors in Computing Systems. 2013. ACM.

[12] Guzman, E. and W. Maalej. How do users like this feature? A fine grained sentiment analysis of app reviews. in Requirements Engineering Conference (RE), 2014 IEEE 22nd International.

[13] Khalid, H. On identifying user complaints of iOS apps. in Software Engineering (ICSE), 2013 35th International Conference on. 2013. IEEE.

[14] Iacob, C. and R. Harrison. Retrieving and analyzing mobile apps feature requests from online reviews. in Mining Software Repositories (MSR), 2013 10th IEEE Working Conference on. 2013. IEEE.

[15] Pagano, D. and W. Maalej. User feedback in the App store: An empirical study. in Requirements Engineering Conference (RE), 2013 21st IEEE International. 2013. IEEE.

[16] Iacob, C., V. Veerappa, and R. Harrison. What are you complaining about? A study of online reviews of mobile applications. In Proceedings of the 27th International BCS Human Computer Interaction Conference. 2013. British Computer Society.

[17] Khalid, H., et al., What do mobile app users complain about? A study on free iOS apps. 2014. Software, IEEE. 10.1109/MS.2014.50

[18] Ha, E. and D. Wagner. Do android users write about electric sheep? Examining consumer reviews in Google play. In Consumer Communications and Networking Conference (CCNC), 2013 IEEE. 10.1109/CCNC.2013.6488439.

[19] Linares-Vásquez, M., et al. API change and fault proneness: a threat to the success of Android apps. in Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering. 2013. ACM.

[20] Hyrynsalmi, s., et al., the relevance of customer reviews in the mobile application marketplaces.

[21] Wolfe, C., Right Now Study Finds Retailers Can Win Back Unhappy Customers through Social Media. 2011. http://www.harrisinteractive.com.

[22] Harleen K. Flora, Xiaofeng Wang, Swati V. Chande, "An Investigation on the Characteristics of Mobile Applications: A Survey Study", IJITCS, vol.6, no.11, pp.21-27, 2014. DOI 10.5815/ijitcs.2014.11.03

[23] Thilaka, N., Amrita Vishwa Vidhyapeetham, and V. Natarajan. "The Dynamics of Online Readers Comments and the Online Media–A Content Analysis." IJITCS.

Muhammad Asif has received PhD in Information Systems Engineering from the Department of Computer and Information Science (IDI), Norwegian University of Science and Technology (NTNU), Norway, 2014. He has done his master degrees in Information system engineering from the Department of Computer and Information Science (IDI), Norwegian University of Science and Technology (NTNU), Norway in 2010. He is Assistant Professor in COMSATS Sahiwal, Pakistan. His research interests are Mobile Information Systems, Pervasive Computing, User modeling, Internet of Things and Model Driven Information Systems Engineering.

Usman Shehzaib holds an MS Computer Science and B.S Computer Science & Engineering from University of Engineering and Technology Lahore Pakistan, currently working as an Assistant Professor in COMSATS Pakistan. He has more than 8 years of experience with 6 years at the university level. His research interests include Data Mining, Cloud Computing, Big Data, Information Retrieval Systems and Information Systems Engineering.

Authors' Profiles

Mubasher Khalid has obtained his B.E Electrical Engineering degree from Air University and MS Computer Science degree from COMSATS Institute of Information Technology in 2011 and 2015 respectively. He is working in Shalimar Group as Chief IT Officer. His research interests include Data Management, Information System, Mobile Apps and Computer Architecture.