App store mining and analysis

7 authors, including:

Afnan Alsubaihin

University College London

24 PUBLICATIONS   **682** CITATIONS

SEE PROFILE

Yue Jia

University College London

69 PUBLICATIONS   **8,970** CITATIONS

SEE PROFILE

Mark Harman

University College London

530 PUBLICATIONS   **37,338** CITATIONS

SEE PROFILE

William Martin

10 PUBLICATIONS   **1,085** CITATIONS

SEE PROFILE

App Store Mining and Analysis (Keynote)

Afnan Al-Subaihin, Anthony Finkelstein, Mark Harman*, Yue Jia,
William Martin, Federica Sarro and Yuanyuan Zhang
Department of Computer Science, University College London, London, UK

ABSTRACT

App stores are not merely disrupting traditional software deployment practice, but also offer considerable potential benefit to scientific research. Software engineering researchers have never had available, a more rich, wide and varied source of information about software products. There is some source code availability, supporting scientific investigation as it does with more traditional open source systems. However, what is important and different about app stores, is the other data available. Researchers can access user perceptions, expressed in rating and review data. Information is also available on app popularity (typically expressed as the number or rank of downloads). For more traditional applications, this data would simply be too commercially sensitive for public release. Pricing information is also partially available, though at the time of writing, this is sadly submerging beneath a more opaque layer of in-app purchasing. This talk will review research trends in the nascent field of App Store Analysis, presenting results from the UCL app Analysis Group (UCLappA) and others, and will give some directions for future work.

General Terms

Design, Experimentation, Measurement

Keywords

App stores, Mining Software Repositories

1. APP STORE MINING AND ANALYSIS

We believe that app stores are scientifically, technically, sociologically and commercially very different from traditional software deployment mechanisms [7, 19, 20]. In particular, they create a software ecosystem [14] that provides researchers with exciting opportunities, not previously available for software engineering research.

---

*This keynote will be given by Mark Harman, but reports the joint work of the UCLappA group: A. Al-Subaihin, A. Finkelstein, Y. Jia, W. Martin, F. Sarro, and Y. Zhang. UCLappA website: http://www0.cs.ucl.ac.uk/staff/F.Sarro/projects/UCLappA/home.html

In 2012, we set out a research agenda for App Store Mining and Analysis, motivated as follows:

> "*never before has there been a nexus of readily available information that combines the users' view, the developers' claims and the sales information pertinent to a large corpus of software products from many different providers. The combination of these three types of information provides a rich and inter-related set of data from which we can analyse and understand this new software engineering paradigm of app development.*" [7]

This keynote, will review our progress and future directions in the development of this research agenda. It will discuss the importance of features [7, 10] as a suitable level of abstraction with which to discuss apps and app stores, presenting initial results about the migration of features through app stores [24]. The keynote will also consider the ways in which genetic improvement [8,13], can be used to improve existing software systems semi-automatically. We will focus on possibilities for improving energy consumption [1], and dynamic adaptivity, which we believe could be applied to Mobile devices [6] and the management and extension of their product lines [5].

Our group is one of many working on App Store Mining and Analysis. The keynote will also attempt to cover some of the exciting work by other researchers on App Store mining and analysis.

Unlike traditional software deployment mechanisms, we have available, in the App Store, considerable information in the form of customer feedback. This has allowed a great deal of App Store Analysis that investigates this feedback [2, 4, 9, 11, 12, 17, 21, 23, 26]. The keynote will also discuss some of the issues raised by the inherent sampling bias in such empirical studies of app stores [18].

There is also considerable potential in the analysis of the source code [16], requested permissions [22], and API calls [3] of the apps themselves, which is enriched by the contextual information from the App Stores in which they reside. Gorla et al. [3] explore API calls as a cheap and effective proxy for apps' semantic behaviour, while Linares-Vasquez et al. [16] study clones in Android apps, and maintenance [15]. Syer et al. [25] investigate the platform dependence of app defects.

We hope that this keynote will serve to stimulate further interest in the App Store Ecosystems, their mining and analysis and the new software engineering challenges and opportunities they create.

2. REFERENCES

[1] B. Bruce, J. Petke, and M. Harman. Reducing energy consumption using genetic improvement. In *Genetic and evolutionary computation conference (GECCO 2015)*, Madrid, Spain, July 2015.

[2] B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh. 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*, KDD '13, pages 1276–1284. ACM, 2013.

[3] A. Gorla, I. Tavecchia, F. Gross, and A. Zeller. Checking app behavior against app descriptions. In *36th International Conference on Software Engineering (ICSE 2014)*, pages 1025–1035, 2014.

[4] E. Guzman and W. Maalej. How do users like this feature? a fine grained sentiment analysis of app reviews. In *Requirements Engineering (RE 2014)*, pages 153–162, Aug 2014.

[5] M. Harman, Y. Jia, J. Krinke, B. Langdon, J. Petke, and Y. Zhang. Search based software engineering for software product line engineering: a survey and directions for future work (keynote paper). In 18th *International Software Product Line Conference (SPLC 14)*, pages 5–18, Florence, Italy, September 2014.

[6] M. Harman, Y. Jia, W. B. Langdon, J. Petke, I. H. Moghadam, S. Yoo, and F. Wu. Genetic improvement for adaptive software engineering (keynote). In 9th *International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS 2014)*, pages 1–4, New York, NY, USA, 2014. ACM.

[7] M. Harman, Y. Jia, and Y. Zhang. App Store Mining and Analysis: MSR for App Stores. In *Proceedings of the 9th IEEE Working Conference on Mining Software Repositories (MSR '12)*, pages 108–111, Zurich, Swiss, June 2012. IEEE.

[8] M. Harman, W. B. Langdon, Y. Jia, D. R. White, A. Arcuri, and J. A. Clark. The GISMOE challenge: Constructing the pareto program surface using genetic programming to find better programs (keynote paper). In 27th *IEEE/ACM International Conference on Automated Software Engineering (ASE 2012)*, pages 1–14, Essen, Germany, September 2012.

[9] L. Hoon, R. Vasa, J.-G. Schneider, and J. Grundy. An analysis of the mobile app review landscape: Trends and implications, 2014. available on line from Swinbourne University of Tethnology, Australia.

[10] C. Iacob and R. Harrison. Retrieving and Analyzing Mobile App Feature Requests from Online Reviews. In *Proceedings of the 10th Working Conference on Mining Software Repositories (MSR '13)*, San Francisco, California, USA, 18-19 May 2013.

[11] H. Khalid. On identifying user complaints of iOS apps. In D. Notkin, B. H. C. Cheng, and K. Pohl, editors, *35th International Conference on Software Engineering (ICSE 2013)*, pages 1474–1476. IEEE/ACM, 2013.

[12] H. Khalid, E. Shihab, M. Nagappan, and A. Hassan. What do mobile app users complain about? A study on free iOS apps. *IEEE Software*, 32(3):70–77, 2014.

[13] W. B. Langdon and M. Harman. Optimising existing software with genetic programming. *IEEE Transactions on Evolutionary Computation (TEVC)*, 2014. To appear.

[14] S. L. Lim and P. J. Bentley. Investigating app store ranking algorithms using a simulation of mobile app ecosystems. In *IEEE Congress on Evolutionary Computation*, pages 2672–2679, 2013.

[15] M. Linares-Vásquez. Supporting evolution and maintenance of android apps. In *36th International Conference on Software Engineering (ICSE 2014) Doctoral Symposium*, pages 714–717, 2014.

[16] M. Linares-Vásquez, A. Holtzhauer, C. Bernal-Cárdenas, and D. Poshyvanyk. Revisiting android reuse studies in the context of code obfuscation and library usages. In *11th Working Conference on Mining Software Repositories (MSR 2014)*, pages 242–251, 2014.

[17] W. Maalej and H. Nabil. Bug report, feature request, or simply praise? on automatically classifying app reviews. In *Requirements Engineering (RE '15)*, 2015. to appear.

[18] W. Martin, M. Harman, Y. Jia, F. Sarro, and Y. Zhang. The app sampling problem for app store mining. In *Mining Software Repositories (MSR'15)*, Florence, Italy, May 2015.

[19] T. Menzies. Beyond data mining; towards "Idea Engineering". In 9th *International Conference on Predictive Models in Software Engineering, PROMISE '13*, Baltimore, MD, USA, Oct. 2013. ACM.

[20] R. Minelli and M. Lanza. Software Analytics for Mobile Applications - Insights & Lessons Learned. In *Proceedings of the 17th European Conference on Software Maintenance and Reengineering (CSMR '13)*, Genova, Italy, 5-8 March 2013. IEEE.

[21] D. Pagano and W. Maalej. User feedback in the appstore: An empirical study. In *requirements engineering (RE 2013)*, pages 125–134. IEEE, 2013.

[22] R. Pandita, X. Xiao, W. Yang, W. Enck, and T. Xie. WHYPER: Towards Automating Risk Assessment of Mobile Applications. In *Proceedings of the 22nd USENIX Security Symposium*, Washington DC, USA, 14-16 August 2013.

[23] I. J. M. Ruiz, M. Nagappan, A. Bra, T. Berger, S. Dienst, and A. E. Hassan. On the relationship between the number of ad libraries in an android app and its rating, 2014. available on line from Queen's University, Canada.

[24] F. Sarro, A. AlSubaihin, M. Harman, Y. Jia, W. Martin, and Y. Zhang. Feature lifecycles as they spread, migrate, remain and die in app stores. In *Requirements Engineering (RE'15)*, Ottawa, Canada, August 2015. To appear.

[25] M. D. Syer, M. Nagappan, B. Adams, and A. E. Hassan. Studying the relationship between source code quality and mobile platform dependence. *Software Quality Journal*, 2014. To appear; available online.

[26] S. E. S. Taba, I. Keivanloo, Y. Zou, J. Ng, and T. Ng. An exploratory study on the relation between user interface complexity and the perceived quality of android applications. In *International Conference on Web Engineering (ICWE 2014)*, 2014. Late Breaking Result.