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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DeMobile 2015, Bergamo
Copyright 20XX ACM X-XXXX-XX-X/XX/XX ...$15.00.

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 consider 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


