

Poster: What are mobile developers asking about? A large scale study using stack overflow

Shantanu Sharma
North Carolina State University
Raleigh, North Carolina
ssharm34@ncsu.edu

ABSTRACT

In this paper, the authors use data from the popular online QA site, Stack Overflow, and analyze 13,232,821 posts to examine what mobile developers ask about. They employ Latent Dirichlet allocation based topic models to help summarize the mobile-related questions. Their findings show that developers are asking about *app distribution, mobile APIs, data management, sensors and context, and mobile tools and user interface development*. Authors also determine what popular mobile-related issues are the most difficult, explore platform specific issues, and investigate the types (e.g., what, how, or why) of questions mobile developers ask. The findings help highlight the challenges facing mobile developers that require more attention from the software engineering research and development communities in the future and establish a novel approach for analyzing questions asked on QA forums.

1 INTRODUCTION

The popularity of mobile devices has seen an exponential growth in the past few years. These smart phones heavily rely on software from the operating systems they use to the applications they run. Therefore, a growing number of software engineering practitioners and researchers are starting to focus on the issues facing mobile software development. One of the main findings of a recent study by Minelli and Lanza [2] concluded that mobile applications are different than most desktop applications. For example, Minelli and Lanza found that mobile applications are much smaller than traditional software systems, that they are complex due to their reliance on third-party libraries, and that they are developed by much smaller teams. The Authors used publicly available Stack Overflow data to investigate, on a large scale, the most common issues encountered by mobile developers. Authors applied Latent Dirichlet Allocation (LDA) based topic models on the mobile-related Stack Overflow posts to determine what mobile developers are asking. They then analyze the topics and rank them based on their views, look at how many questions in each topic receive answers that are accepted by the questioner and their response time. Last, we look at the popular issues for the three major mobile platforms: Android, iOS, and Windows Phone. The authors have formalized the study in the following four research questions

RQ1: What issues are mobile developers asking about?

RQ2: What issues are the most difficult?

RQ3: Do developers of different mobile platforms face different issues? What are the most popular issues for the Android, iOS, and Windows Phone mobile development platforms?

RQ4: What types of questions are mobile developers asking?

Hardware	Platform	Development
iPhone	symbian	xcode
iPad	blackberry	cocoa
galaxy	android	iOS sdk
nexus	windows phone	adobe air
tablet	palm os	blackberry-cascades
nokia	bada	phonegap
	maemo	qt sdk
	meego	appceleator
	blackberry-10	mosync
	webos	javaME
	tizen	
	windowsPhone	
	palmOS	

Figure 1: List of words used to identify mobile-related posts

2 METHODOLOGY

2.1 Data Extraction and Processing

Our first step was to download the Stack Overflow data dump (last updated March 2013) in XML format. After parsing the Stack Overflow data the initial data set had had a total of 13,232,821 posts.

2.2 Identifying Mobile Posts

Posts on Stack Overflow can be about any topic, so ways to identify mobile-related posts is needed.

2.2.1 Search. Start with Initial list

2.2.2 Add tags. Add additional tags from the step above and expand list

2.2.3 Remove Irrelevant Tags. In order to determine which of our tags are relevant to mobile related posts exclusively, use a tag relevance threshold $TRT_{tag} = (\frac{No.of\ mobile\ posts}{Total\ no.\ posts})$

Once irrelevant tags are filtered out using the TRT value, we may end up with some tags that had very few posts associated with it that were related to a very specific problem, hence another value to filter out was used and is defined as the tag significance threshold $TST_{tag} = (\frac{No.of\ mobile\ posts}{No.of\ mobile\ posts\ for\ the\ most\ popular\ tag})$.

Experiments showed $TRT > 45$ and $TST > 1$ gave a good set of posts.

2.2.4 Extract all posts. Extract all post identifiers of the questions that contained these mobile tags. After these questions were identified, authors also extracted the posts that answer these questions. The question and answer posts make up our final mobile-related

Classification	# Ques.	Average Views
App Distribution	7,298	1,439
Connectivity	19,616	1,123
Tools	66,935	1,079
User Interface	58,793	1,069
Data Formatting	9,735	1,055
Phone Orientation	10,034	1,047
Lists	31,010	963
Media	15,673	947
File Operations	21,190	929
Input	36,779	906

Figure 2: Top 10 topics discovered by LDA based on average views

Topic	Mean Time (Days)	Median Time (Mins)	% Accepted	Avg. # of
App Distribution	10.59	46	55	1.59
Connectivity	13.27	89	45	1.43
Tools	11.03	66	56	1.62
User Interface	9.80	30	58	1.58
Data Formatting	4.48	17	68	1.82
Phone Orientation	12.78	46	52	1.50
Lists	7.91	32	56	1.49
Media/Images	10.47	40	52	1.42
File Operations	8.62	45	55	1.48
Input	9.93	56	55	1.52
Exceptions	7.14	35	56	1.56
Contacts	7.82	21	59	1.68
Data Structures	3.07	21	71	1.93
HTML5/Browser	13.49	132	47	1.39

Figure 3: Average time until accepted answer and percent of questions w/ accepted answers for classifications(trunc.)

dataset that we use in the remainder of our experiments. In total, our mobile-related corpus reduced to 1,642,602 posts.

2.3 Topic Modeling Using LDA

Experimented with number of topics(k) and converged on k=40. Other hyper parameters were experimented and parameters were optimized by running the tool many times and manually inspecting the results. After many trials, we chose the parameters that generally gave topic keywords that are more semantically related and where the instances had a higher average probability assigned to the most relevant topic.

3 RESULTS

The results are summarized in Figure 2 - 5

RQ1: What issues are mobile developers asking about? Figure 2

RQ2: What issues are the most difficult? Figure 3

RQ3: Do developers of different mobile platforms face different issues? What are the most popular issues for the Android, iOS, and Windows Phone mobile development platforms? Figure 4

RQ4: What types of questions are mobile developers asking? Figure 5

Android			iOS		
Avg. Views	# Ques.	Issue	Avg. Views	# Ques.	Issue
1,278	1,574	App Distribution	1,592	4,851	App Distribution
1,247	14,945	Connectivity	1,244	14,395	User Interface
1,120	3,821	Orientation	1,098	29,018	Tools
1,091	24,731	Tools	1,097	5,502	Data Formatting
1,053	39,278	User Interface	1,029	5,082	Orientation
999	10,332	File Operations	1,021	10,628	HTML5
975	24,935	Data Formatting	1,015	2,584	Phone/Sensors
971	7,582	Media/Images	986	6,682	Media/Images
946	17,333	Input	969	13,703	Input
933	6,792	Media/Streaming	922	19,052	Tables

Figure 4: Top topic for android and iOS

Platform	% Why	% How	% What	% Other
iOS	27	56	10	7
Android	30	52	8	10
Windows phone	27	62	6	5

Figure 5: Type of questions asked

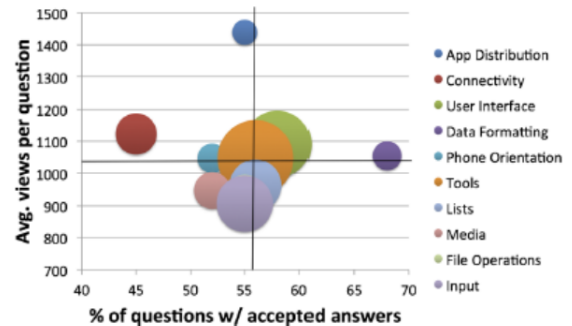


Figure 6: Significance of top mobile related issues

4 CONCLUSIONS

In this paper, authors conclude that the most popular questions include those relating to app distribution, mobile tools, and user interface development. Further they find that mobile questions are more difficult to answer than the none mobile questions on Stack Overflow. They also do a comparison between the different popular mobile platforms (Android, iOS, and Windows Phone), and find questions related to app distribution, user interface, and input are the most popular for all of them.

REFERENCES

- [1] Rosen, C. and Shihab, *What are mobile developers asking about, A large scale study using stack overflow*. Empirical Software Engineering June 2016, Volume 21, Issue 3, pp 1192-1223
- [2] Minelli R, Lanza *Software analytics for mobile applications insights and lessons learned*. 2013 17th European conference on software maintenance and reengineering (CSMR), pp 144-153