UNDERSTANDING PRIVACY-RELATED QUESTIONS ON STACK OVERFLOW

Mohammad Tahaei, Kami Vaniea, Naomi Saphra School of Informatics, University of Edinburgh {mohammad.tahaei, kami.vaniea, naomi.saphra}@ed.ac.uk

Comic by @zezhong_w



Privacy is challenging for developers. they see it as security, someone else's job, or full of legal jargon.



One of them is like this: You must provide a valid Privacy Policy URL in order take your app Live. Go to App Details and make sure it is valid sked 5 years, 5 months ago Active 3 months ago Viewed 24k times i can't publish my facebook application, when i click "yes" on Status and Reviews of developers platform see this message "You must provide a valid Privacy Policy URL in order take your app Live. Go to App Details and make sure it is valid." in privacy field i have a right url and i tried also to change it with others but continue to see the messsage asked Nov 15 '14 at 9:59

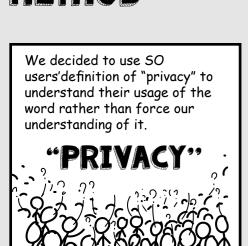
RESEARCH QUESTIONS

1. What topics do SO users associate with the word "privacy"?

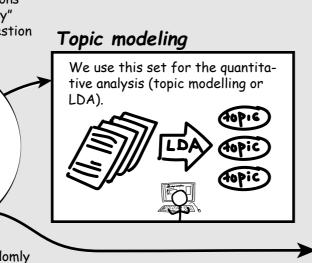


2. What or who is pushing 50 users to engage with privacy-related topics?

METHOD







And also reviewed the output of the topic modelling. LDA HOPIC) 40pic

DRIVERS

Laws & Regulations

5 (2%)

Platforms

to fix it ?" [53097654 - 2018].

all [...]" [54952842 - 2019].

camera

5 (2%)

Encrtption

10 (3%)

Versions and code issues

18 (6%)

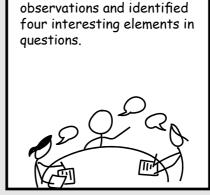
Developers as end-users

21 (7%)

app under the privacy photos list." [32646366 - 2015].

<u>Develop</u>ers ignoring PbD principles

11 (4%)



"I was unable to find the answers on the dialogflow/google terms of use/privacy statement, so I ask

there a comprehensive information available which data and how it is processed within the DialogFlow API (wrt data protection)? (This would be needed in order to inform our users of our ChatBot about

"I am submitting my app on App Store Connect\My App page and when I submit for review, it shows

error on App Information: "You must provide a Privacy Policy URL." even I have pasted the link to

book.com/tools/debug/sharing/ and they show no error. Do you know what could be the reason and how

"How would you design a privacy setting on a per-user basis? For example, Id like my email to be shown

only to connections where as my alias can be shown to everybody and my phone number to be hidden to

• They define what is a sensitive content such as access to contacts, location, and

"We use HPE to check the code potential risks, i got one critical issue below in Log util class "The method d() in LogUtil.java mishandles confidential information, which can compromise user privacy

"Since the salt is used to add a huge range of password possibilities [...] what is the purpose of letting

the salt insecure? [...] Is there something that I don't understand? I know that knowing the salt don't

break the security but, say- ing that it "need not be kept secret" sounds strange to me." [6176848 -

"But sometimes iPhone 5s running iOS 8.4 and always iPhone 6 Plus running iOS 9 does not show my

"How should an app communicate with a server operated by its developer without android.permis-

sion.INTERNET? Or is there a reliable source stating that this is impossible in Android?" [29545251

the website show the privacy policy there. I have checked the link using https://developers.face-

here: - Are all the Dialogflow API Servers located in the EU - i.e. do GDPR Rules apply here? - is

their data usage here... At least, we have to follow GDPR.)" [53052121 - 2018].

Personal concerns, client or company requirements

• Platforms are gatekeepers of publishing apps

PRIVACY ASPECTS

Privacy and code issues

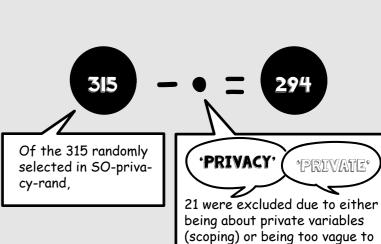
Examples include Apple iOS and Google Android

• Developers try to follow the requirements set by platforms

and is often illegal". how can i do to fix this?" [44410004 - 2017].

Then during multiple discus-

sion sessions, they shared



Number of questions (294 in total)

Number of questions (294 in total)

136 (46%)

144 (49%)

understand.

CODING FINDINGS

Qualitative Coding

independently read through

40 questions drawn random-

40 questions

40 questions

Two researchers first

ly from SO-privacy.

QUESTION TYPES

Unexpected

12 (7%)

Number of questions (294 in total)

"Currently I am developing audio recording app which record the audio hence I set microphone permission in info.plist file so record audion permission alert displaying in iOS 10.3.2 but its not appearing in iOS 10.3.3 devices" [46297966 - 217].

Error

46 (16%)

"I can't publish my facebook application, when i click "yes" on Status and Reviews of developers platform i see this message "You must provide a valid Privacy Policy URL in order take your app Live. Go to App Details and make sure it is valid." in privacy field i have a right url and i tried also to change it with others but continue to see the messsage. this happens not just for one application but also for others. maybe for this reason i think is it a temporary issue of developers platform?" [26944634 - 2014].

Abstract or conceptual

50 (17%)

"I've used my personal email address for [Git repository] commits and I'm trying to set it to another one, before I make the repository public. [...] Is there a way to remove it from there, too, without losing my history?" [13323759 - 2012].

How

186 (63%)

"What is the hidden cost using these CDN services? If the script is not cached by the browser and it loads the script from google what could google potentially do with the information? Could it be usefully extrapolated in conjunction with other services such as search, analytics or adsense? Nothing is free, what's the catch?" [10133816 - 2012].

Many Stack Overflow users are asking questions that are abstract or conceptual. This may signal that users ask questions about privacy related topics, in the early stages of software development, and not necessary in the implementation phases.

ACCEPTED ANSWERS

With a link to another SO question

4 (3%)

Number of questions (130 in total)

"You should take a look at my answer here which addresses some of your concerns. The tl;dr of that answer is when using 3rd party channels, the connectors don't store any of the messages, they just convert and route them" [46787477 - 2017].

With a link to unofficial resources such as websites, blogs, Wikipedia, an app, or a GitHub project

17 (13%)

"Speaking generally, I think you should do both. Any private data you collect is simply a liability for you, and not just because you intend to publish your databases. The less you can collect, the better. By the same token, however, you probably realize that it is not just IDs and passwords which are sensitive. Remember the AOL search data leak? Or the Netflix database publication? Even without having IDs, people managed to figure out the real identities of some of the accounts, simply by piecing together trails of user behavior, and corresponding that with data from other places. Some people are embarrassed by their search histories and their movie rentals. Go figure" [547545 - 2009].

With a link to official documentations, websites, blogs

settings] All Google OAuth2 scopes are listed in this table: https://developers.google.com/identity/protocols/googlescopes#oauth2v2" [47511216 - 2017].

Answered by SO users

"All you can get is **an identifier of the device**. You can get the **telephone number**, but then your **app** would be rejected. So in short, all you can do is ask the user for his personal data (or to choose his contact from the address book) - no automatic way to get these. All you can get is less personal things like IP address of the device, device type etc" [13474477 -

A lot of questions are being answered by Stack Overflow users without any links to other resources, which could mean that the community is building knowledge around privacy related topics on Stack Overflow.

TOPIC MODELING RESULTS

Topic label

- 1. Access to and read contents 2. Set the privacy field
- 3. App purchase and user registration 4. Privacy and permission settings and dialogues 5. Crash reporting, analytics tools, and trackers
- 6. PPs in Google Play and Android 7. Concerns about using Google services 8. Publicity of sensitive data in code repositories
- 9. Design a db schema with privacy settings 10. Privacy values in Facebook, YouTube, and plists
- 11. Image privacy statements in Instagram and Windows
- 12. Store users' sensitive data securely 13. Access to, create, and upload photos and albums 14. Private and public variables

15. Browsers errors (cookies and security settings)

Top five words app, application, use, android, privacy

user, privacy, like, page, facebook device, ios, cloud, feature, access app, user, privacy, access, ios crashlytics, tracker, integrate, news, advertiser app, policy, privacy policy, privacy, store google, button, use google, ad, click analytics, firebase, repository, google analytics, git table, privacy, column, transaction, mysql privacy, post, set, facebook, api image, windows, statement, instagram, privacy

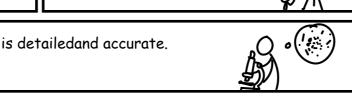
photo, album, picture, save, access photo file, private, privacy, use, code use, privacy, website, browser, site

Can give a quick overview of the data

datum, user, use, address, information

Topic modeling is cheap and fast		
Manual coding is slower	1	





SUPPORT FOR DEVELOPERS -

- # Identification of third-party libraries
- # Walk them through setting up a PP which correctly describes how they collect, store, and use data # Could automatically detect and adapt to changes and updates that occur in third-party libraries
- # Better integration of privacy checks into code development process such as IDEs

Support for privacy requirements from platforms during development stages

36 (28%)

"You can configure Google OAuth scope to not request the client's email from Google. [Example

76 (58%)

- 2015].

"What files/folders should I **ignore** in a **git repository** of an iOS app? [...] Do the files generated by cocoapods contain some of my private information? Does info.plist file contain my private stuff as well? Also, when I was putting Firebase into my app, I downloaded a GoogleService-Info.plist. Should I ignore it as well? What things should I ignore?" [37479924 - 2016].

Privacy policies

39 (13%)

"i can't publish my facebook application, when i click "yes" on Status and Reviews of developers platform i see this message "You must provide a valid Privacy Policy URL in order take your app Live. Go to App Details and make sure it is valid."

in privacy field i have a right url and i tried also to change it with others but continue to see the messsage. this happens not just for one application but also for others. maybe for this reason i think is it a temporary issue of developers platform?" [26944634 - 2014]

Developers as privacy concerns

71 (24%)

"Can I trust react-devtools not to breach my privacy? [...] The tool (and react) is made by Facebook, a company infamously known for their complete lack of moral when it comes to data gathering and creepy surveilance of us all. And it requires the ability to access everything you are browsing (which is probably needed to work it's magic), in order to be installed." [54549807 - 2019].

Access control

186 (63%)

"How to change a Vimeo's video privacy via API (PHP)? [...] I've followed every step specified by the Vimeo's API Documentation but I can't get it to work. What am I doing wrong?" [52080930 - 2018].



Privacy policies are challenging for developers. Potential causes of privacy policy (PP) problems for developers

- 1. They do not know that they need a PP.
- 2. They do not see a reason for adding a PP.
- 3. They do not know what language needs to be in a PP for their app's unique profile.
- 4. They are trying to add a PP but cannot do it because of complicated procedures as well as unhelpful user interfaces.
- 5. They see PPs as a wall that is blocking their app being published, with the resulting frustration leading to reluctance to prepare a well written PP.