

ANDROID APPS AND USER FEEDBACK: A DATASET FOR SOFTWARE EVOLUTION AND QUALITY IMPROVEMENT

WORKSHOP ON APP MARKET ANALYTICS - WAMA 2017

**G.GRANO, A. DI SORBO, F. MERCALDO, C. VISAGGIO
G. CANFORA, S. PANICHELLA**



GRANO@IFI.UZH.CH



GIOGRANO90

OUTLINE

- CONTEXT**
- MOTIVATION AND RELEVANCE**
- DESCRIPTION OF THE DATASET**
- ENABLED RESEARCH**

GOOGLE PLAY STORE
3 MILLIONS OF APPS
65 BILLIONS OF DOWNLOADS
~ 13\$ BILLIONS REVENUES



**APP STORES → NEW PARADIGM
RICH SOURCE OF INFORMATION:
APP DESCRIPTIONS, CHANGELOGS
USER REVIEWS**

FINDINGS FROM MOBILE STORE: DIRECT AND ACTIONABLE IMPACTS FOR APP DEVELOPER TEAMS¹

¹ Martin, Sarro, Jia, Zhang, Harman, A Survey of App Store Analysis for Software Engineering, TSE 16

INITIAL RESEARCH FOCUSED ON CLASSIFICATION² AND SUMMARIZATION³ OF USER REVIEWS

² Panichella, Di Sorbo, Guzman, Visaggio, Canfora, Gall, How can i improve my app?
Classifying user reviews for software maintenance and evolution, ICSME 15

³ Di Sorbo, Panichella, Alexandru, Shimagaki, Visaggio, Canfora, Gall, What would
users change in my app? Summarizing app reviews for recommending software changes,
FSE 16

EVOLUTION IS GUIDED BY REQUESTS IN USER REVIEWS^{4,5} STORES LACK IN FUNCTIONALITIES

⁴ Palomba, Salza, Ciurumelea, Panichella, Gall, Ferrucci, De Lucia, Recommending and localizing change requests for mobile apps based on user reviews, ICSE 17

⁵ Palomba, Linares-Vásquez, Bavota, Oliveto, Di Penta, Poshyvanyk, Lucia, User reviews matter! Tracking crowdsourced reviews to support evolution of successful apps, ICSME 15

OUR DATASET:

- ~ 280K USER REVIEWS**
- 395 APPLICATIONS**
- 22 CODE QUALITY METRICS**
- 8 CODE SMELLS**

DATASET CONSTRUCTION

We built the dataset in two phases:

→ DATA COLLECTION

FDroid + Google Play Store

→ ANALYSIS PHASE

Classification + apk analysis

DATA COLLECTION

→ FDROID

Crawler for meta-data ~ 1,929 apps

→ PLAY STORE MATCHING

Removed not matched apps or older than
2014

DATA COLLECTION

→ REVIEW CRAWLER

Mining reviews for 965 apps

→ VERSION MATCHING

Based on release and post date

→ FILTERING

Version with less than 10 review.

288k reviews for 629 versions of 395 apps!

ANALYSIS

→ USER REVIEWS CLASSIFICATION

- » Two-level taxonomy

→ CODE ANALYSIS

- » Code Quality Indicators

- » Code Smells

USER REVIEWS CLASSIFICATION

URM TAXONOMY MODEL

Two-level taxonomy

- » Intention

ARDOC⁶: reviews classifier based on NLP+SA+TA

- » Topic

SURF³: topic classifier based on topics-related keyword and n-grams

⁶ Panichella, Sorbo, Guzman, Visaggio, Canfora, Gall, ARdoc: app reviews development oriented classifier, FSE 16

³ Di Sorbo, Panichella, Alexandru, Shimagaki, Visaggio, Canfora, Gall, What would users change in my app? Summarizing app reviews for recommending software changes, FSE 16

INTENTION CATEGORIES

CATEGORY	DEFINITION
Information Giving	Informs users or developers about app aspects
Information Seeking	Attempts to obtain informations or help
Feature Requests	Expresses idea, suggestions for enhancing the app
Problem Discovery	Unexpected behaviour or issues
Other	Anything not in previous categories

EXAMPLES

PROBLEM DISCOVERY, UPDATE/VERSION

I can't access my SD card with the new update which makes this app and the ery money I donated worthless.

FEATURE REQUEST, FEATURE FUNCTIONALITY

I would give 5 stars if there was a way to move emails from the delete folder back into the inbox folder.

SOME NUMBERS...

TOPIC	SENTENCES	FR	PD	IS	IG	OTHER
App	117,409	4,879	11,089	1,600	11,943	87,898
GUI	37,620	3,381	5,034	705	3,560	2,4940
Contents	16,819	1,315	1,973	434	1,620	11,477
Download	7,853	333	1,346	363	830	4,981
Company	1672	118	190	57	152	1,155
Feature	173,847	15,480	27,810	4,342	14,972	111,243
Improvement	8,281	1,005	304	54	755	6,163
Pricing	4,016	142	216	62	559	3,037
Resources	3071	155	375	50	263	2228
Update/ Version	21,669	1,358	3,886	548	2,423	13,454
Model	22,044	1,308	3,397	459	2,055	14,825
Security	2,392	212	313	65	218	1,584
Other	189,784	630	2,019	1,402	2,842	182,891
TOTAL	606,477	30,316	57,952	10,141	42,192	465,876

CODE ANALYSIS

APKS → APKTOOL → SMALI BYTECODE

SMALI BYTECODE → PYTHON SCRIPTS → METRICS

AVAILABLE METRICS @ GITHUB WIKI

CODE METRICS

- **DIMENSIONAL METRICS**
- **COMPLEXITY METRICS**
- **OBJECT-ORIENTED METRICS**
- **ANDROID-ORIENTED METRICS**

CODE ANALYSIS

SMALI BYTECODE → PAPRIKA → SMELLS

- » Blob Class (BLOB)
- » Swiss Army Knife (SAK)
- » Long Method (LM)
- » Complex Class (CC)
- » Internal Getter/Setter (IGS)
- » Member Ignoring Method (MIM)
- » No Low Memory Resolver (NLMR)
- » Leaking Inner Class (LIC)

CODE SMELLS @ GITHUB WIKI

DATA SHARING
→ **CSV FILES**
→ **RELATIONAL DATABASE**

CSV FILES

→ VERSIONS

ID, PACKAGE NAME, CATEGORY, VERSION, RELEASE DATE

1125, org.tomdroid, Productivity, 0.7.5, January 16 2014

→ REVIEWS

ID, PACKAGE NAME, TEXT ,CATEGORY, VERSION, RELEASE DATE, STARS, VERSION ID

7bd1c70a-afc9-11e6-93ea-c4b301cdf627

org.tomdroid

Don't sync it online. The whole app crashed. I had to reinstall it.

Lost my notes. As long as you keep it in ur sd card it works good

August 24 2015

3

1125

→ SENTENCES

ID, TEXT, INTENTION, TOPIC

7bd1c70a-afc9-11e6-93ea-c4b301cdf627

Don't sync it online.

INFORMATION GIVING, Other

7bd1c70a-afc9-11e6-93ea-c4b301cdf627

The whole app crashed.

PROBLEM DISCOVERY, App

7bd1c70a-afc9-11e6-93ea-c4b301cdf627

I had to reinstall it.

OTHER, App-Update/Version

7bd1c70a-afc9-11e6-93ea-c4b301cdf627

Lost my notes.

OTHER, Contents-Feature/Functionality

7bd1c70a-afc9-11e6-93ea-c4b301cdf627

As long as you keep it in ur sd card it works good

OTHER, Feature/Functionality

→ USER METRICS

**ID, PACKAGE NAME, NO.REVIEWS, NO.SENTENCES, RATING,
FR, %FR, PD, % PD**

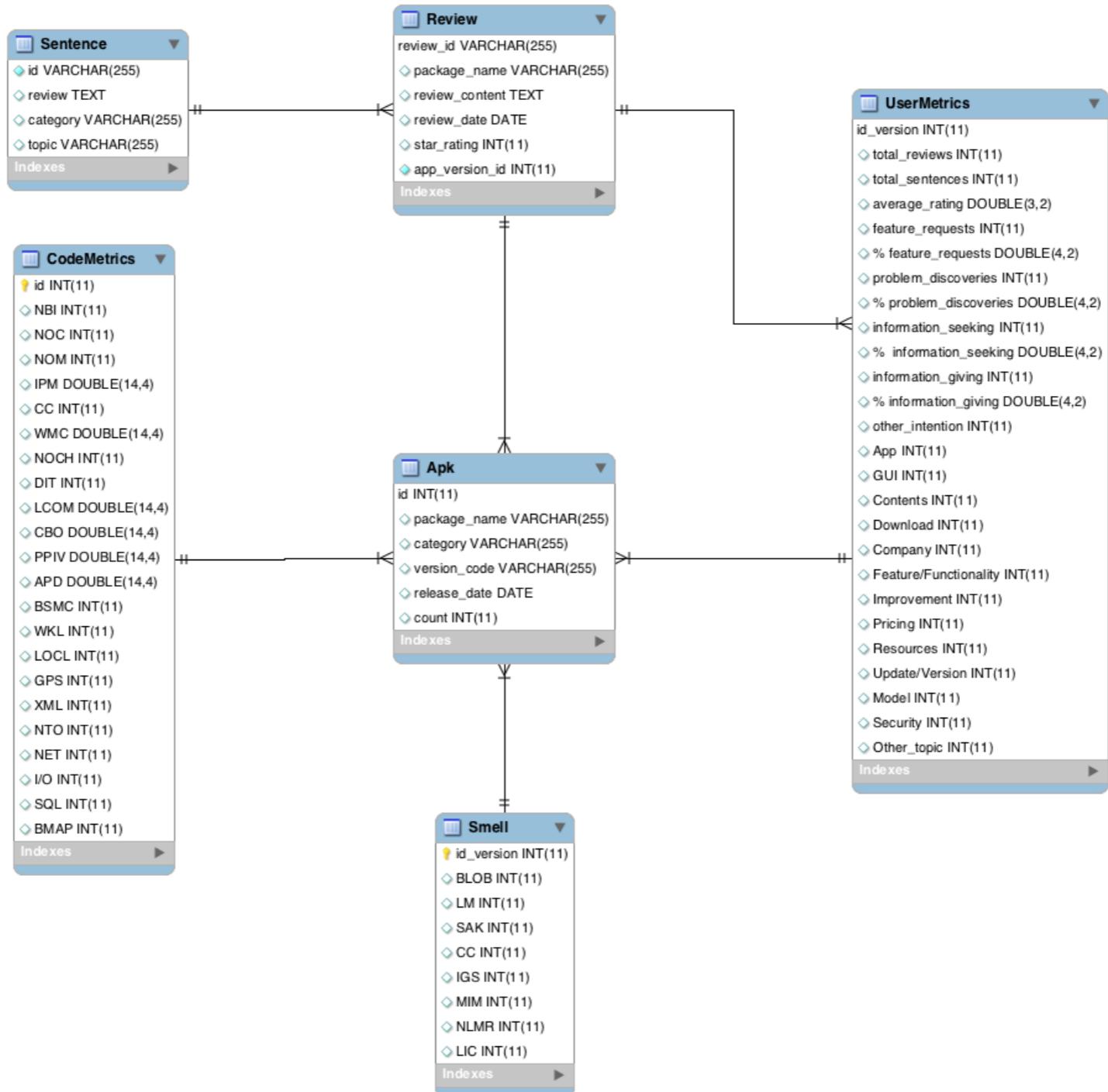
→ CODE METRICS

ID, PACKAGE NAME, <ALL METRIC NAMES>

→ CODE SMELLS

ID, PACKAGE NAME, <ALL SMELL NAMES>

RELATIONAL DB



RESEARCH OPPORTUNITIES

UNDERSTANDING HOW CODE QUALITY AFFECTS REVIEWS AND RATING FOR DIFFERENT CATEGORIES

**OBSERVE CONSEQUENCES
ON CODE QUALITY
WHILE INTEGRATING USER FEEDBACK
INTO THE APP CODEBASE**

STUDY CO-EVOLUTION TRENDS OF QUALITY METRICS, CODE SMELLS AND USER FEEDBACK FOR SEQUENTIAL RELEASES

THANKS FOR YOUR ATTENTION

DATASET @ GITHUB

“That's all Folks!”