



ON AUTOMATICALLY DETECTING SIMILAR ANDROID APPS

Mario Linares-Vásquez
Andrew Holtzhauer
Denys Poshyvanyk



You Own a Start-Up

In **two weeks**, you are meeting with this venture capitalist.

He might invest **>\$1 million** in you.



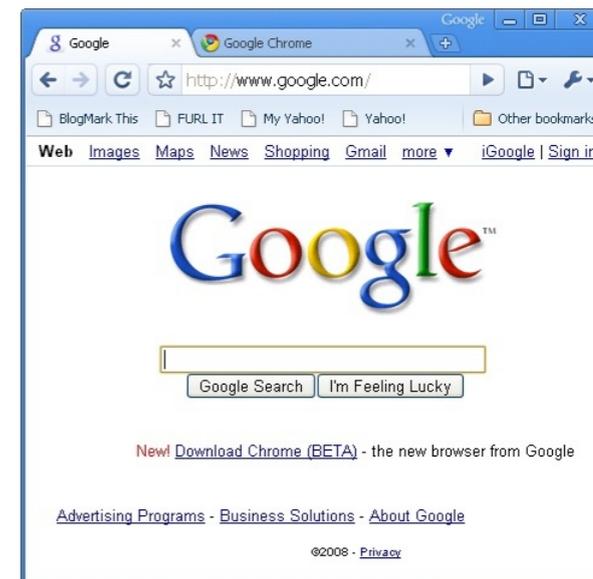
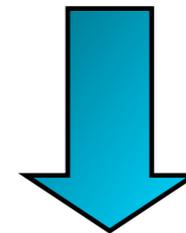
What Should You Do?

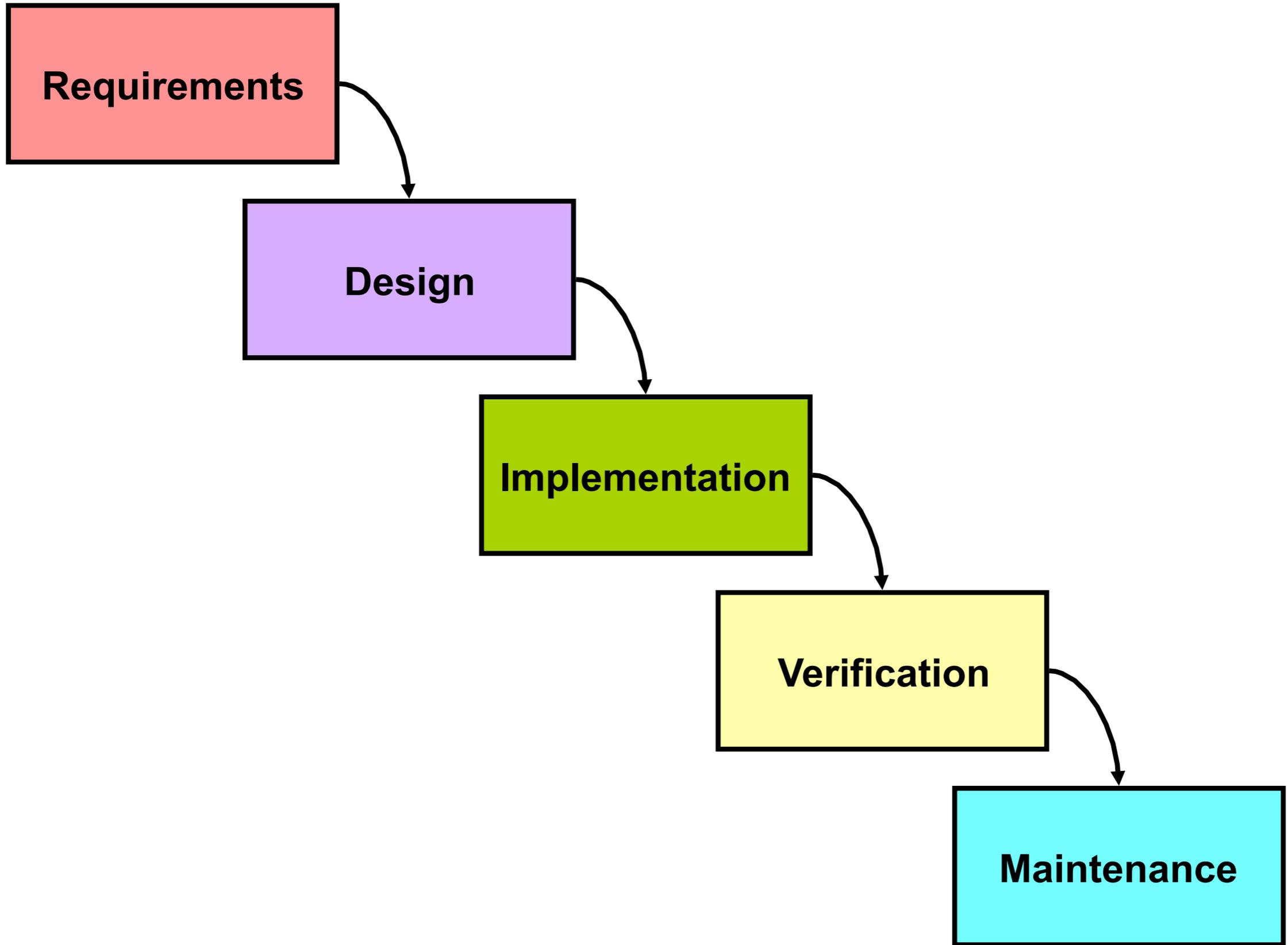
Your Product

A new web-enabled device is on the market.

You can build the first and only **web browser** for it.

What do we do?





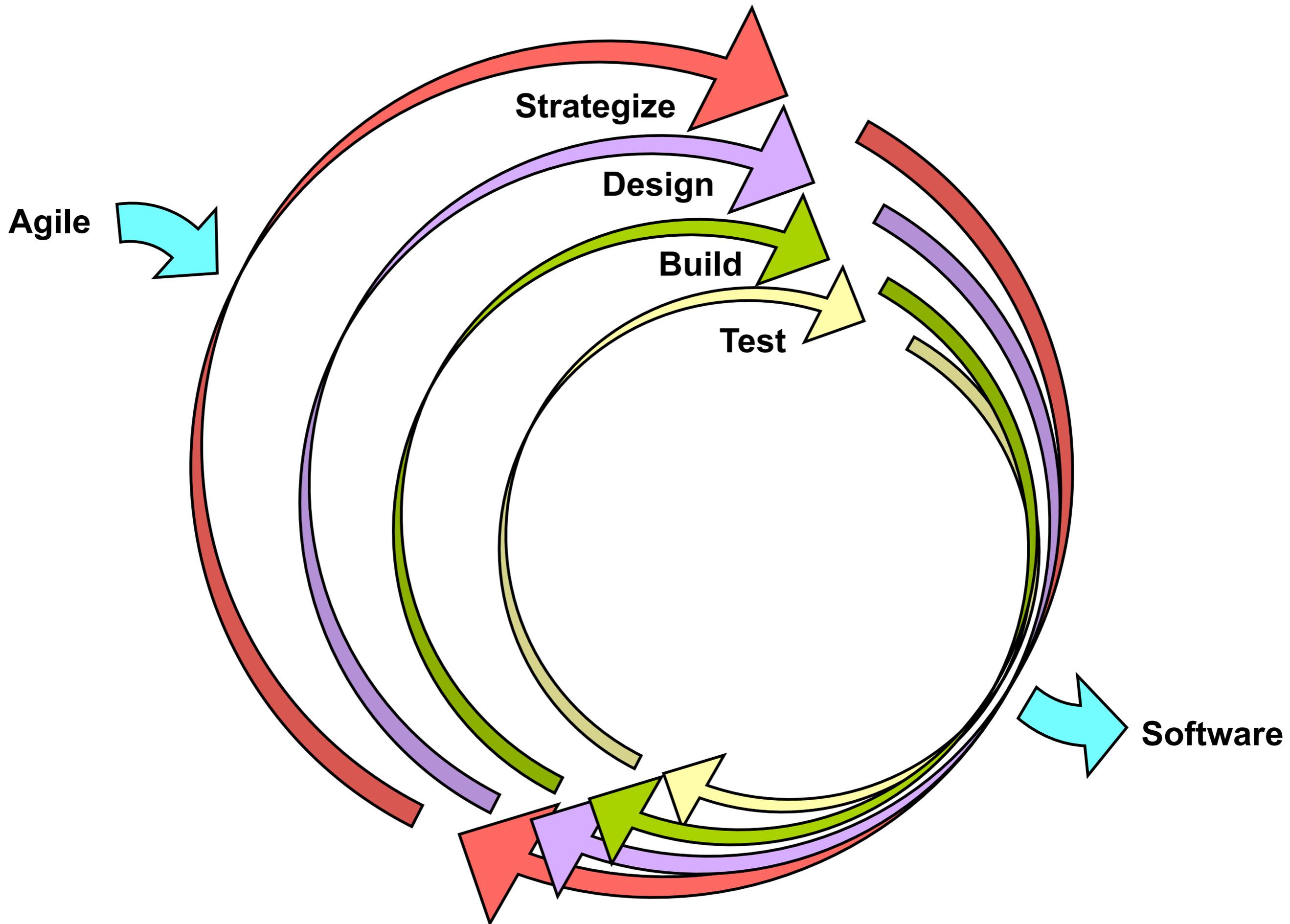
Requirements

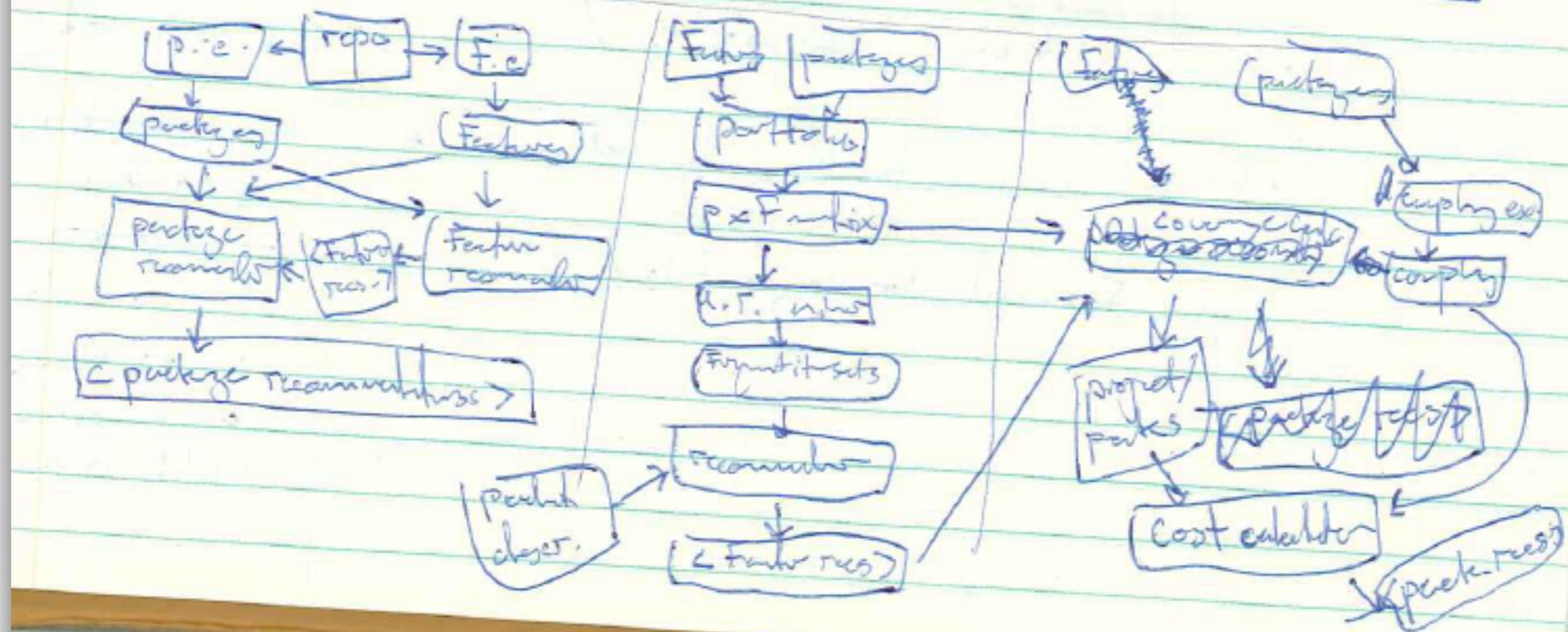
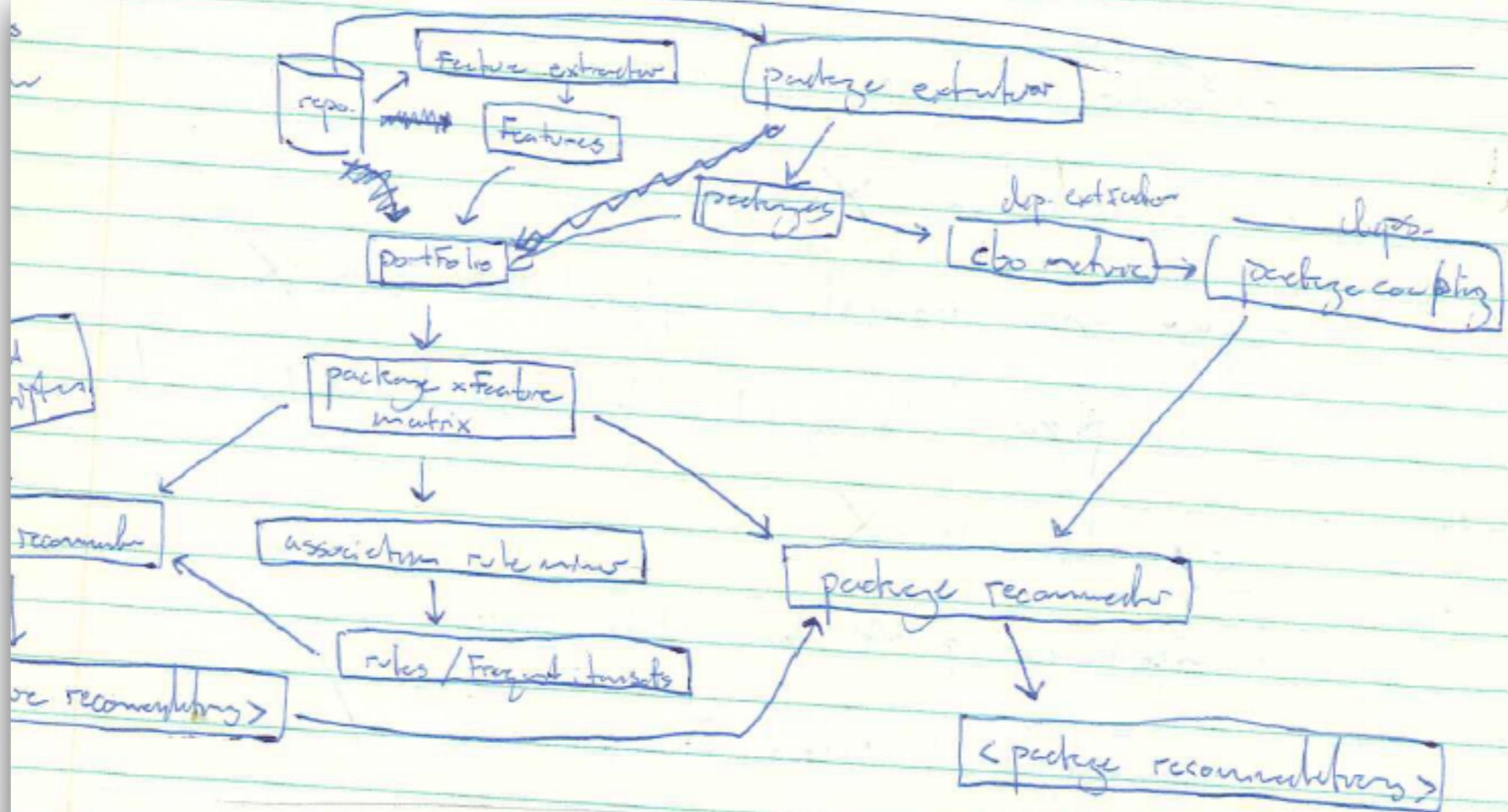
Design

Implementation

Verification

Maintenance





Study Existing Solutions

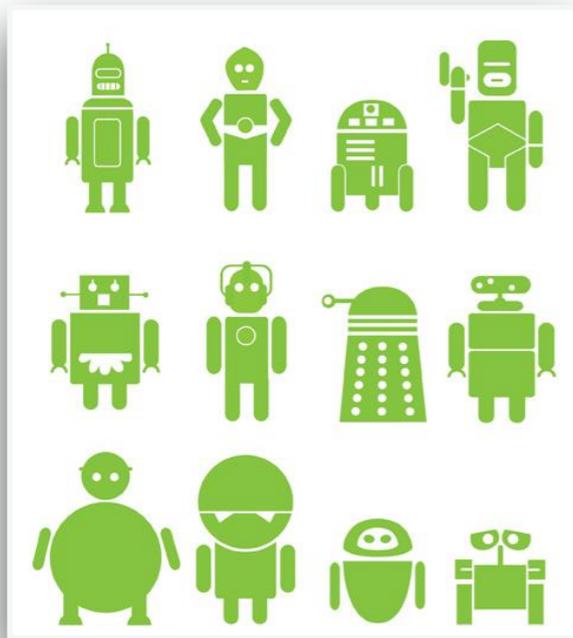
Feature 1
“browser tabs”



Feature 2
“search bar”

Feature 3
“image downloading”

Importance of Detecting Similar Apps



Plagiarism/clones

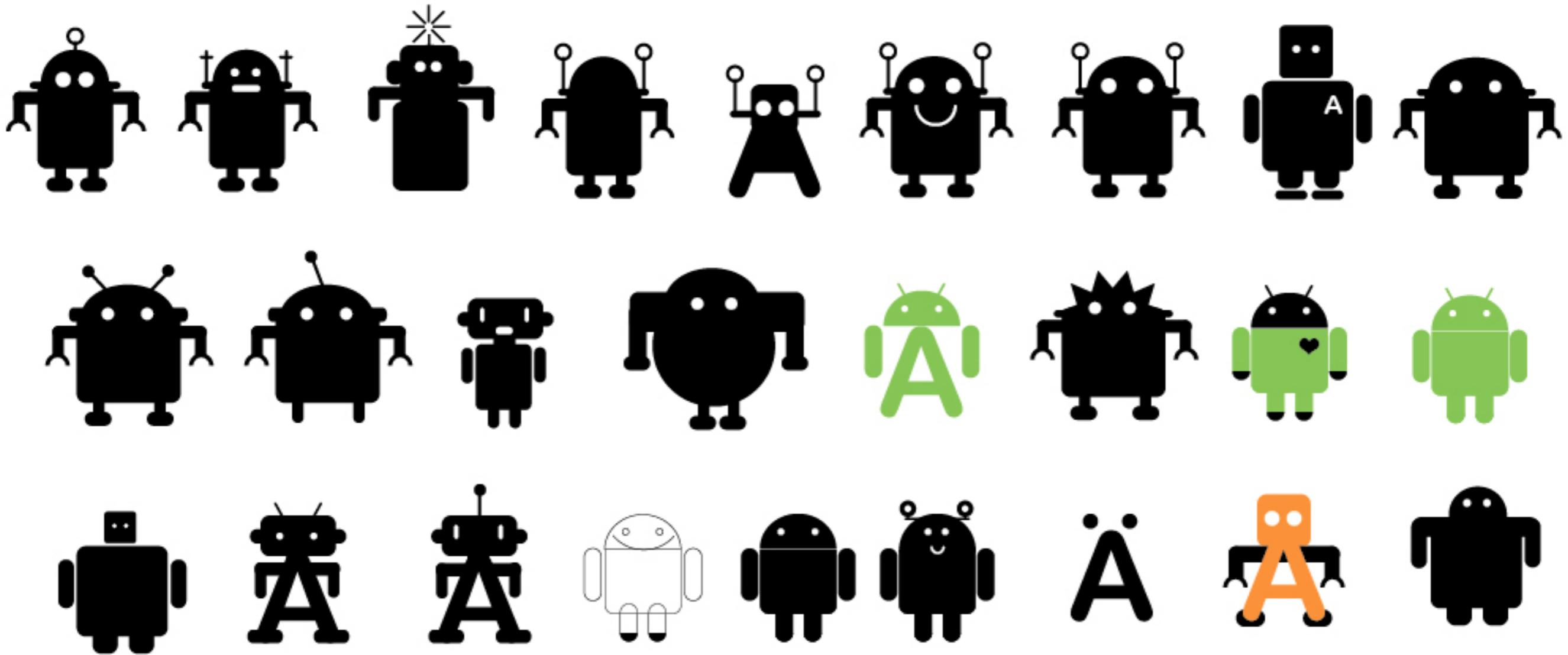


Salient features
in successful apps



API usages

Detecting Similar Apps is Challenging



1.9M apps - Google Play

50B downloads - Google Play

23 releases (Android) since 2008

My apps

Shop

Games

Family

Editors' Choice



Account

My Play activity

My wishlist

Redeem

Send gift

Add credit

Parent Guide



Angry Birds

Top Developer

Rovio Entertainment Ltd. Arcade

★★★★★ 4,775,901

Everyone

Offers in-app purchases

This app is compatible with your device.

Add to Wishlist

Install



Similar

See more



Angry Birds Rio

Rovio Entertainment Ltd

The most exciting of avian adventures continues!

★★★★★ FREE



Subway Surfers

Kiloo

Help Jake, Tricky & Fresh escape from the grumpy Inspector and his dog!

★★★★★ FREE



Fruit Ninja Free

Halfbrick Studios

Be the ultimate bringer of sweet, tasty destruction with every slash!

★★★★★ FREE

Similar to Angry Birds



Angry Birds Rio
Rovio Entertainment Ltd

★★★★★ FREE



Subway Surfers
Kiloo

★★★★★ FREE



Fruit Ninja Free
Halfbrick Studios

★★★★★ FREE



Angry Birds Star Wars
Rovio Entertainment Ltd

★★★★★ FREE



Zombie Tsunami
Mobigame S.A.R.L.

★★★★★ FREE



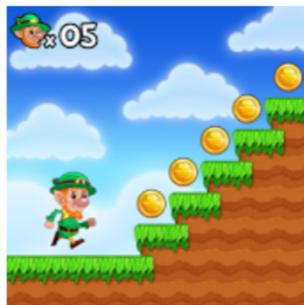
Angry Birds Space
Rovio Entertainment Ltd

★★★★★ FREE



Angry Birds Transformers
Rovio Entertainment Ltd

★★★★★ FREE



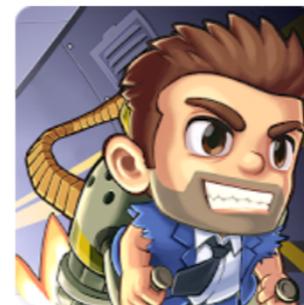
Lep's World 2
nerByte GmbH

★★★★★ FREE



Piano Tiles 2 (Don't Clean)
Master Games

★★★★★ FREE



Jetpack Joyride
Halfbrick Studios

★★★★★ FREE



Hungry Shark Evolution
Future Games of London

★★★★★ FREE



Sonic Dash
SEGA

★★★★★ FREE



JOIN THE ANGRY BIRDS IN THEIR BIGGEST ADVENTURE YET!

A long time ago in a galaxy far, far away... a group of desperate rebel birds faced off against a galactic menace: the Empire's evil Pigtroopers!

Rebel birds, striking from a hidden base, have won their first victory against the evil Imperial Pigs. During the battle, Rebel spies managed to steal secret plans to the Empire's ultimate weapon, the PIG STAR, and are racing to deliver the plans to the Rebel Birds. Now they need your help!

Join an epic adventure with the Angry Birds in the legendary Star Wars™ universe! Use the Force, wield your lightsaber, and blast away Pigtroopers on an intergalactic journey from the deserts of Tatooine to the depths of the Pig Star -- where you'll face off against the terrifying Darth Vader, Dark Lord of the Pigs! Can you become a Jedi Master and restore freedom to the galaxy?

Time to grab your lightsaber and join the adventure! May the birds be with you!

HOURS AND HOURS OF ENGAGING GAMEPLAY - Explore more than 200 levels in iconic locations like Tatooine and the Pig Star. Can you dodge Imperial pigs, laser turrets, Tusken Raider pigs, and the dark side of the Force to get all three stars?

NEW GAMEPLAY MECHANICS - Use lightsabers, Blasters and Jedi powers to wreak havoc on the Imperial Pigs!

LEVEL UP YOUR BIRDS - Keep playing and level up your birds to improve their skills!

SECRETS AND HIDDEN GOODIES - Can you unlock all the R2-D2 and C-3PO bonus levels?

THE MIGHTY FALCON

Stuck on a tricky level? Earn stars and call the Mighty Falcon to rain down the destruction. New goals, achievements and gameplay!

Use the unique powers of the Angry Birds to destroy the greedy pigs' defenses. The survival of the Angry Birds is at stake. Dish out revenge on the greedy pigs. Use the unique powers of each bird to destroy the pigs' defenses. Angry Birds features physics-based gameplay and hours of replay value. Each level requires logic.

If you get stuck in the game, you can purchase the Mighty Eagle! Mighty Eagle is a purchase in Angry Birds that gives unlimited use. This phenomenal creature can wreak havoc and smash the pesky pigs into oblivion. There's just one catch: you need one Mighty Eagle to pass a level once per hour. Mighty Eagle also includes all the achievements!

In addition to the Mighty Eagle, Angry Birds now has power-ups! Boost your star levels to unlock secret content! Angry Birds now has the following amazing



CLANDROID

CLANDroid: Key Ideas

Similar apps *share* some *semantic anchors* (e.g., API calls)

Requirements or features are implemented by a *combination* of *different semantic anchors*

Some semantic anchors are *more expressive* in terms of describing *silent features*

Semantic Anchors

APIs

JAVA / ANDROID APPS

Identifiers

Permissions

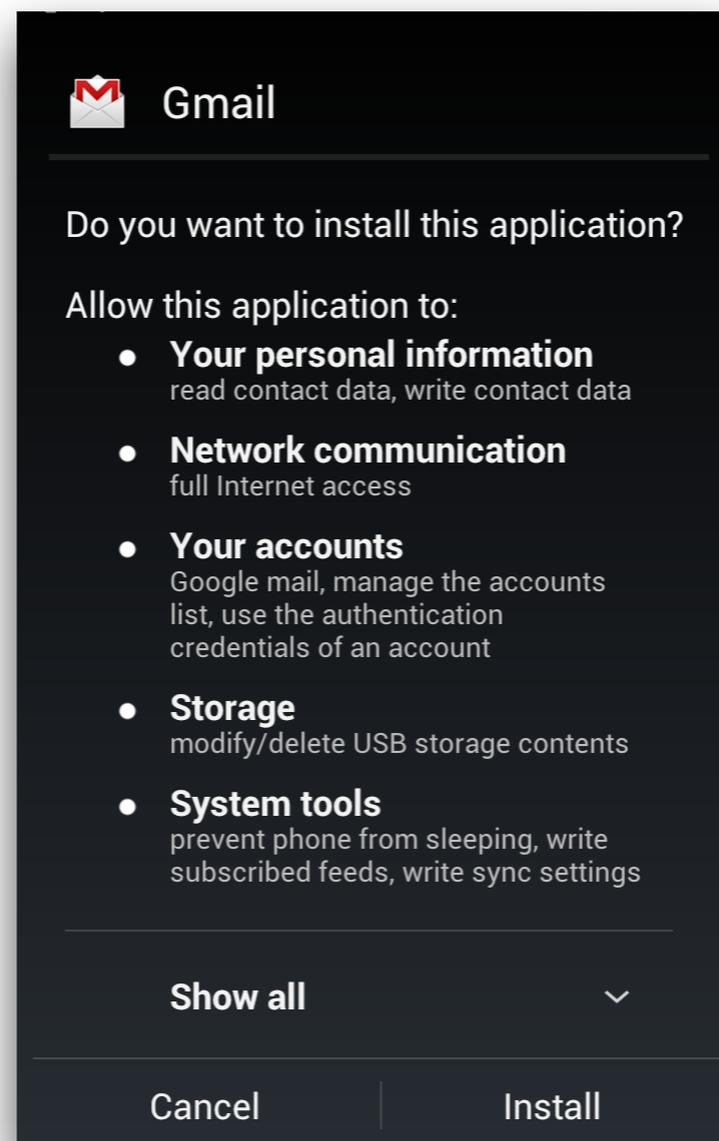
Intents

ANDROID SPECIFIC

Sensors

Permissions

Access to system resources is controlled with install time permissions



Permissions

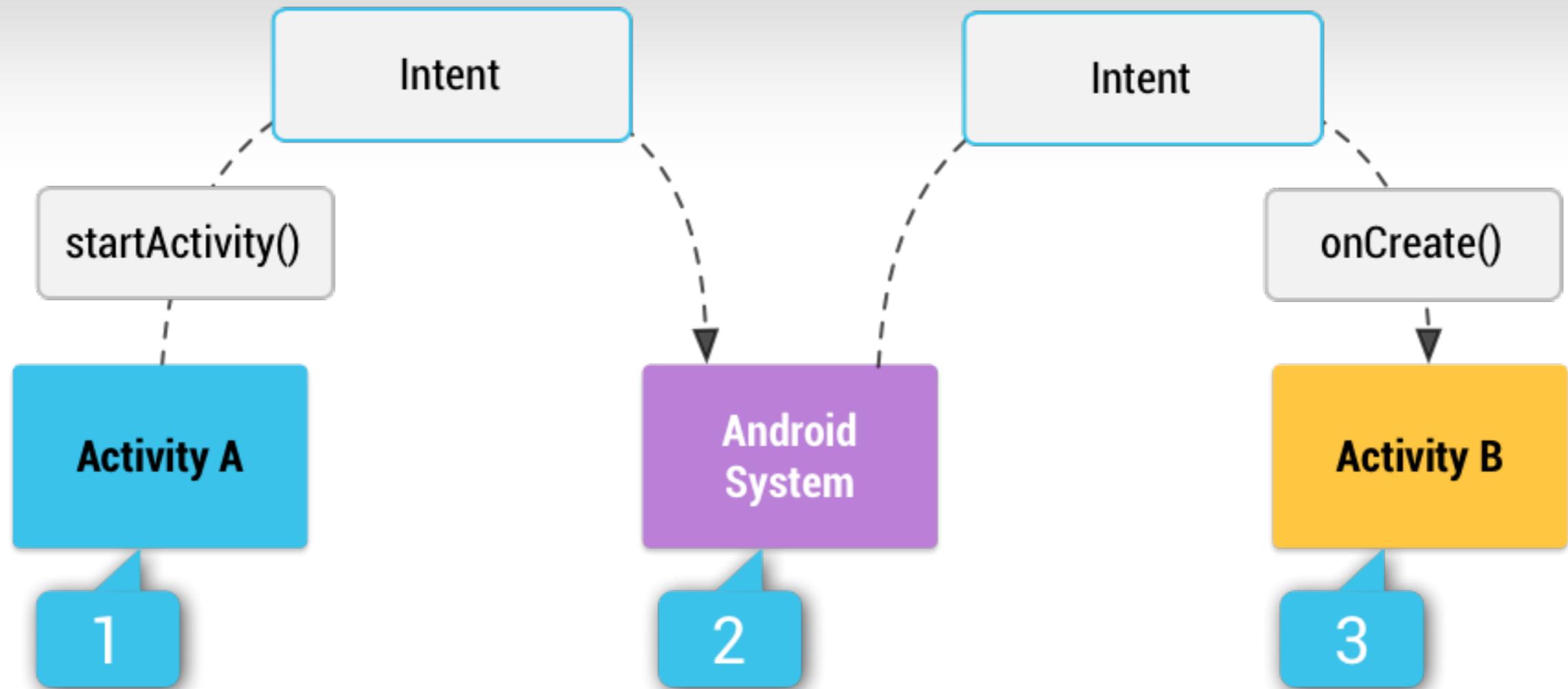
Required for:

- APIs/Features (e.g., internet connection)
- Content providers (e.g.,
READ_CONTENTS)
- Databases
- Message-passing system

Permissions

```
<uses-permission android:name="com.android.vending.CHECK_LICENSE" />  
<uses-permission android:name="android.permission.INTERNET" />  
<uses-permission android:name="android.permission.WAKE_LOCK" />  
<uses-permission android:name="android.permission.ACCESS_NETWORK_STATE" />  
<uses-permission android:name="android.permission.ACCESS_WIFI_STATE"/>  
<uses-permission android:name="android.permission.WRITE_EXTERNAL_STORAGE" />
```

Intents



Intents: asynchronous messages that describe operations to perform

Intents



Explicit intent

```
Intent intent = new Intent(MainActivity.this, OtherActivity.class);  
intent.putExtra("vo", vo);  
intent.putExtra("services_quantity", quantity);
```

....

```
startActivity(intent);
```

Asynchronous call

Sensors

Sensor	Android 4.0 (API Level 14)	Android 2.3 (API Level 9)	Android 2.2 (API Level 8)	Android 1.5 (API Level 3)
TYPE_ACCELEROMETER	Yes	Yes	Yes	Yes
TYPE_AMBIENT_TEMPERATURE	Yes	n/a	n/a	n/a
TYPE_GRAVITY	Yes	Yes	n/a	n/a
TYPE_GYROSCOPE	Yes	Yes	n/a ¹	n/a ¹
TYPE_LIGHT	Yes	Yes	Yes	Yes
TYPE_LINEAR_ACCELERATION	Yes	Yes	n/a	n/a
TYPE_MAGNETIC_FIELD	Yes	Yes	Yes	Yes
TYPE_ORIENTATION	Yes ²	Yes ²	Yes ²	Yes
TYPE_PRESSURE	Yes	Yes	n/a ¹	n/a ¹
TYPE_PROXIMITY	Yes	Yes	Yes	Yes
TYPE_RELATIVE_HUMIDITY	Yes	n/a	n/a	n/a
TYPE_ROTATION_VECTOR	Yes	Yes	n/a	n/a
TYPE_TEMPERATURE	Yes ²	Yes	Yes	Yes

Challenges

Different programmers can use the same words to describe different requirements (i.e., the synonymy problem)

Different programmers can use different words to describe the same requirements (i.e., the polysemy problem)

Latent Semantic Indexing

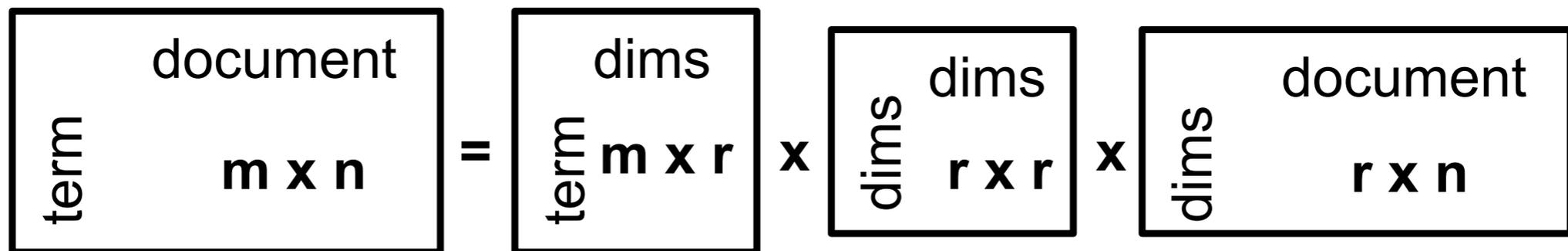
LSI reveals latent concepts

The JDK contains close to 115,000 API calls that are exported by a little more than 13,000 classes and interfaces that are contained in 721 packages

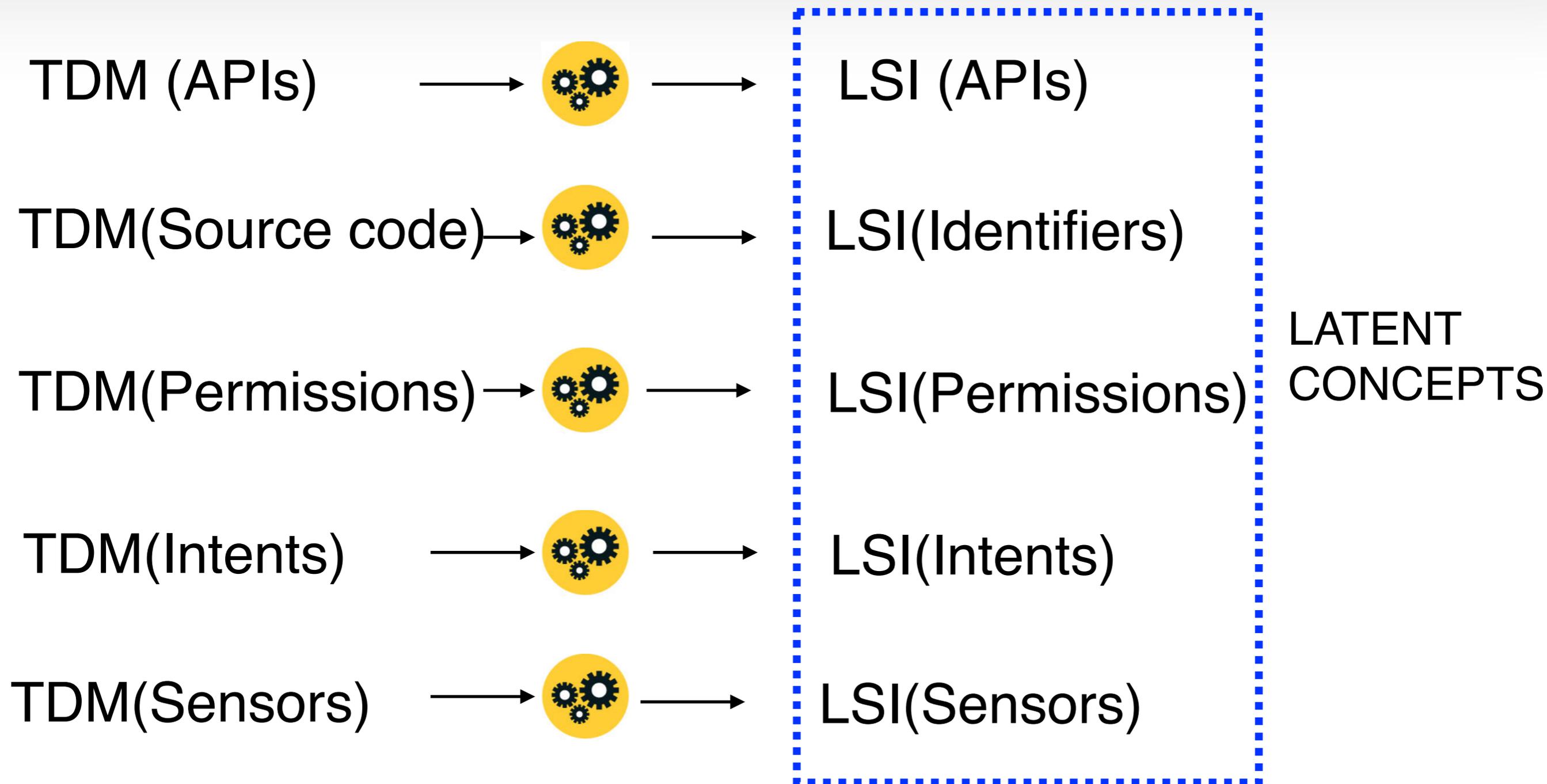
Similarities among terms and words in LSI

Singular Value Decomposition (SVD)

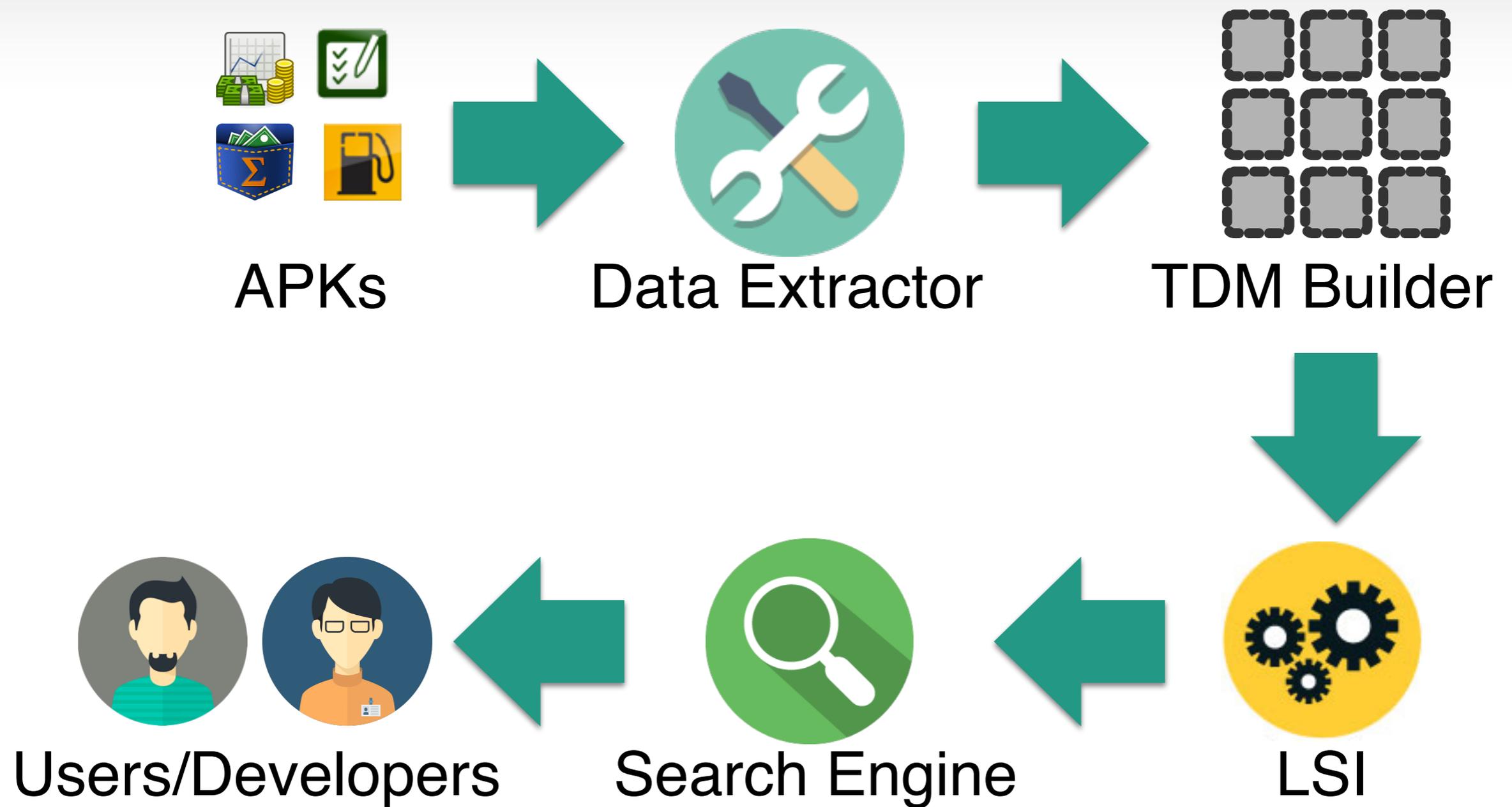
Latent Semantic Indexing



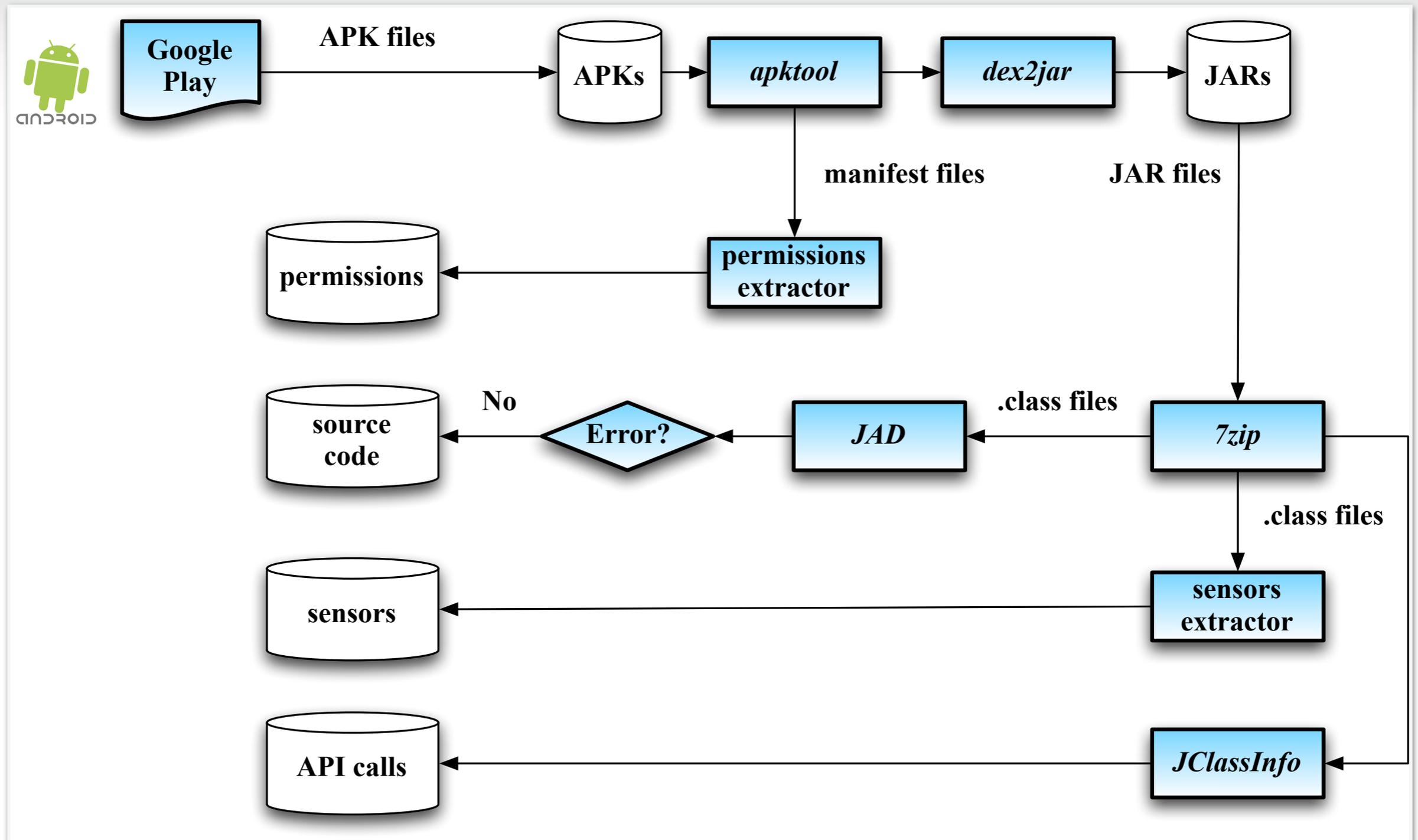
LSI and Semantic Anchors



CLANDroid



Data Extraction





EVALUATION

Research Questions

RQ1: What semantic anchors used in CLANDroid produce better results when compared to the others?

RQ2: How orthogonal are the apps detected by CLANDroid as compared to Google Play?

RQ3: Do third-party libraries and obfuscated apps impact the accuracy of CLANDroid?

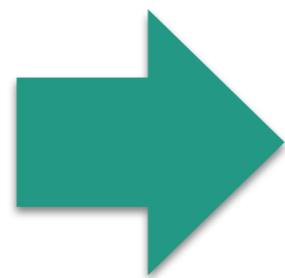
RQ1: Survey



6 versions of the survey
(one for each semantic
anchor + combined)



A survey



12 Questions, 4 apps

1

+

3

Target app

Similar

RQ1: Survey

27

Participants

48

Apps



WILLIAM & MARY

CHARTERED 1693

Intro.

Dear participant,

Thanks for your help and your time. In the following survey you will evaluate whether a set of mobile apps are similar to 12 target Android applications. In order to evaluate to similarity of the applications you will use the following Likert scale:

- 1) Completely dissimilar – The participant is highly confident that the app is dissimilar to the target app.
- 2) Mostly dissimilar – It is unclear if the app is similar to the target app.
- 3) Mostly similar – There are some similarities between the app and the target app.
- 4) Highly similar – The participant is highly confident that the app is similar to the target app.

We encourage the user to read/check the information of the apps available at Google Play (i.e., description, permissions, screenshots, what's new summary, etc) . Please do not base your judgement using only the name of the app or the icon. We want you to judge similarities based on the features you think each app provides to the users.

RQ1: Survey



WILLIAM & MARY

CHARTERED 1693

Intro.

Dear participant,

Thanks for your help and your time. In the following survey you will evaluate whether a set of mobile apps are similar to 12 target Android applications. In order to evaluate to similarity of the applications you will use the following Likert scale:

- 1) Completely dissimilar – The participant is highly confident that the app is dissimilar to the target app.
- 2) Mostly dissimilar – It is unclear if the app is similar to the target app.
- 3) Mostly similar – There are some similarities between the app and the target app.
- 4) Highly similar – The participant is highly confident that the app is similar to the target app.

We encourage the user to read/check the information of the apps available at Google Play (i.e., description, permissions, screenshots, what's new summary, etc) . Please do not base your judgement using only the name of the app or the icon. We want you to judge similarities based on the features you think each app provides to the users.

RQ1: Survey



WILLIAM & MARY

CHARTERED 1693

Q1. Given the following app: `air.BasketballDoodFree` ([description](#))

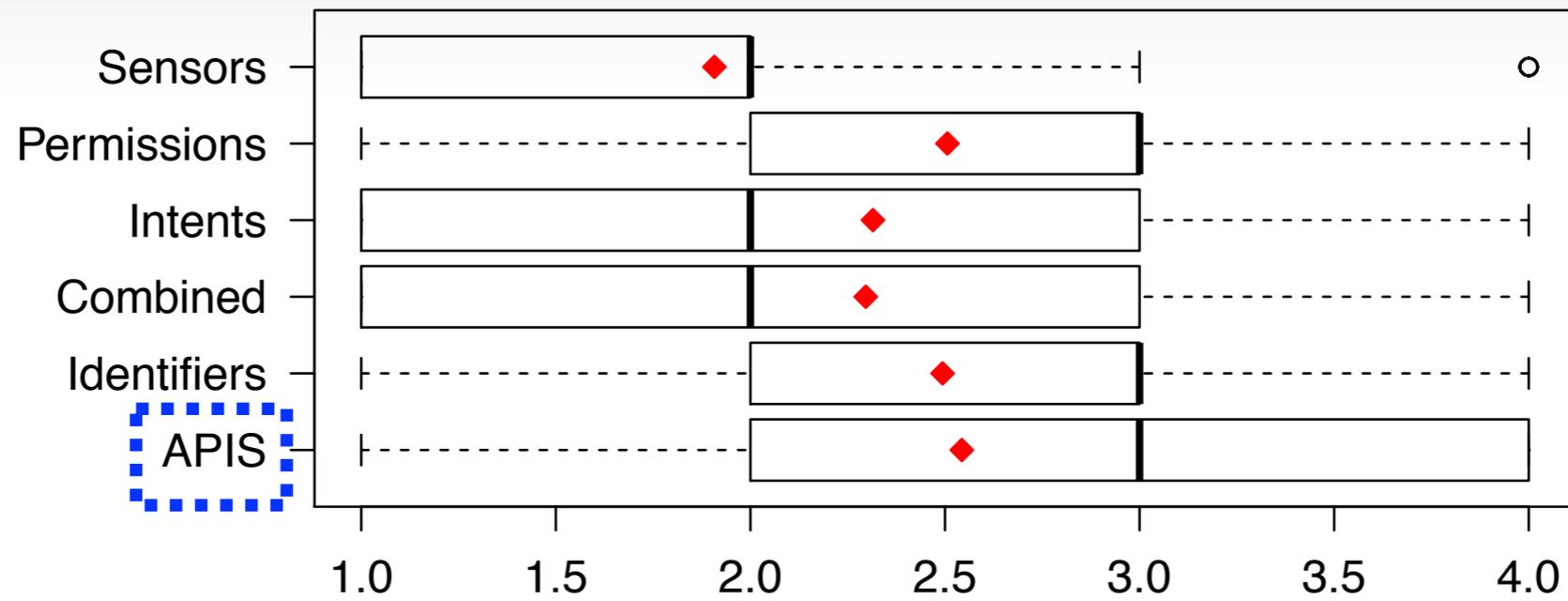
Please rate the similarity of the following set of apps to the app described above:

- **App 1:** `air.SoccerDoodFree` ([description](#))
- **App 2:** `air.RugbyDoodFree` ([description](#))
- **App 3:** `air.FieldHockeyDoodFree` ([description](#))

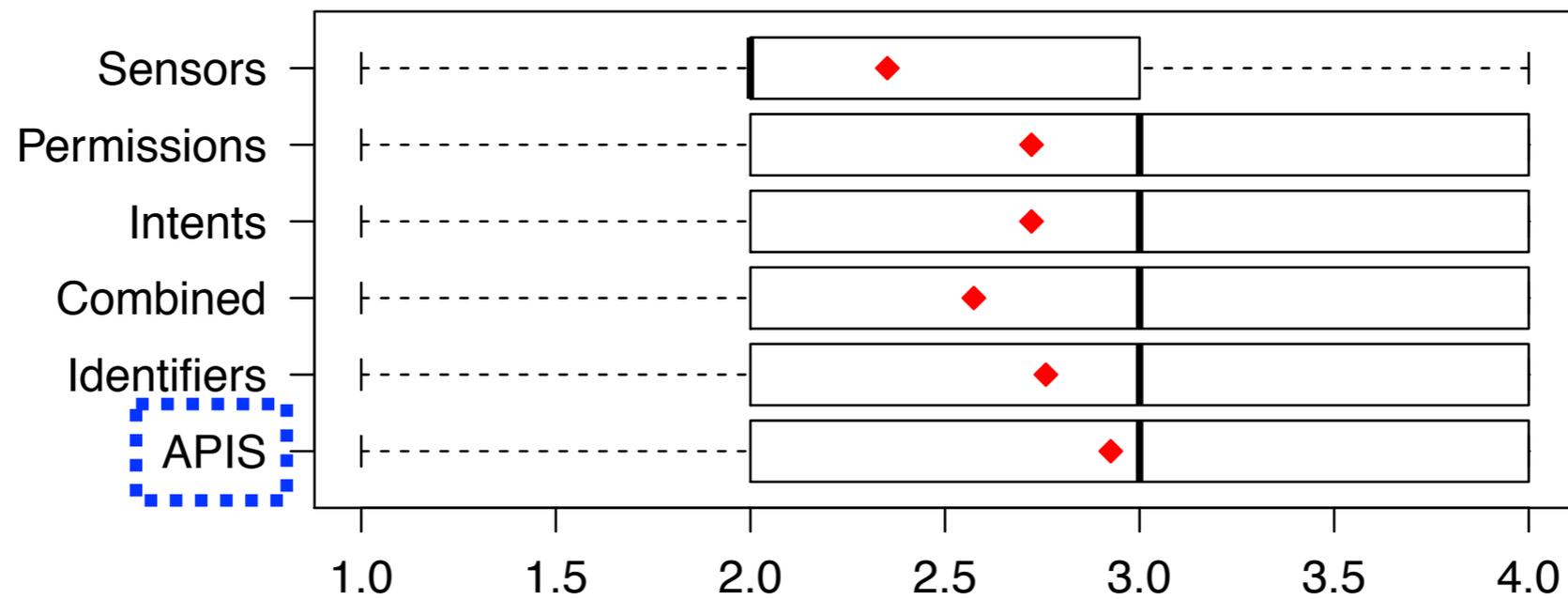
	Completely dissimilar	Mostly dissimilar	Mostly similar	Highly Similar
App 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
App 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
App 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

RQ1: Results

a) Similarity of Top-3 Similar Apps



b) Similarity of Top-1 Similar App



Target App Example



Race, Stunt, Fight, Lite!

Adrenaline Crew Racing

★★★★★ 14,144

Unrated

Offers in-app purchases

 This app is compatible with your device.

 Add to Wishlist

Install



Similar

See more



SouzaSim - Moped

Euangelos Oganosov

SouzaSim is a wheelie Simulator, Yeap! thats right!

★★★★★

FREE



Race, Stunt, Fight, Lite!

Adrenaline Crew

Race stunt and fight your way across the all new motorcycle racing game!

★★★★★

FREE



Traffic Rider

Soner Kara

The Next-Gen of Endless Motorbike Racing

★★★★★

FREE

Top-1 Similar App (CLANDroid)



Crazy Monster Truck - Escape

GAMEANAX Racing

★★★★★ 43,852

E Everyone

Offers in-app purchases

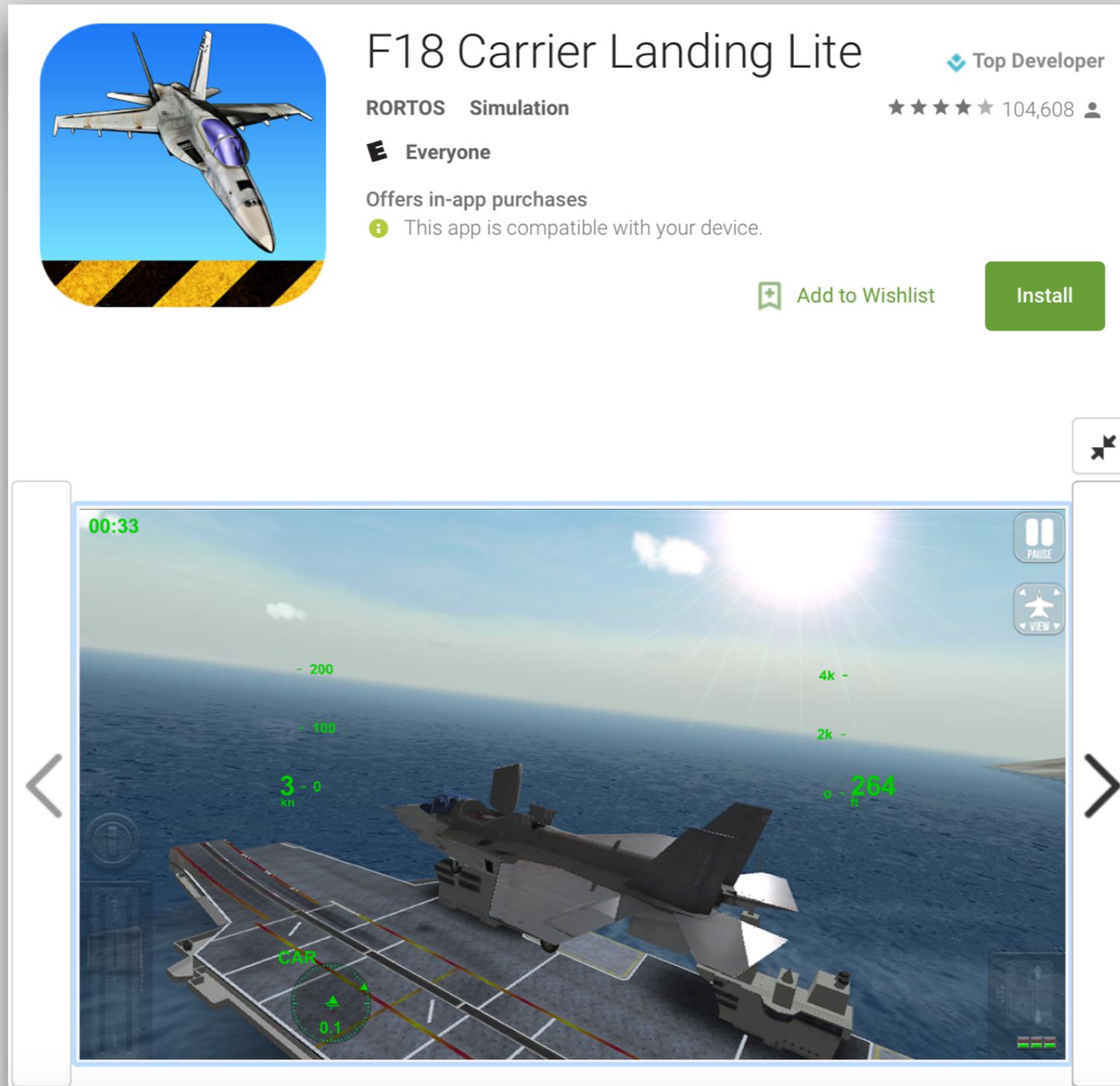
i This app is compatible with your device.

 Add to Wishlist

Install



Top-2 Similar App (CLANDroid)



F18 Carrier Landing Lite Top Developer

RORTOS Simulation ★★★★★ 104,608

E Everyone

Offers in-app purchases

i This app is compatible with your device.

Add to Wishlist Install

00:33

PAUSE

VIEW

- 200

- 100

3 - 0

4k -

2k -

264

0.1

The screenshot shows the Google Play Store listing for the app 'F18 Carrier Landing Lite' by RORTOS. The app is categorized as a Simulation and is rated 'Everyone'. It has a 5-star rating from 104,608 reviews and is marked as a 'Top Developer'. The app offers in-app purchases and is compatible with the user's device. Below the listing is a video preview of the app, showing an F18 fighter jet on a carrier deck. The video includes a timer at 00:33, a pause button, and a view button. The video also displays various HUD elements such as altitude (200, 100, 3-0, 4k, 2k), speed (264), and a heading indicator (0.1).

Top-3 Similar App (CLANDroid)



Cat Escape

Ignacio Sastre Racing

★★★★★ 5,667

E Everyone

i This app is compatible with your device.

 Add to Wishlist

Install



RQ1: Summary

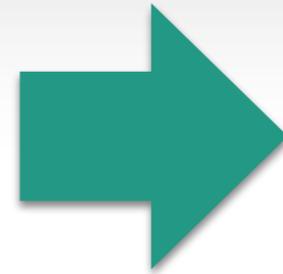
CLANDroid is an *effective tool* for detecting similar apps, in particular when using *APIs, identifiers, permissions, and intents* as semantic anchors

Using *sensors* for detecting similar apps appears to be *ineffective*

APIs provided the *highest number* of apps rated as “*highly similar*”

RQ2: Ranking of Goldset Apps

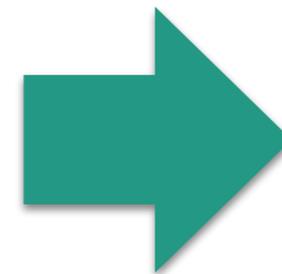
14K+



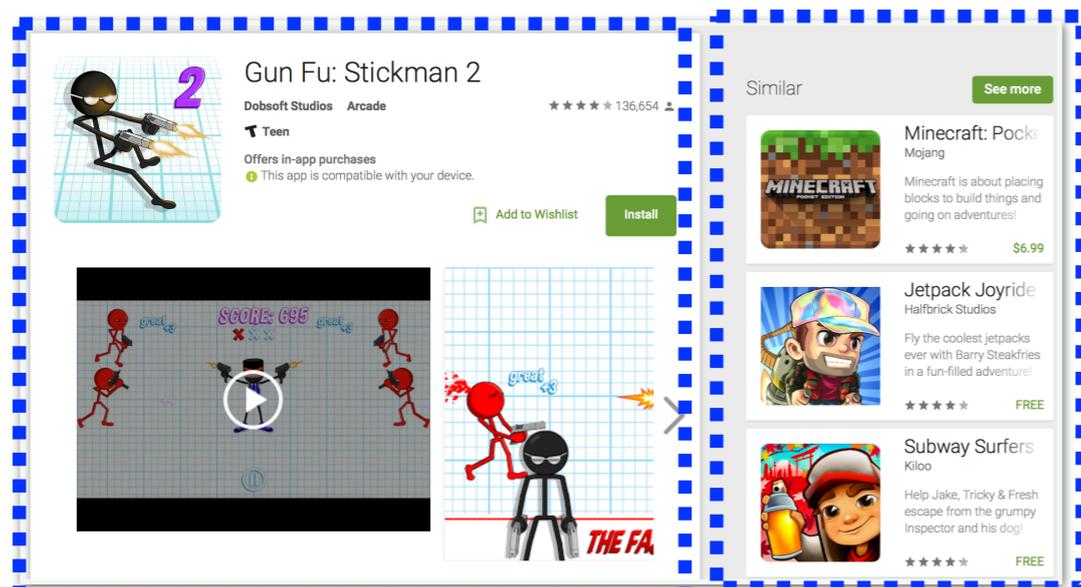
Target



Rankings using
6 CLANDroid instances



Rankings of
goldset apps



Target App Goldset

RQ2: Summary

Mann-Whitney tests show *statistically significant differences* in all the cases except for the comparison between *Identifiers* and *Permissions*, and *Intents* and *Combined*

Magnitude of the *differences* (Cliff's delta) is *small* or *negligible*

CLANdroid is able to retrieve *similar apps belonging to different categories*, while Google Play lists as similar only apps in the same category

RQ3: Results -TPL



The rankings of
goldset apps
significantly improved

~490

Positions

RQ3: Examples



Night Vision Camera

Fingersoft Photography

Unrated

 This app is compatible with all of your devices.

 Top Developer

★★★★★ 93,635 

 Add to Wishlist

Install



LiveKey™ Camera

Sony Mobile Communications Tools

Unrated

 This app is compatible with all of your devices.

★★★★★ 4,670 

 Add to Wishlist

Install

CLANDroid API with and without TPL: 4,928 vs. 50

CLANDroid Ident with and without TPL: 6,852 vs. 266

RQ3: Summary

CLANDroid is significantly (negatively) impacted by the inclusion of *third-party libraries*

There are differences in the rankings when *excluding* applications we detected as *obfuscated*, however, the magnitude of these differences is *negligible* in most cases

CLANDroid (online)



[What is CLANDroid?](#)

app package name

Same Category: All Categories:

- Whole Dataset
- Dataset Excluding Third-party Libraries
- Dataset Excluding Obfuscated Apps



Same Category: All Categories:

- Whole Dataset
- Dataset Excluding Third-party Libraries
- Dataset Excluding Obfuscated Apps

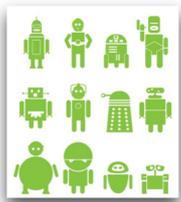
Current query: [com.rovio.angrybirds](#), All categories, whole dataset

APIs	Identifiers	Combined	Intents	Permissions	Sensors
Rank: 1, Similarity: 0.9999450000, Name: com.rovio.angrybirdsspace.ads					
Rank: 2, Similarity: 0.9999417000, Name: com.rovio.amazingalex.trial					
Rank: 3, Similarity: 0.9978808000, Name: com.ea.games.simsfreeplay_na					
Rank: 4, Similarity: 0.9799003000, Name: com.gameloft.android.ANMP.GloftFDHM					
Rank: 5, Similarity: 0.9753396000, Name: com.zynga.scramble					

<http://www.semeru.info/clandroid/>

Summary

Importance of detecting similar apps



Plagiarism/clones



Salient features in successful apps



API usages

*Images from <http://howtowedesign.org/excellent-android-logo-showcase/>

CLANDroid



RQ1: Survey



6 versions of the survey
(one for each semantic anchor + combined)



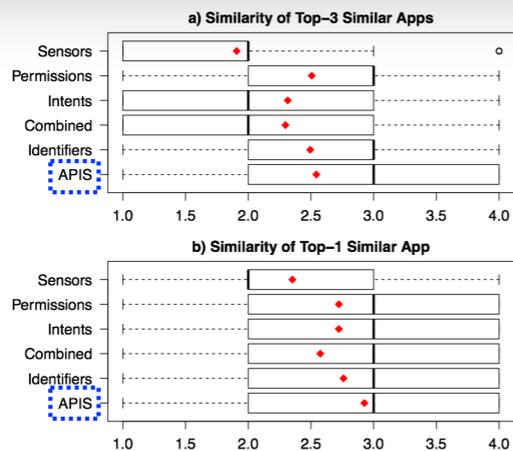
A survey

12 Questions, 4 apps each

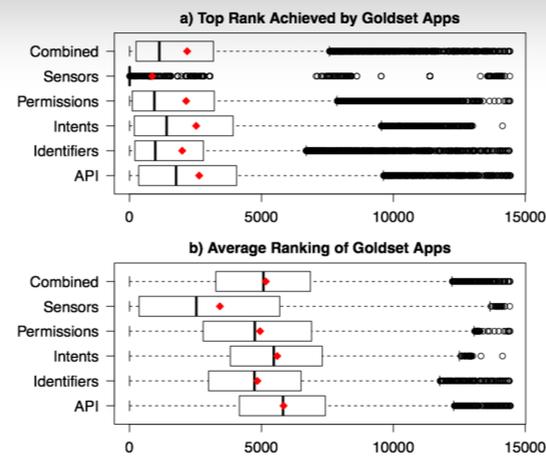


1 + 3
Target app + Similar apps

RQ1: Results



RQ2: Results



RQ3: Results -TPL



Ranking of goldset apps was improved with significant differences

~490 Positions