Mining Apps for Anomalies

Andreas Zeller
Saarland University, Saarbrücken, Germany

Joint work with Alessandra Gorla, Ilaria Tavecchia, Vitalii Avdiienko, Konstantin Kuznetsov, Eric Bodden, Steven Arzt, Siegfried Rasthofer, Isabelle Rommelfanger, and Florian Gross


Saarbrücken

Saarbrücken

1700
BSc + MSc students

375
PhD students

200
Researchers (post PhD)

8
New buildings since 2001

10
ERC Grant holders

6
Leibniz Awardees

4
ACM Fellows

1
Software Engineer

UNIVERSITÄT DES SAARLANDES | mpi max planck institut informatik | Max Planck Institute for Software Systems | DFKI | intel Visual Computing Institute | CISPA Center for Information Security, Privacy and Accountability

Specifications

$$
\begin{array}{l}
\textbf{removeChild} \\
\hline
\Delta XMLElement \\
child? : XML\_ELEMENT \\
\hline
child? \in enumerateChildren \\
child? \neq null \\
enumerateChildren' = enumerateChildren \setminus child? \\
getChildrenCount' = getChildrenCount - 1 \\
\hline
\end{array}
$$

Specifications

removeChild

$\Delta XMLElement$
$child? : XML\_ELEMENT$

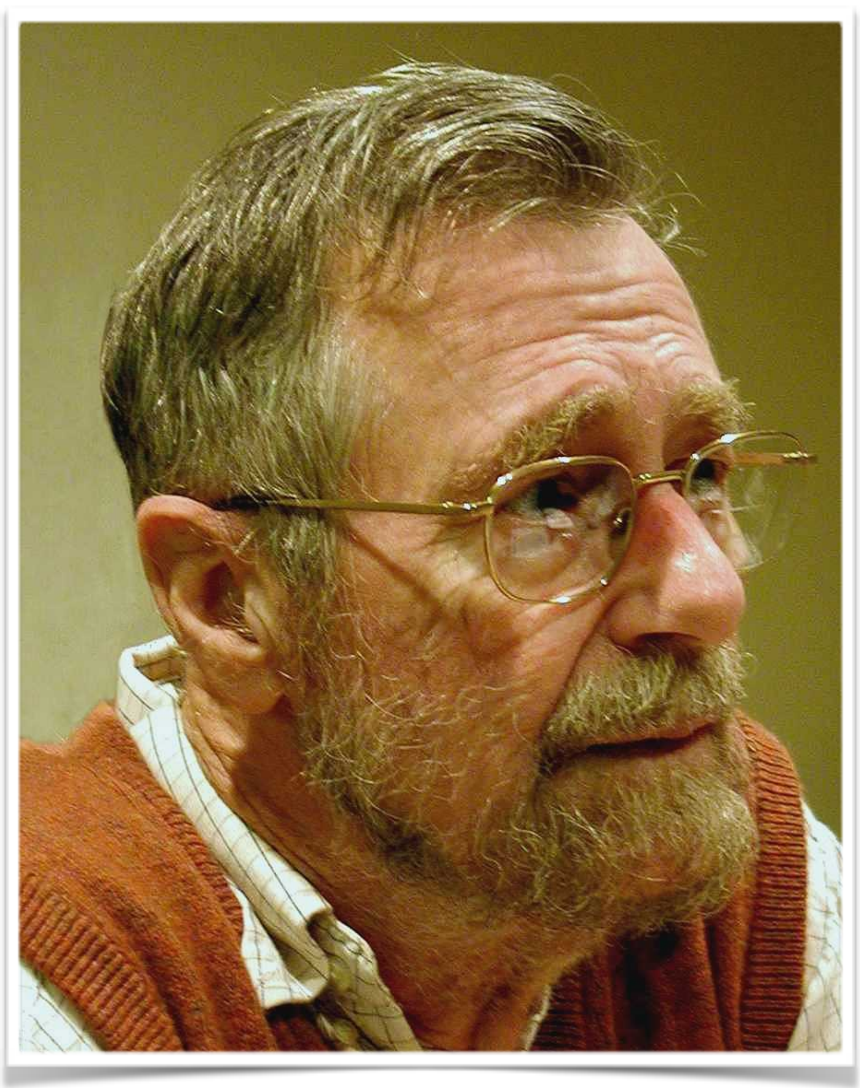
$child? \in enumerateChildren$
$child? \neq null$
$enumerateChildren' = enumerateChildren \setminus child?$
$getChildrenCount' = getChildrenCount - 1$

fully automated testing

fully automated debugging

widely automated verification

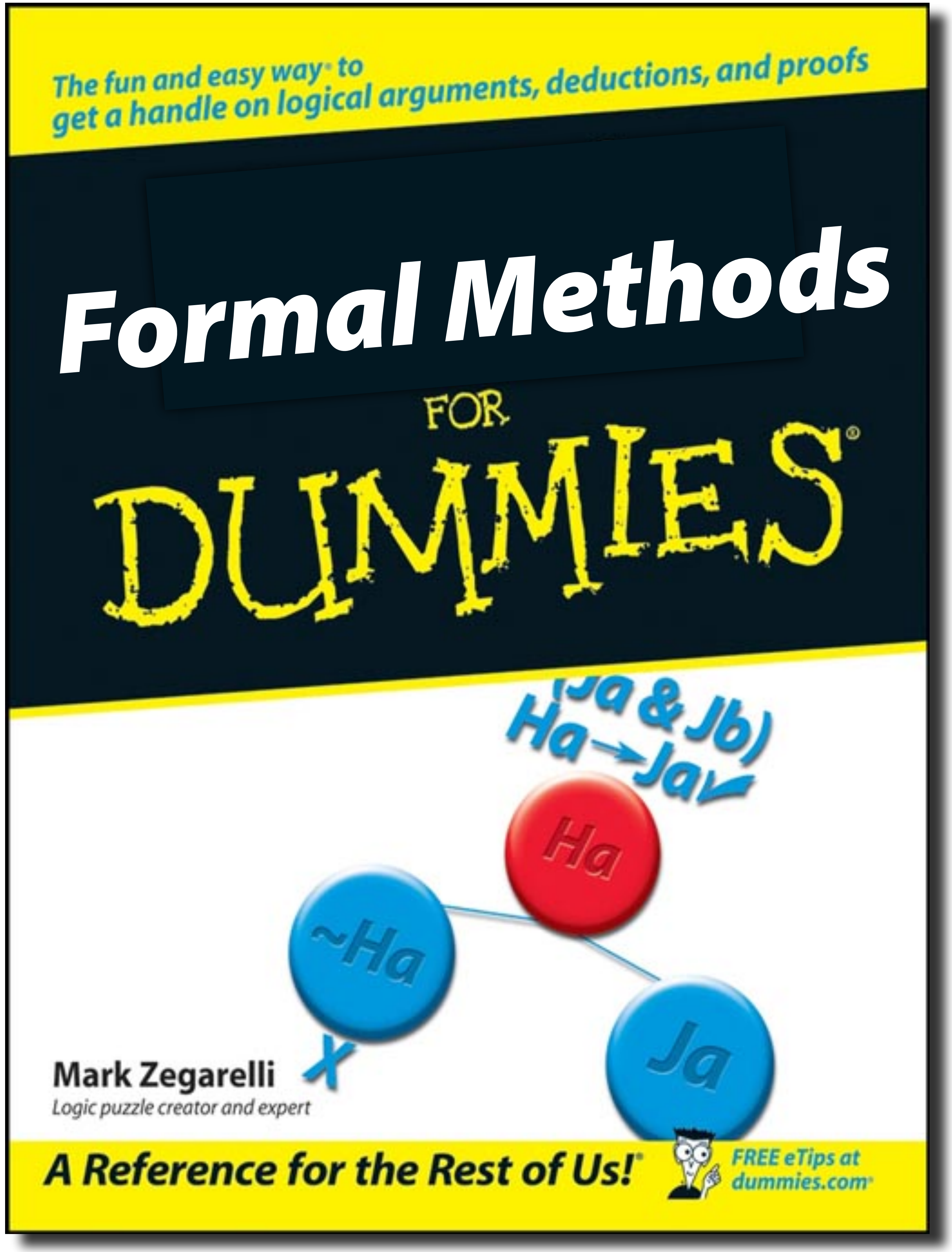
Specifying Correctness



removeChild

$\Delta XMLElement$
$child? : XML\_ELEMENT$

$child? \in enumerateChildren$
$child? \neq null$
$enumerateChildren' = enumerateChildren \setminus child?$
$getChildrenCount' = getChildrenCount - 1$


The fun and easy way® to
get a handle on logical arguments, deductions, and proofs
Formal Methods
FOR
DUMMIES®
Ja & Jb)
Ha → Ja
Ha
~Ha
Ja
Mark Zegarelli
Logic puzzle creator and expert
A Reference for the Rest of Us!®
FREE eTips at
dummies.com®


Microsoft Outlook
Unknown error
OK
Was this information helpful?

Normality


ABC
Aa
Σ
x + y = z
A
W
Abc
0100
1101
1001
First option
Second option

Mining Normality

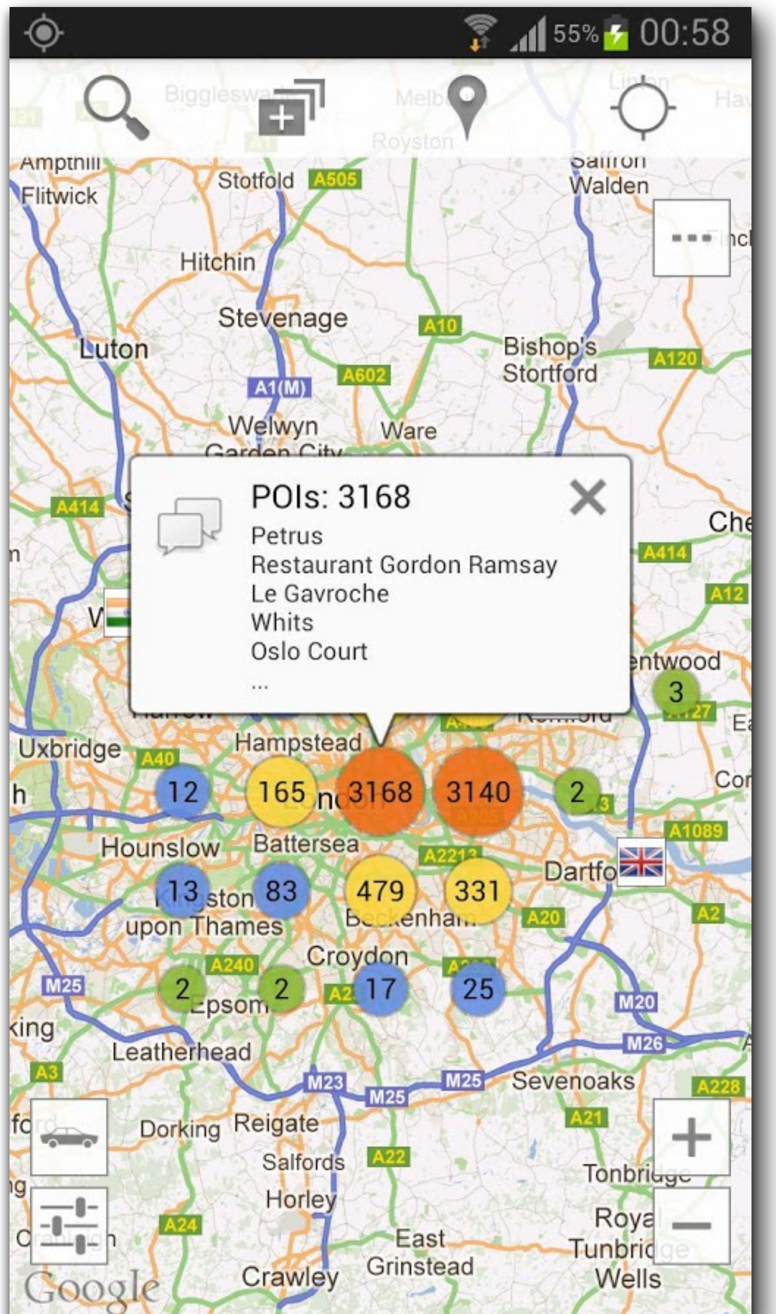
removeChild

$\Delta XMLElement$
$child? : XML\_ELEMENT$

$child? \in enumerateChildren$
$child? \neq null$
$enumerateChildren' = enumerateChildren \setminus child?$
$getChildrenCount' = getChildrenCount - 1$

Outliers

London Restaurants



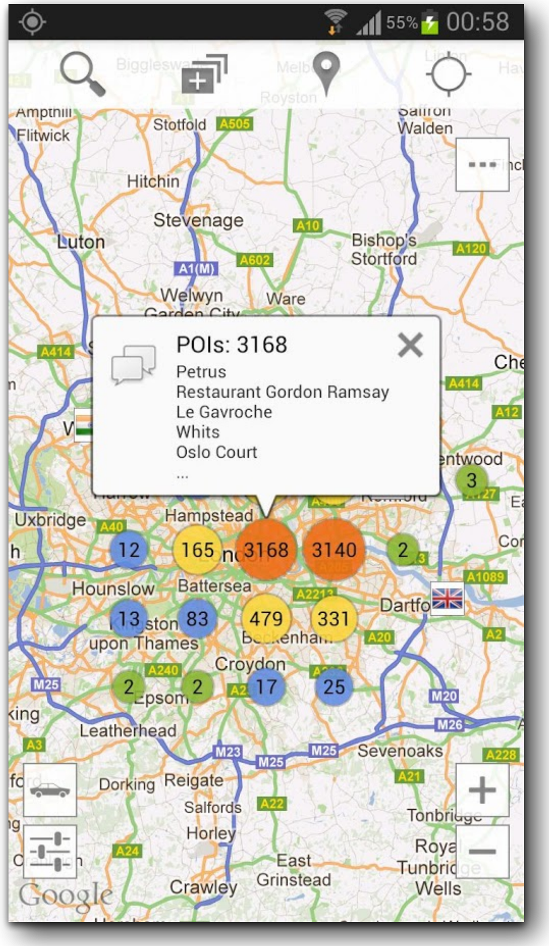
Looking for a restaurant, a bar, a pub or just to have fun in London? Search no more! This application has all the information you need:

- You can search for every type of food you want: french, british, chinese, indian etc.
- You can use it if you are in a car, on a bicycle or walking
- You can view all objectives on the map
- You can search objectives
- You can view objectives near you
- You can view directions (visual route, distance and duration)
- You can use it with Street View
- You can use it with Navigation

Keywords: london, restaurants, bars, pubs, food, breakfast, lunch, dinner, meal, eat, supper, street view, navigation

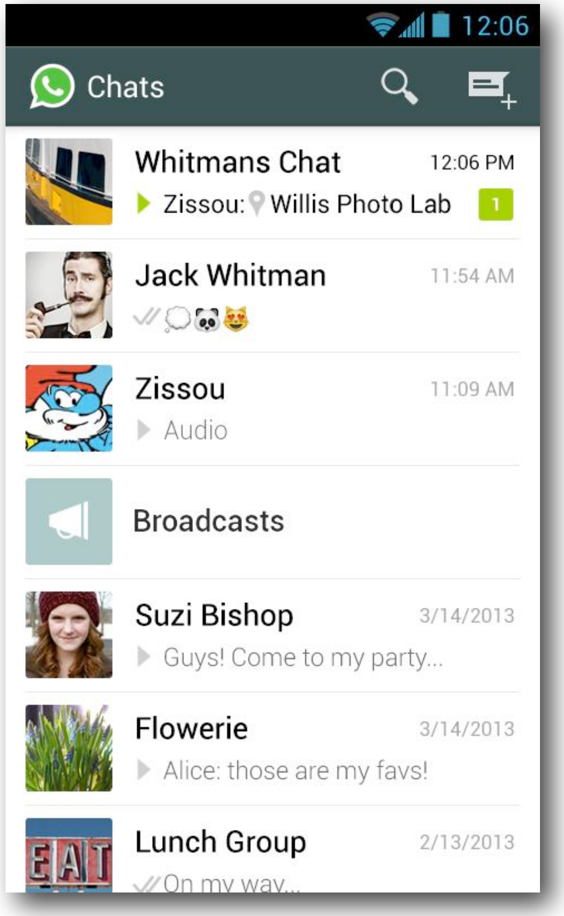
Also sends out *account info*
Also sends out *mobile phone number*
Also sends out *your device ID*

What is malicious?



London Restaurants

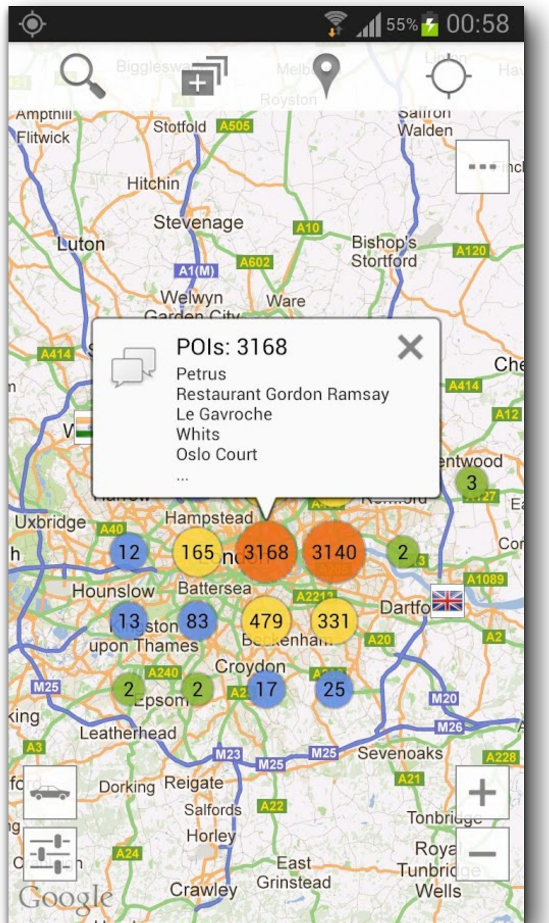
Also sends out account info
Also sends out mobile phone number
Also sends out your device ID



Also sends out account info
Also sends out mobile phone number
Also sends out your device ID

WhatsApp messenger

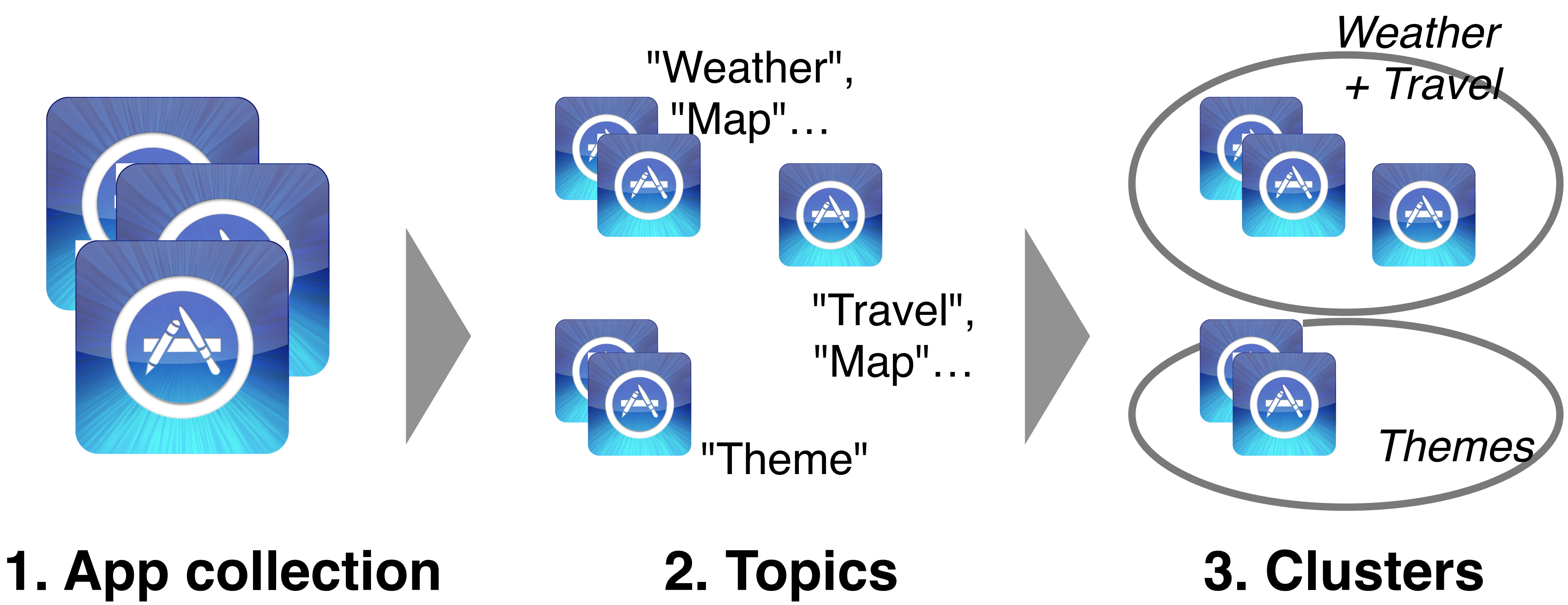
What is normal?



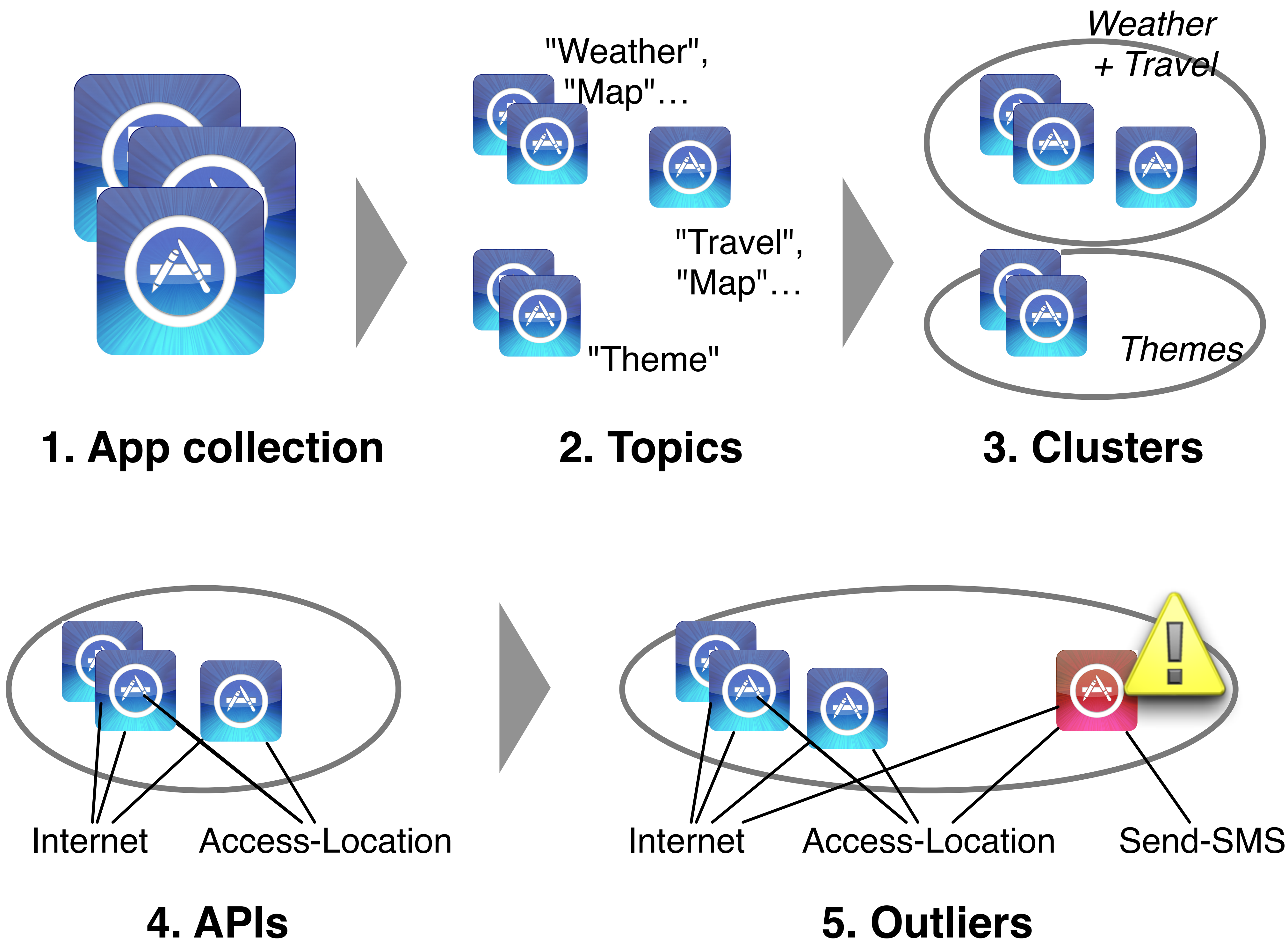
London Restaurants

- “London Restaurants” is a “travel” app
- For “travel” apps, sending account infos is *abnormal*
- For “messaging” apps, this is far more likely

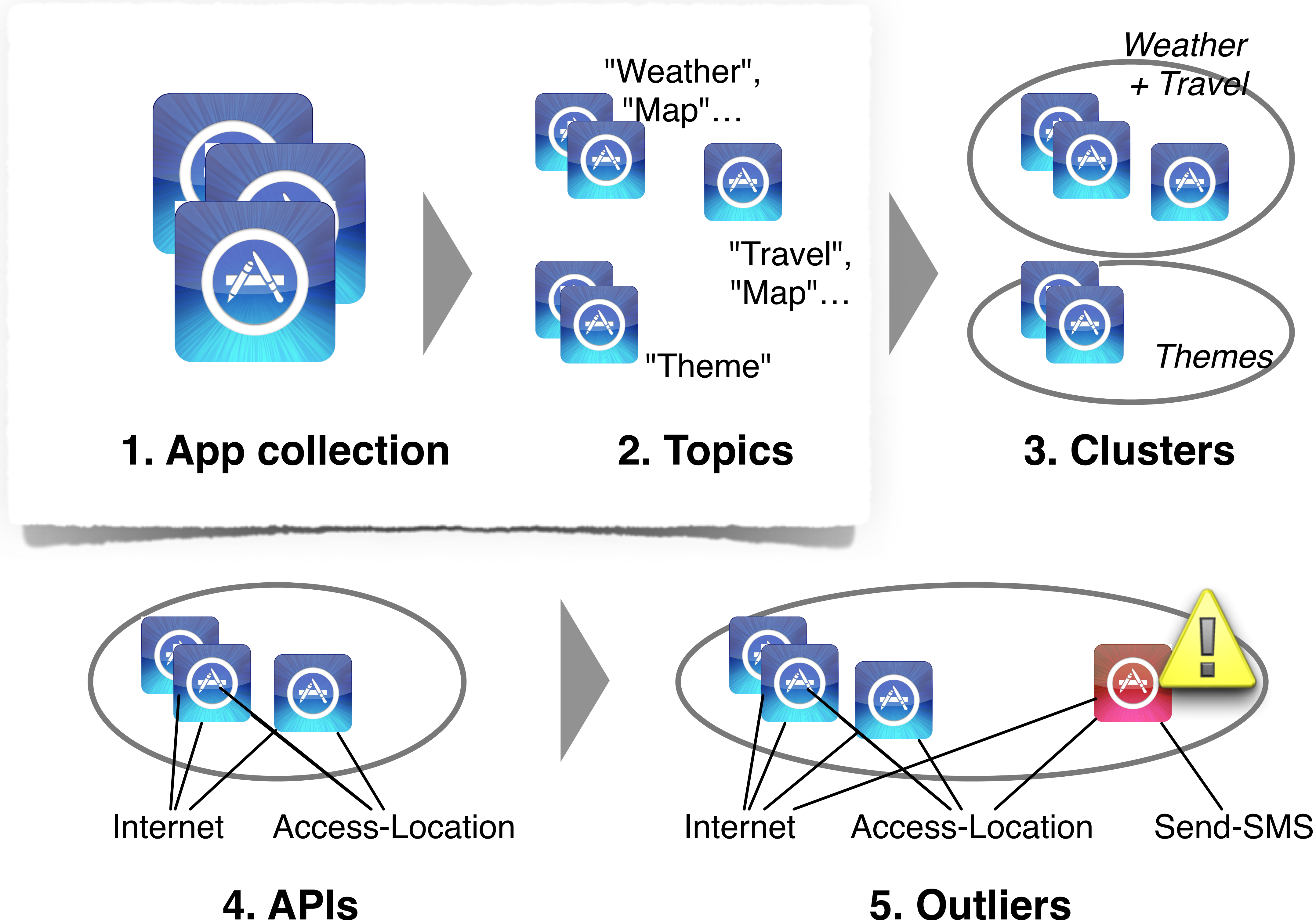
CHABADA


"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters

CHABADA


"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters
Internet
Access-Location
Send-SMS
4. APIs
5. Outliers

CHABADA

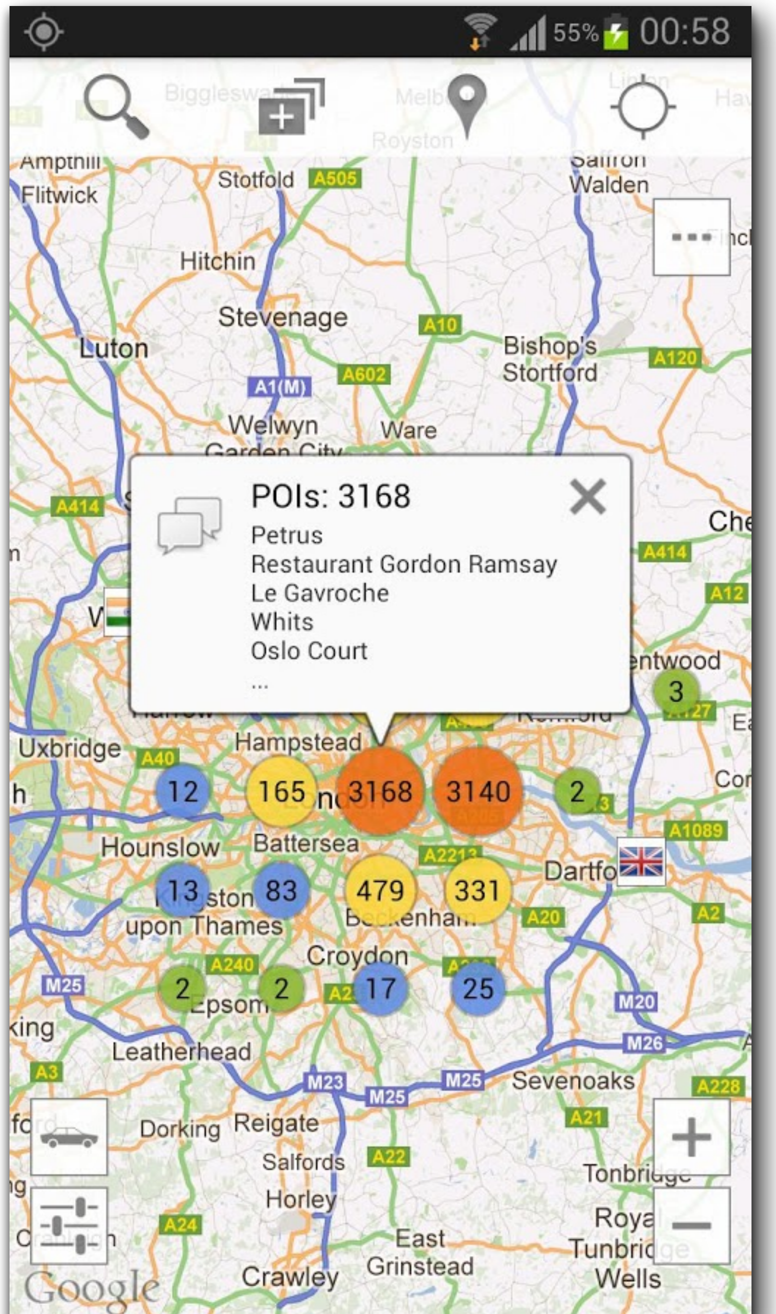


"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters
Internet
Access-Location
Internet
Access-Location
Send-SMS
4. APIs
5. Outliers

App Collection

- Source: Google Play Store
- Downloaded top 150 apps + metadata from each of the 30 categories
- Time frame: Winter to Spring 2013
- Total: 32,136 apps
- Data package available on Web site

Stemming

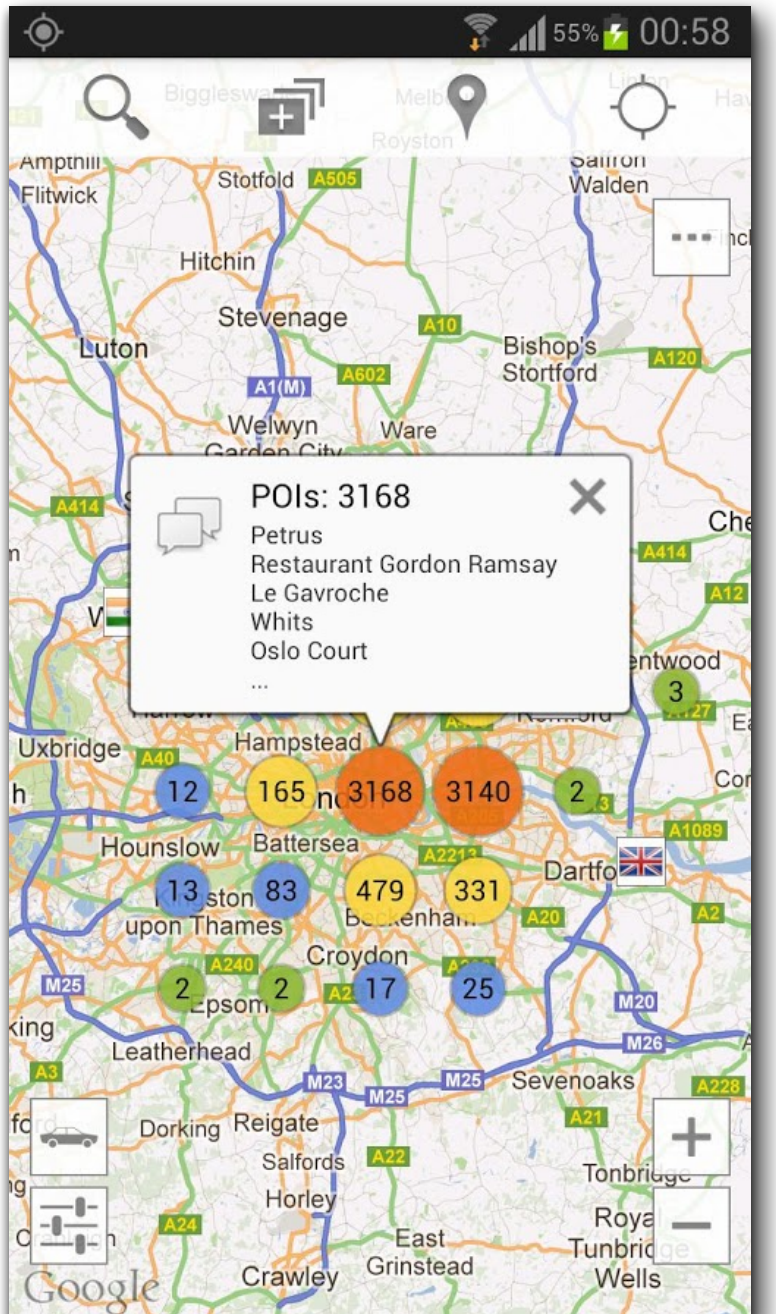


looking for a restaurant, a bar, a pub or just to have fun in london? search no more! this application has all the information you need:

- you can search for every type of food you want: french, british, chinese, indian etc.
- you can use it if you are in a car, on a bicycle or walking
- you can view all objectives on the map
- you can search objectives
- you can view objectives near you
- you can view directions (visual route, distance and duration)
- you can use it with street view
- you can use it with navigation

keywords: london, restaurants, bars, pubs, food, breakfast, lunch, dinner, meal, eat, supper, street view, navigation

Stemming

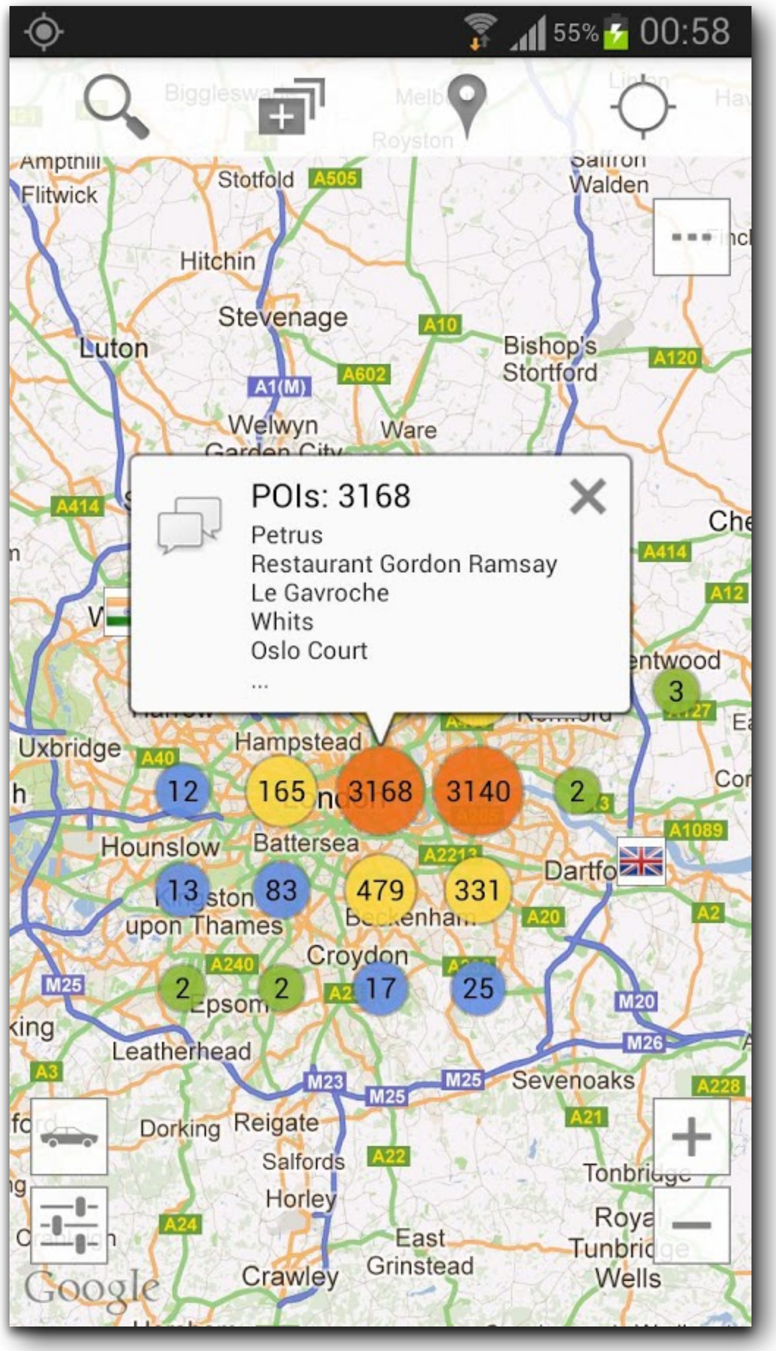


looking for a restaurant, a bar, a pub or just to have fun in london? search no more! this application has all the information you need:

- you can search for everi type of food you want: french, british, chinese, indian etc.
- you can use it if you are in a car, on a bicycle or walking
- you can view all objectives on the map
- you can search objectives
- you can view objectives near you
- you can view directions (visual route, distance and duration)
- you can use it with street view
- you can use it with navigation

keywords: london, restaurants, bars, pubs, food, breakfast, lunch, dinner, meal, eat, supper, street view, navigation

Stemming



look london restaur search bar pub just applic fun

inform can search need everi type food want french

british chines indian etc car bicycl walk

can us can view object map visual rout

can search object search can view distanc

durat can view direct object near

can us street view can us navig

keyword london restaur bar pub food view

breakfast lunch dinner meal eat supper street navig

Topic Analysis

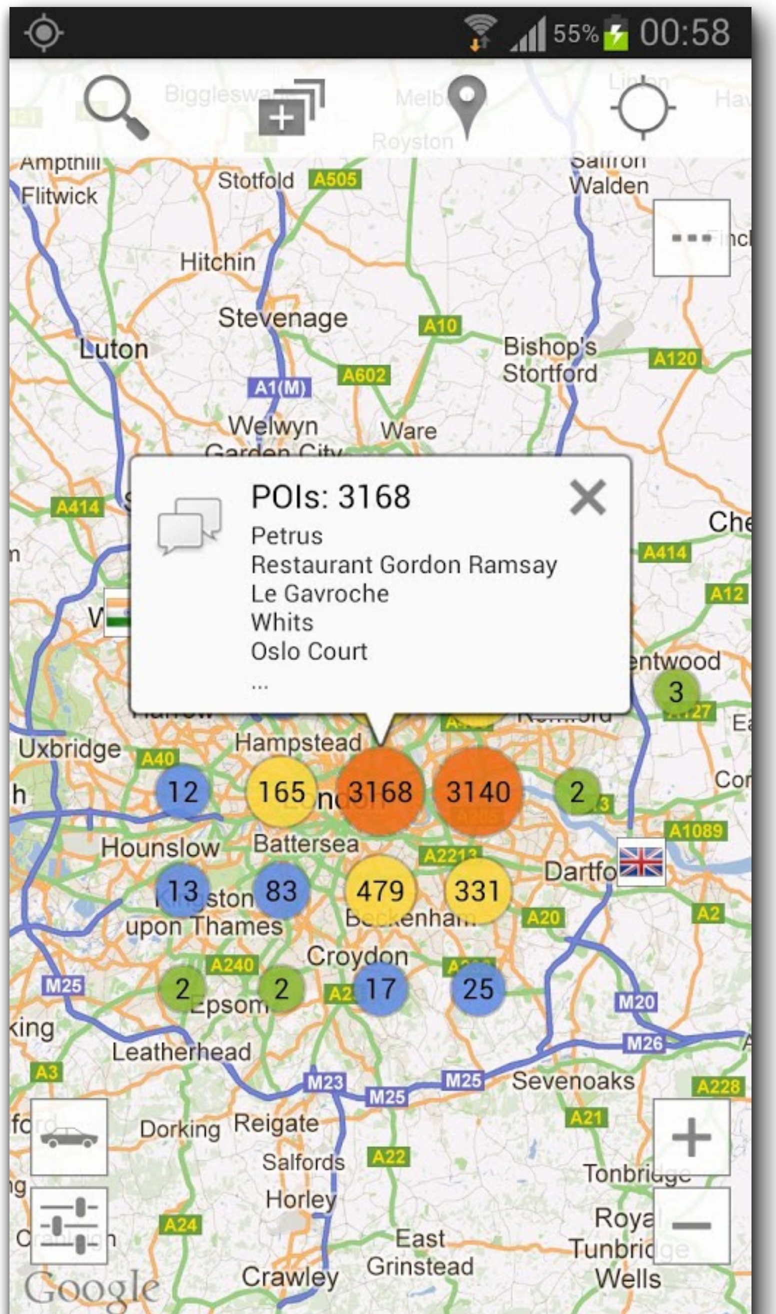
- Eliminated all apps with ≤10 words, now 22,521 apps
- Want to discover the *topics* that occur in a collection of unlabeled text
- A *topic* consists of a cluster of words that frequently occur together
- Used *Latent Dirichlet Allocation* (LDA) to identify 30 topics

Topics

Id	Assigned Name	Most Representative Words (stemmed)
0	“personalize”	galaxi, nexu, device, screen, effect, instal, customis
1	“game and cheat sheets”	game, video, page, cheat, link, tip, trick
2	“money”	slot, machine, money, poker, currenc, market, trade, stock, casino coin, finance
3	“tv”	tv, channel, countri, live, watch, germani, nation, bbc, newspap
4	“music”	music, song, radio, play, player, listen
5	“holidays” and religion	christmas, halloween, santa, year, holiday, islam, god
6	“navigation and travel”	map, inform, track, gps, navig, travel
7	“language”	language, word, english, learn, german, translat
8	“share”	email, ad, support, facebook, share, twitter, rate, suggest
9	“weather and stars”	weather, forecast, locate, temperatur, map, city, light
10	“files and video”	file, download, video, media, support, manage, share, view, search

13	"design and art"	life, peopl, natur, form, feel, learn, art, design, uniqu, effect, modern
14	"food and recipes"	recip, cake, chicken, cook, food
15	"personalize"	theme, launcher, download, install, icon, menu
16	"health"	weight, bodi, exercise, diet, workout, medic
17	"travel"	citi, guid, map, travel, flag, countri, attract
18	"kids and bodies"	kid, anim, color, girl, babi, pictur, fun, draw, design, learn
19	"ringtones and sound"	sound, rington, alarm, notif, music
20	"game"	game, plai, graphic, fun, jump, level, ball, 3d, score
21	"search and browse"	search, icon, delet, bookmark, link, homepag, shortcut, browser
22	"battle games"	story, game, monster, zombi, war, battle
23	"settings and utils"	screen, set, widget, phone, batteri
24	"sports"	team, football, leagu, player, sport, basketbal
25	"wallpapers"	wallpap, live, home, screen, background, menu
26	"connection"	device, connect, network, wifi, blootooth, internet, remot, server
27	"policies and ads"	live, ad, home, applovin, notif, data, polici, privacy, share, airpush, advertis
28	"popular media"	seri, video, film, album, movi, music, award, star, fan, show, gangnam, top, bieber
29	"puzzle and card games"	game, plai, level, puzzl, player, score, challeng, card

London Restaurant Topics



look london restaur search bar pub just applic fun

inform can search need everi type food want french

british chines indian etc car bicycl walk

can us can view object map visual rout

can search object search can view distanc

durat can view direct object near

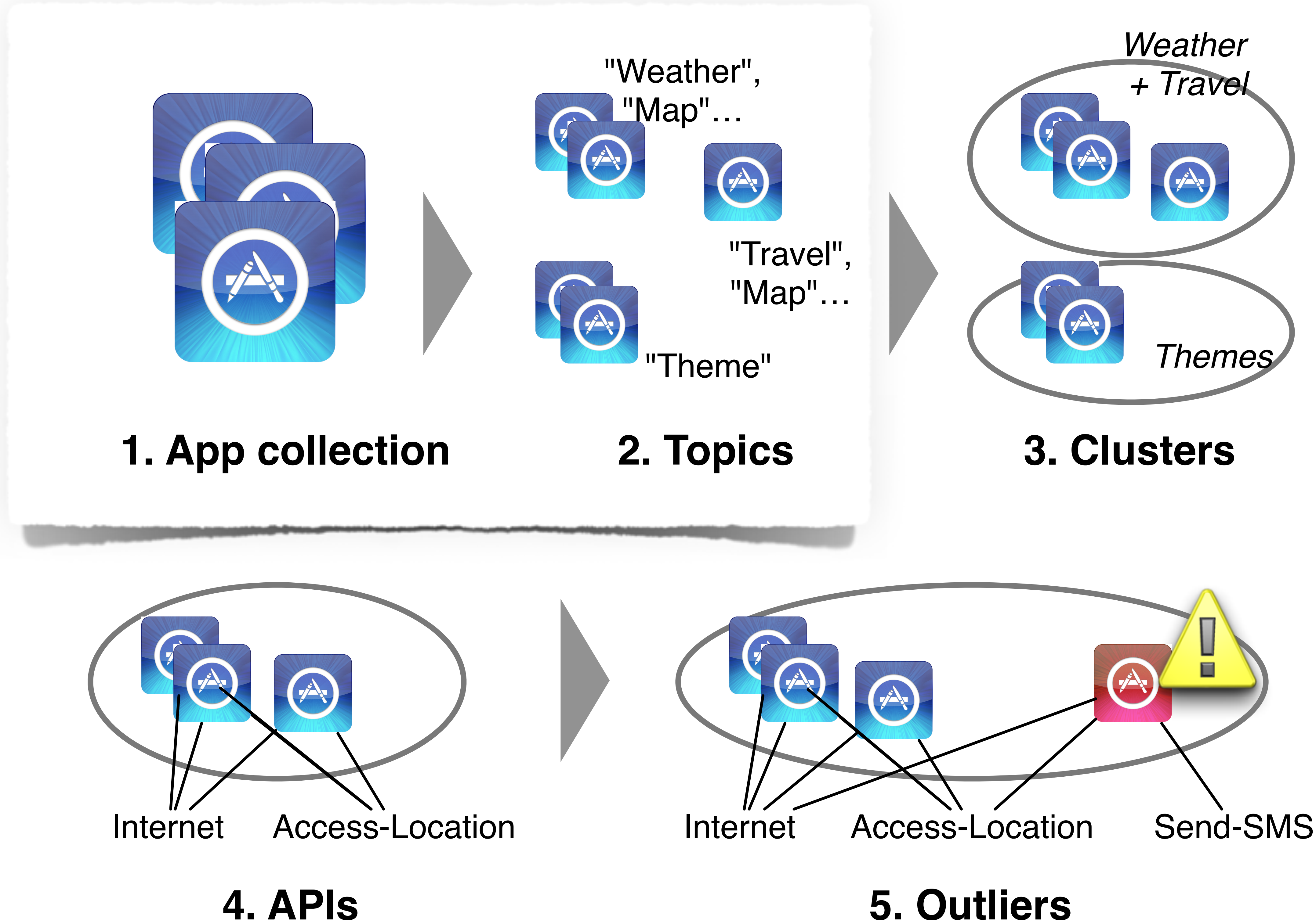
can us street view can us navig

keyword london restaur bar pub food view

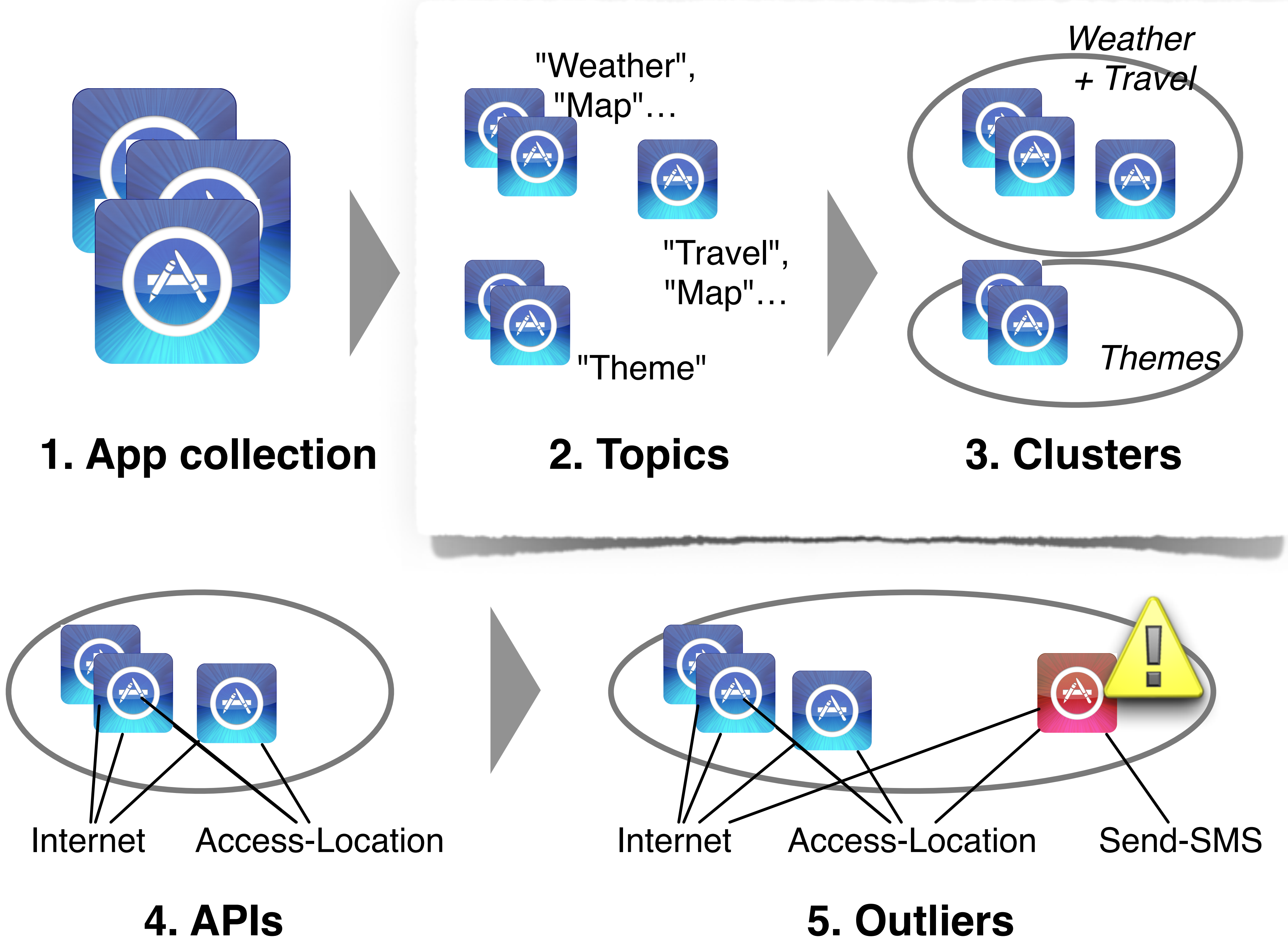
breakfast lunch dinner meal eat supper street navig

“navigation and travel” (59.8%)
“food and recipes” (19.9%)
“travel” (14.0%)

CHABADA


"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters
Internet
Access-Location
Internet
Access-Location
Send-SMS
4. APIs
5. Outliers

CHABADA


"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters
Internet
Access-Location
4. APIs
Internet
Access-Location
Send-SMS
5. Outliers

Clustering

- Want to identify *groups of applications* that are similar according to their descriptions.
- Used *K-Means* to identify such clusters
- Used *elements silhouette*
  to identify best number K of clusters

Clusters

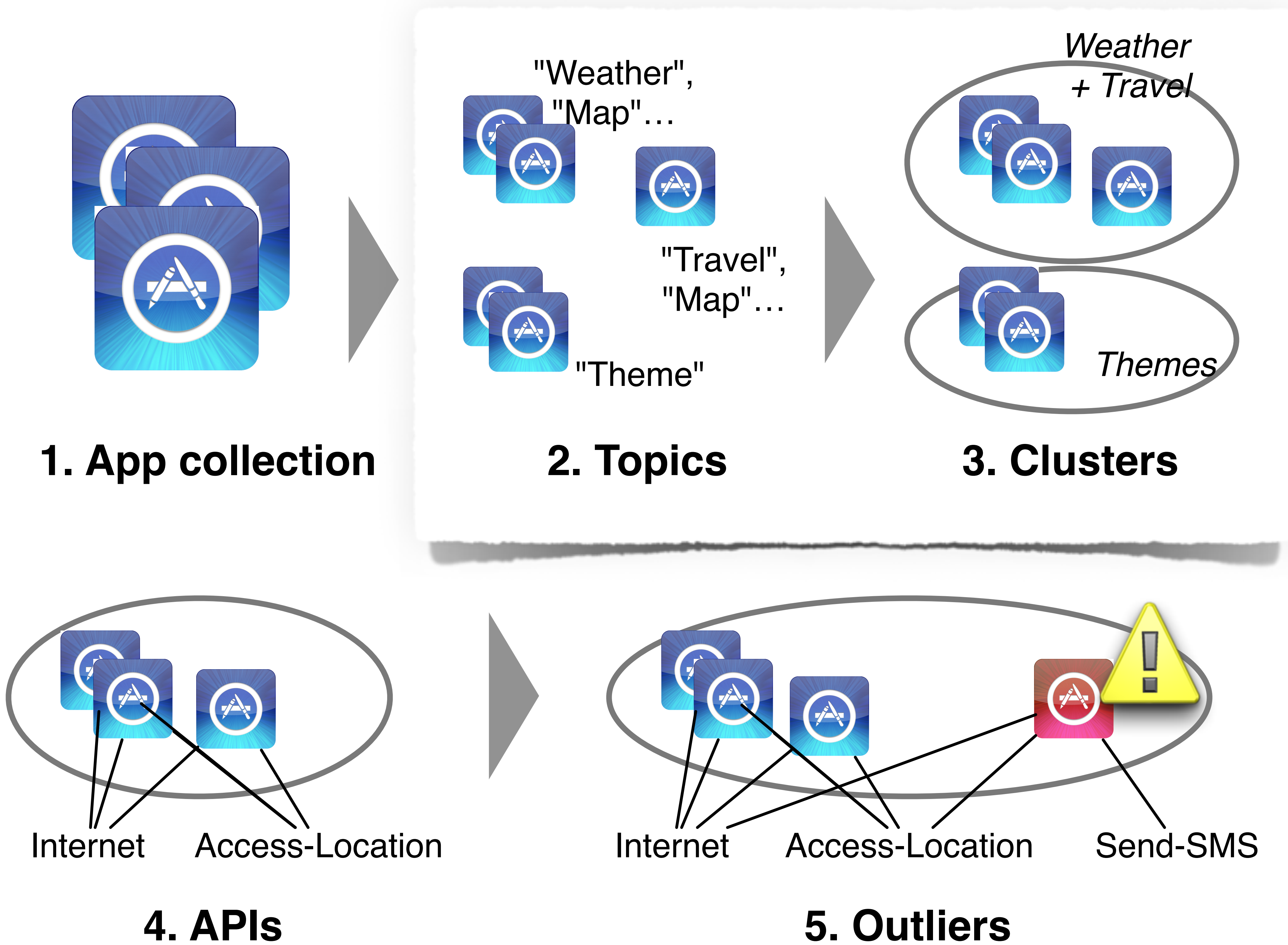
Id	Assigned Name	Size	Most Important Topics
1	"sharing"	1,453	**share** (53%), settings and utils, navigation and travel
2	"puzzle and card games"	953	**puzzle and card games** (78%), share, game
3	"memory puzzles"	1,069	**puzzle and card games** (40%), game (12%), share
4	"music"	714	**music** (58%), share, settings and utils
5	"music videos"	773	**popular media** (44%), **holidays and religion** (20%), share
6	"religious wallpapers"	367	**holidays and religion** (56%), design and art, wallpapers
7	"language"	602	**language** (67%), share, settings and utils
8	"cheat sheets"	785	**game and cheat sheets** (76%), share, popular media
9	"utils"	1,300	**settings and utils** (62%), share, connection
10	"sports game"	1,306	**game** (63%), battle games, puzzle and card games
11	"battle games"	953	**battle games** (60%), **game**

			and travel
19	“sports”	580	**sports** (62%), share, popular media
20	“files and videos”	679	**files and videos** (63%), share, settings and utils
21	“search and browse”	363	**search and browse** (64%), game, puzzle and card games
22	“advertisements”	380	**policies and ads** (97%)
23	“design and art”	978	**design and art** (48%), share, game
24	“car games”	449	**cars** (51%), game, puzzle and card games
25	“tv live”	500	**tv** (57%), share, navigation and travel
26	“adult photo”	828	**photo and social** (59%), share, settings and utils
27	“adult wallpapers”	543	**wallpapers** (51%), share, kids and bodies
28	“ad wallpapers”	180	**policies and ads** (46%), wallpapers, settings and utils
29	“ringtones and sound”	662	**ringtones and sound** (68%), share, settings and utils
30	“theme wallpapers”	593	**wallpapers** (90%), holidays and religion, share
31	“personalize”	402	**personalize** (86%), share, settings and utils
32	“settings and wallpapers”	251	**settings and utils** (37%), **wallpapers** (37%), personalize

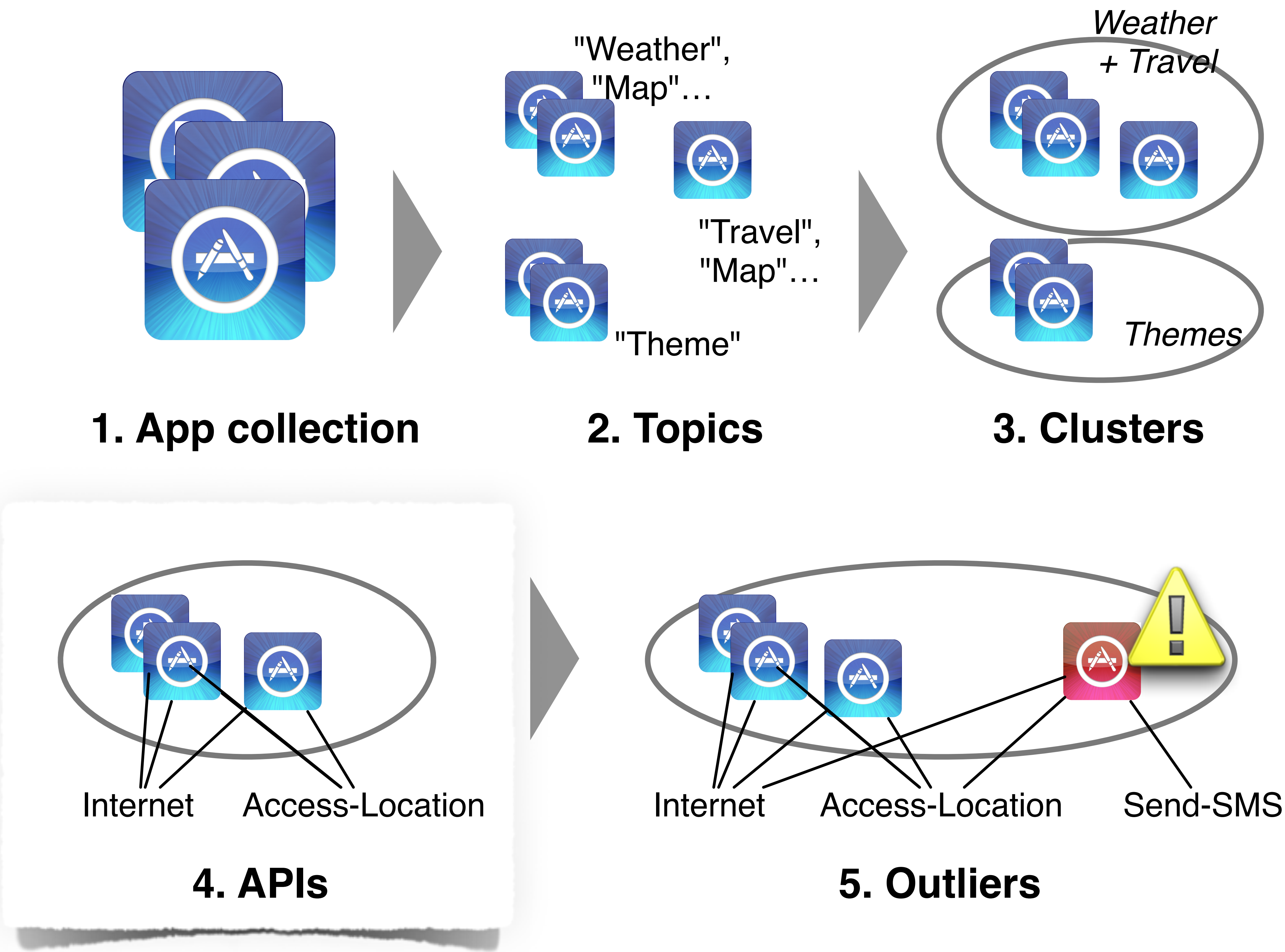
“Personalize” Cluster

“Travel” Cluster

CHABADA


"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters
Internet
Access-Location
Internet
Access-Location
Send-SMS
4. APIs
5. Outliers

CHABADA

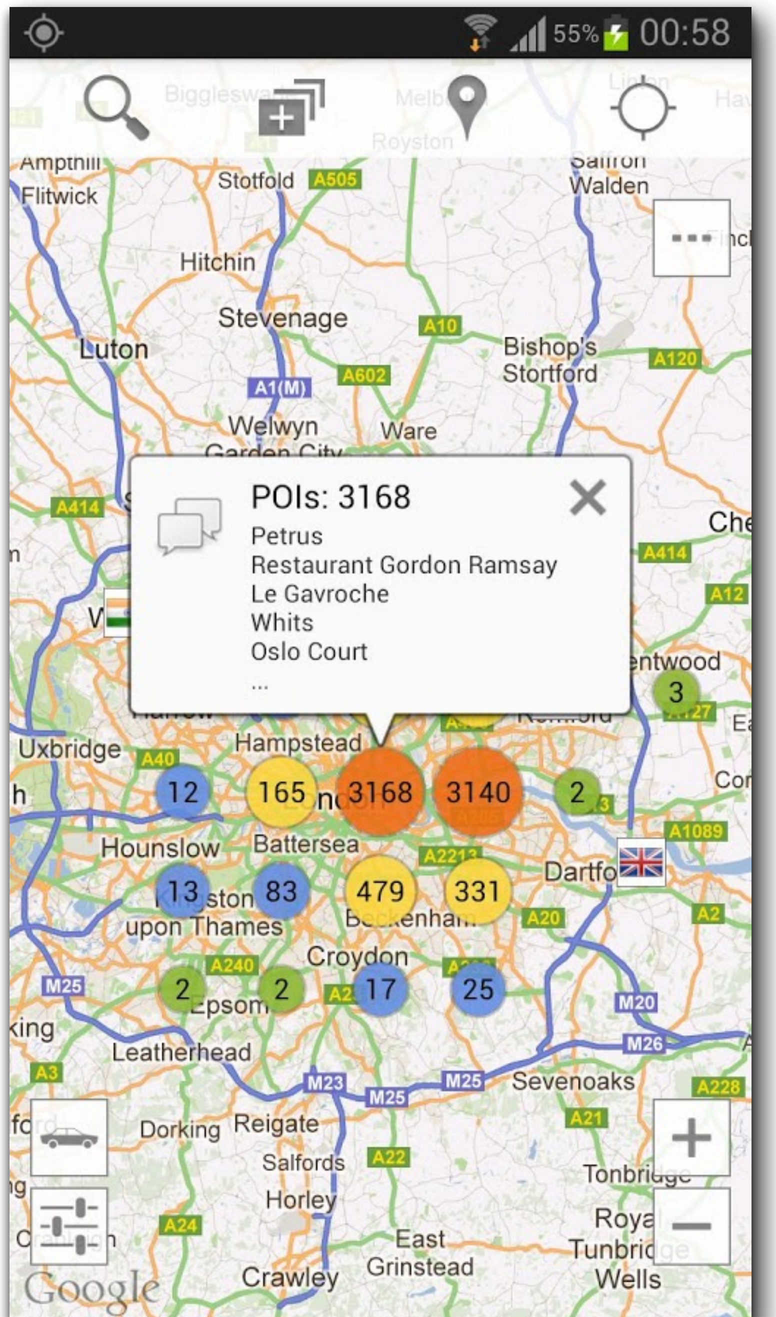


"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters
Internet
Access-Location
4. APIs
Internet
Access-Location
Send-SMS
5. Outliers

API Analysis

- For each APK, we identified the APIs used
- Used simple static analysis
- Only considered *sensitive APIs* which would be governed by *permissions*

London Restaurants



android.net.ConnectivityManager.getActiveNetworkInfo()
android.webkit.WebView()

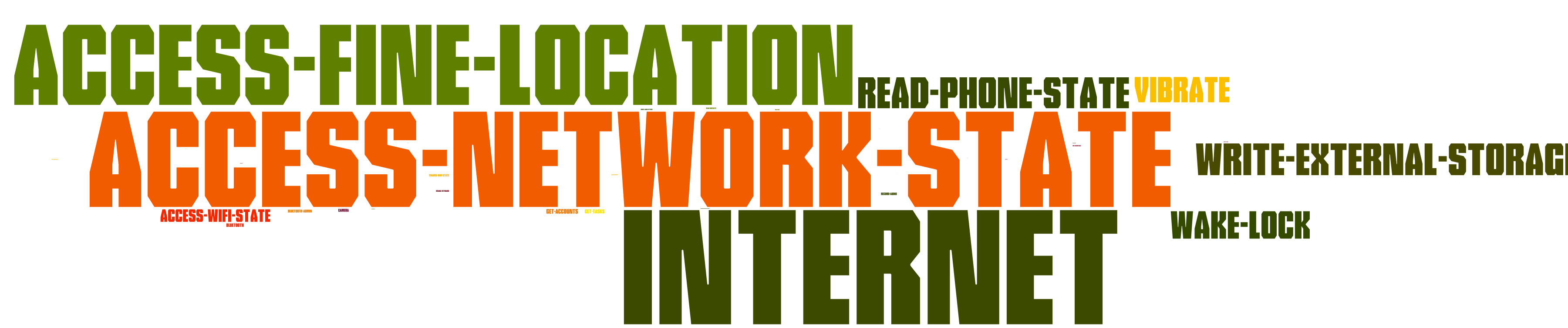
INTERNET
GET-ACCOUNTS
ACCESS-WIFI-STATE
ACCESS-NETWORK-STATE
ACCESS-FINE-LOCATION
READ-PHONE-STATE
VIBRATE

android.net.NetworkInfo.isConnectedOrConnecting()
android.net.ConnectivityManager.getAllNetworkInfo()

"Travel" Cluster

Description

Permissions of APIs used

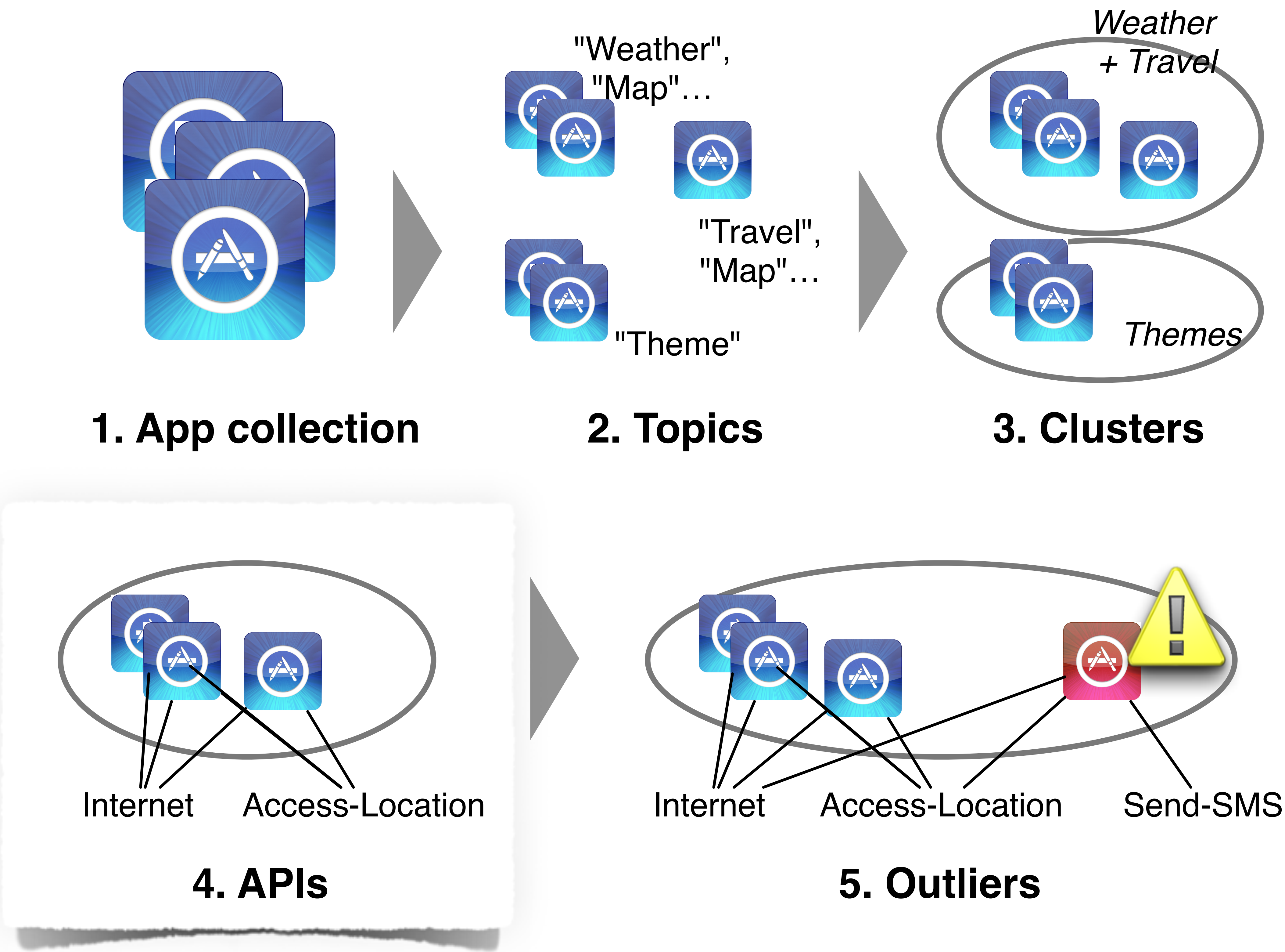


“Personalize” Cluster

Description

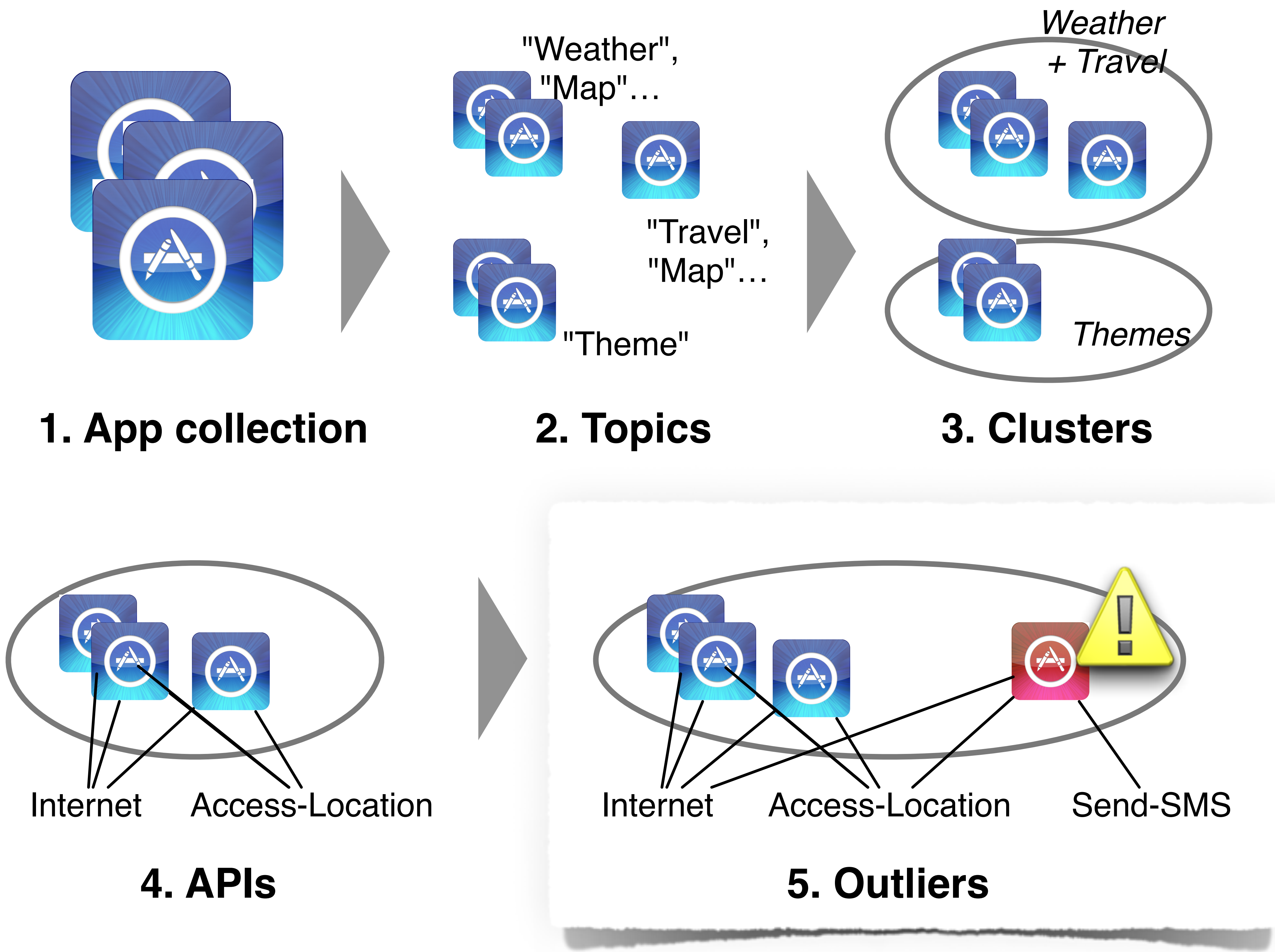
Permissions of APIs used

CHABADA



"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters
Internet
Access-Location
4. APIs
Internet
Access-Location
Send-SMS
5. Outliers

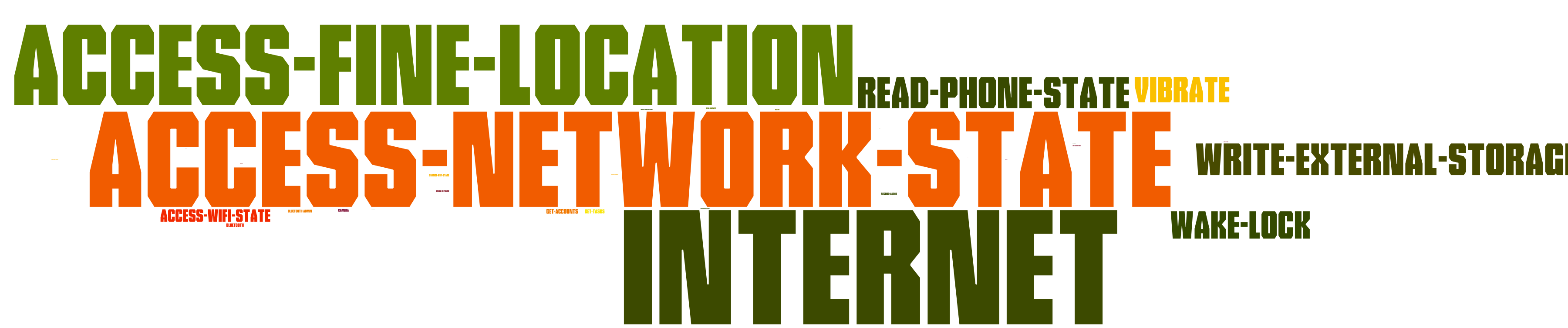
CHABADA


"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters
Internet
Access-Location
Internet
Access-Location
Send-SMS
4. APIs
5. Outliers

“Travel” Cluster

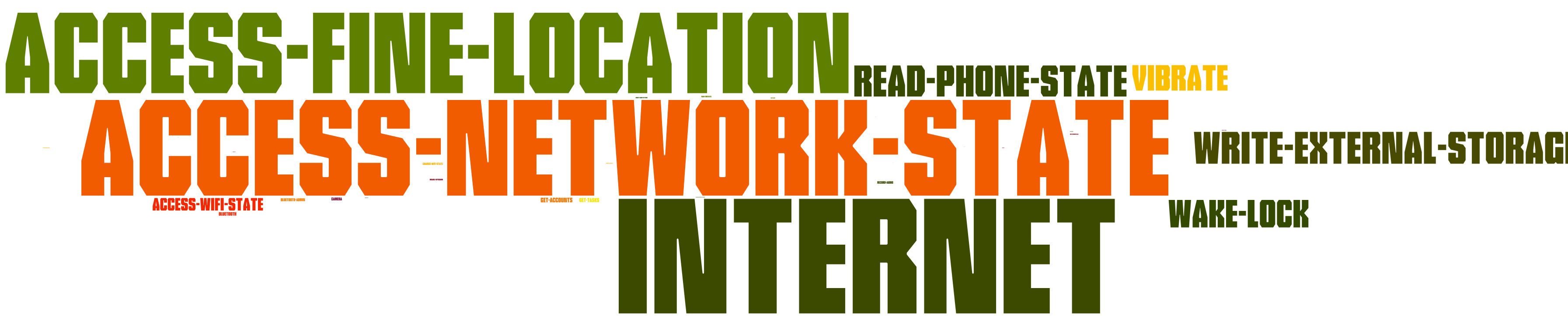
Description

Permissions of APIs used



“Travel” Cluster

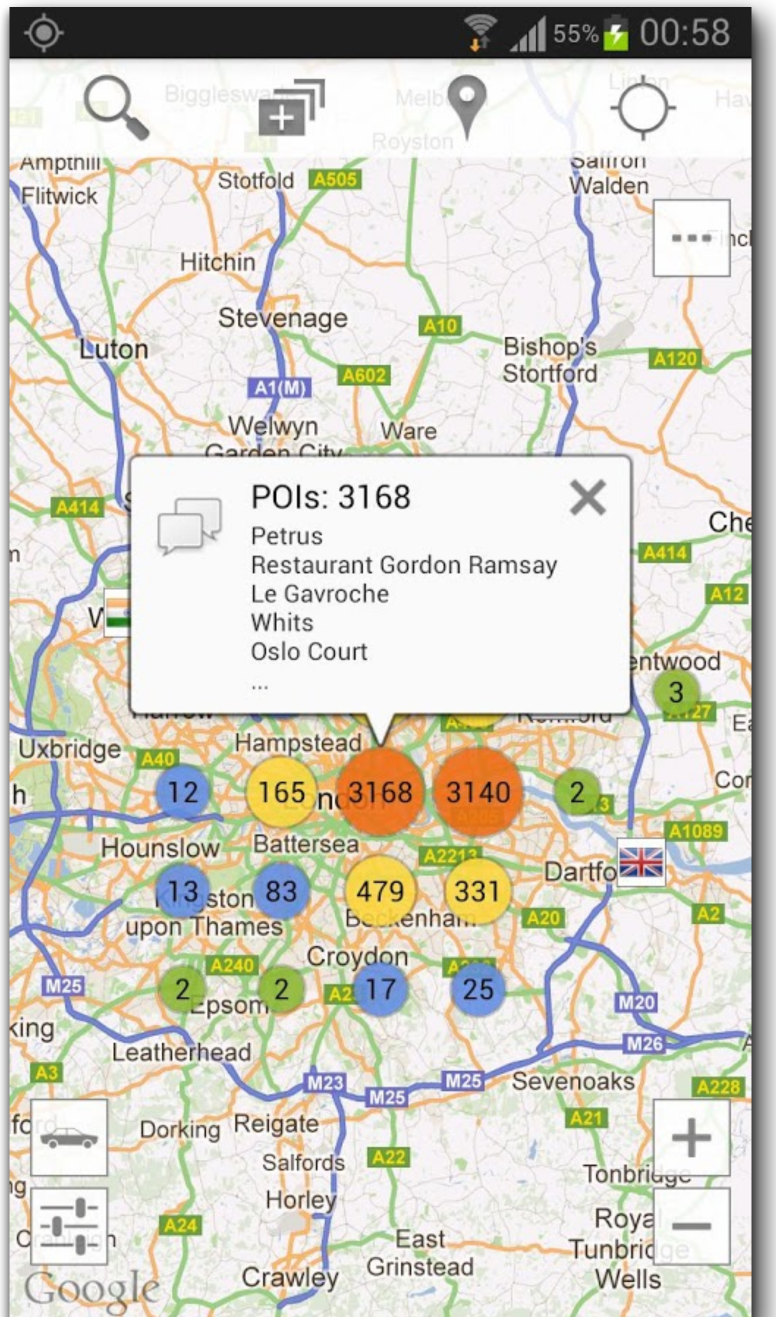
Permissions of APIs used



London Restaurants

Permissions of APIs used

London Restaurants

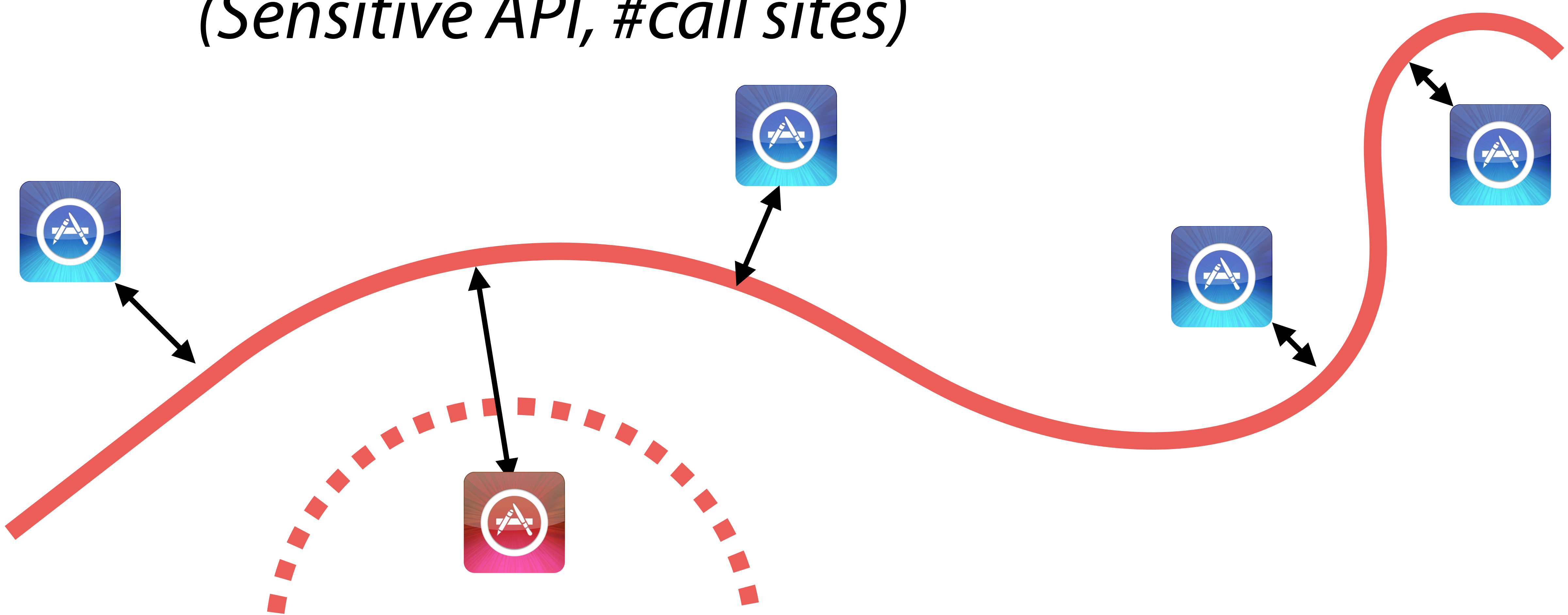


android.net.ConnectivityManager.getActiveNetworkInfo()
android.webkit.WebView()
java.net.HttpURLConnection.connect()
android.app.NotificationManager.notify()
java.net.URL.openConnection()
android.telephony.TelephonyManager.getDeviceId()
org.apache.http.impl.client.DefaultHttpClient()
org.apache.http.impl.client.DefaultHttpClient.execute()
android.location.LocationManager.getBestProvider()
android.telephony.TelephonyManager.getLine1Number()
android.net.wifi.WifiManager.isWifiEnabled()
android.accounts.AccountManager.getAccountsByType()
android.net.wifi.WifiManager.getConnectionInfo()
android.location.LocationManager.getLastKnownLocation()
android.location.LocationManager.isProviderEnabled()
android.location.LocationManager.requestLocationUpdates()
android.net.NetworkInfo.isConnectedOrConnecting()
android.net.ConnectivityManager.getAllNetworkInfo()

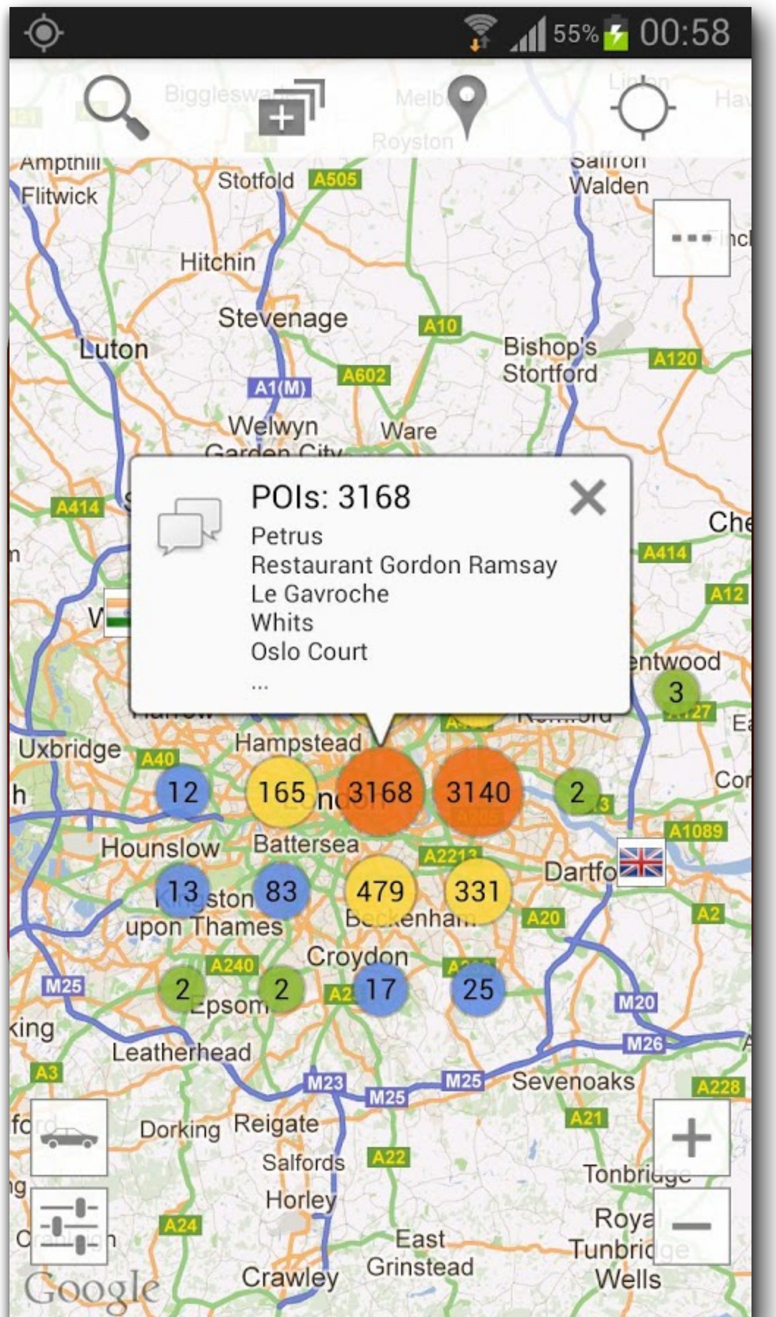
→ An *Outlier* in the “Travel” Cluster

Outlier Analysis

- In each cluster, identified outliers through *one-class support vector machine* (OC-SVM)
- Features of each APK: a vector of *(Sensitive API, #call sites)*



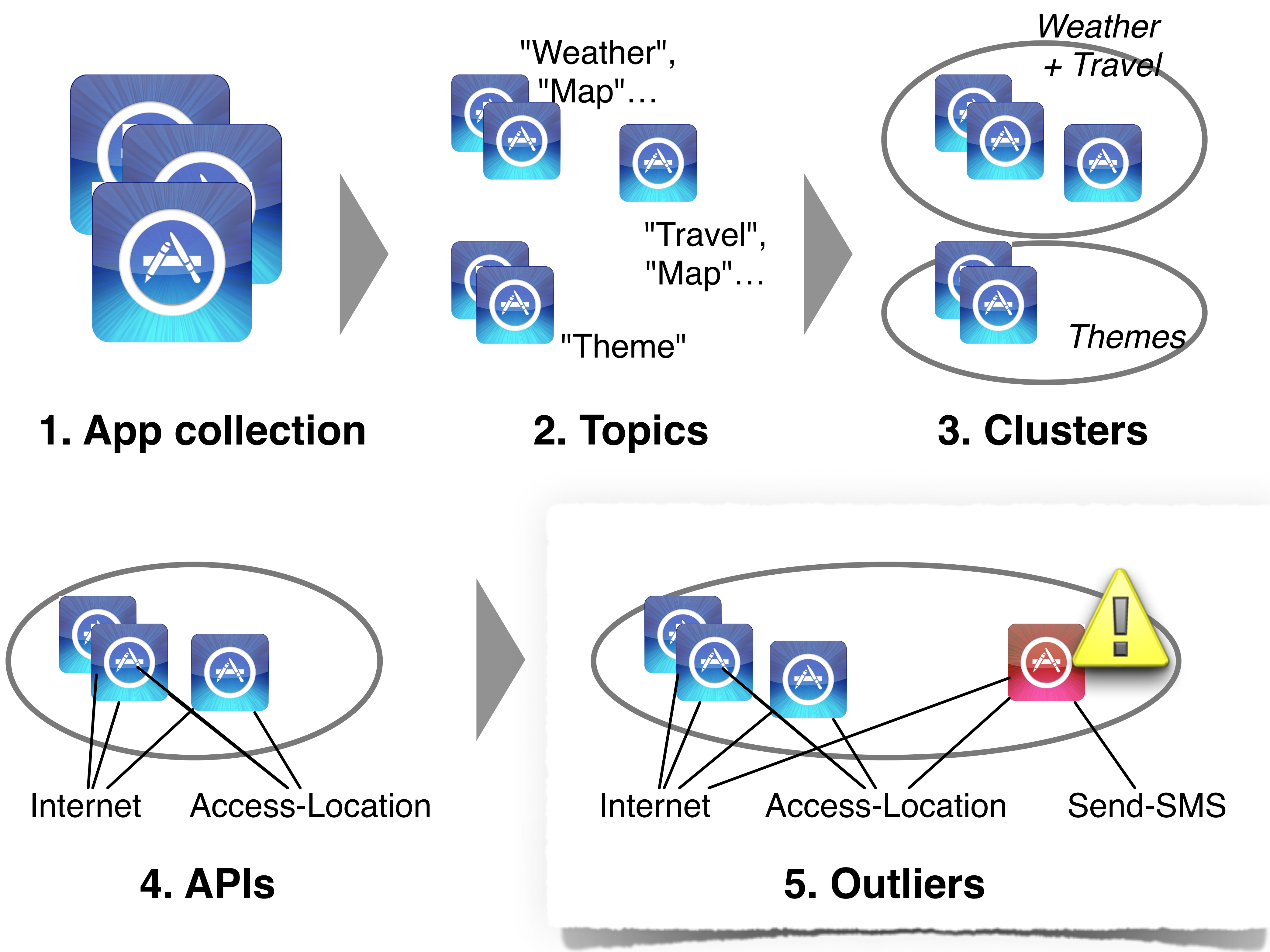
London Restaurants



android.net.ConnectivityManager.getActiveNetworkInfo()
android.webkit.WebView()
java.net.HttpURLConnection.connect()
android.app.NotificationManager.notify()
java.net.URL.openConnection()
android.telephony.TelephonyManager.getDeviceId()
org.apache.http.impl.client.DefaultHttpClient()
org.apache.http.impl.client.DefaultHttpClient.execute()
android.location.LocationManager.getBestProvider()
android.telephony.TelephonyManager.getLine1Number()
android.net.wifi.WifiManager.isWifiEnabled()
android.accounts.AccountManager.getAccountsByType()
android.net.wifi.WifiManager.getConnectionInfo()
android.location.LocationManager.getLastKnownLocation()
android.location.LocationManager.isProviderEnabled()
android.location.LocationManager.requestLocationUpdates()
android.net.NetworkInfo.isConnectedOrConnecting()
android.net.ConnectivityManager.getAllNetworkInfo()

→ Identified as Outlier

CHABADA


"Weather",
"Map"…
"Travel",
"Map"…
"Theme"
Weather
+ Travel
Themes
1. App collection
2. Topics
3. Clusters
Internet
Access-Location
4. APIs
Internet
Access-Location
Send-SMS
5. Outliers

Evaluation: Outliers

- *Can our technique effectively identify anomalies (i.e., mismatches between description and behavior) in Android apps?*
- Manually checked
  top 5 outliers in each cluster (160 total)
- 26% showed *covert behavior using sensitive APIs that acts against the interest of its users.*

What makes an outlier?

- Ad frameworks (apploving, airpush)
- Dubious behavior (UNO, WICKED, Yahoo!)
- Uncommon behavior (SoundCloud)
- Benign outliers (Mr. Will's Stud Poker)

Evaluation: Malware

- *Can our technique be used to identify malicious Android applications?*
- In each cluster, trained OC-SVM on 90% of "benign" apps
- Used TF-IDF as classifier on sets with remaining "benign" apps and 173 known malware apps

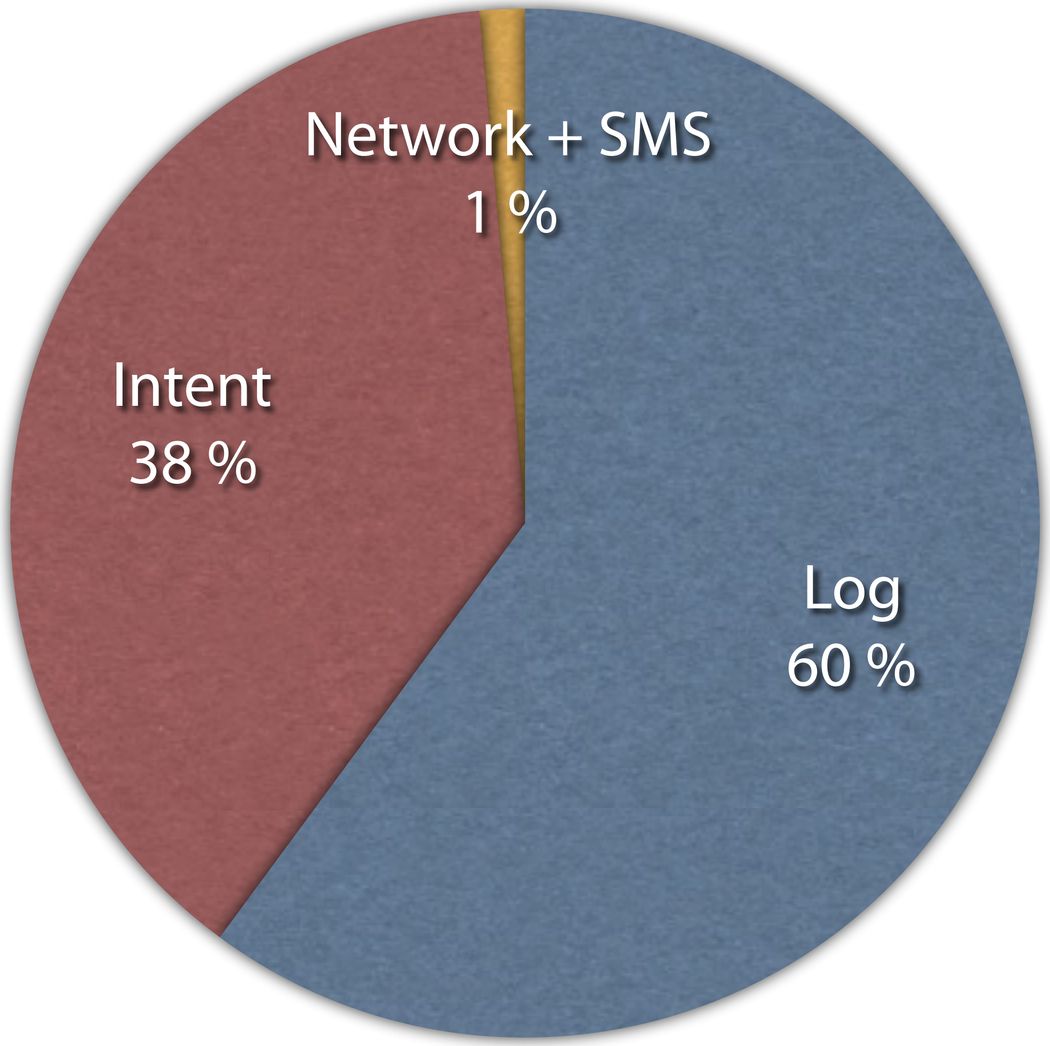
Malware recognition rate >80%

Úlfar Erlingsson

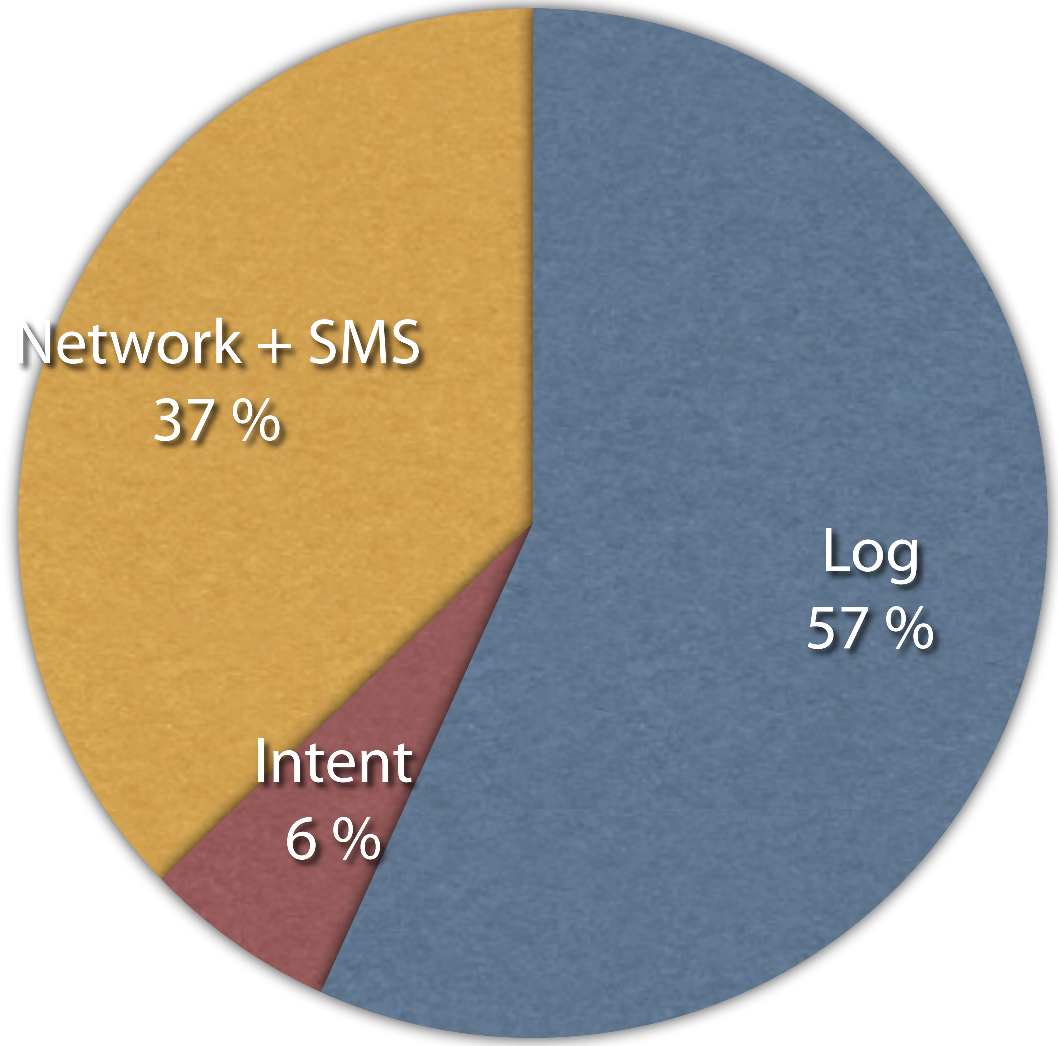
Information Flow

- Which sensitive APIs does the *device ID* flow to?

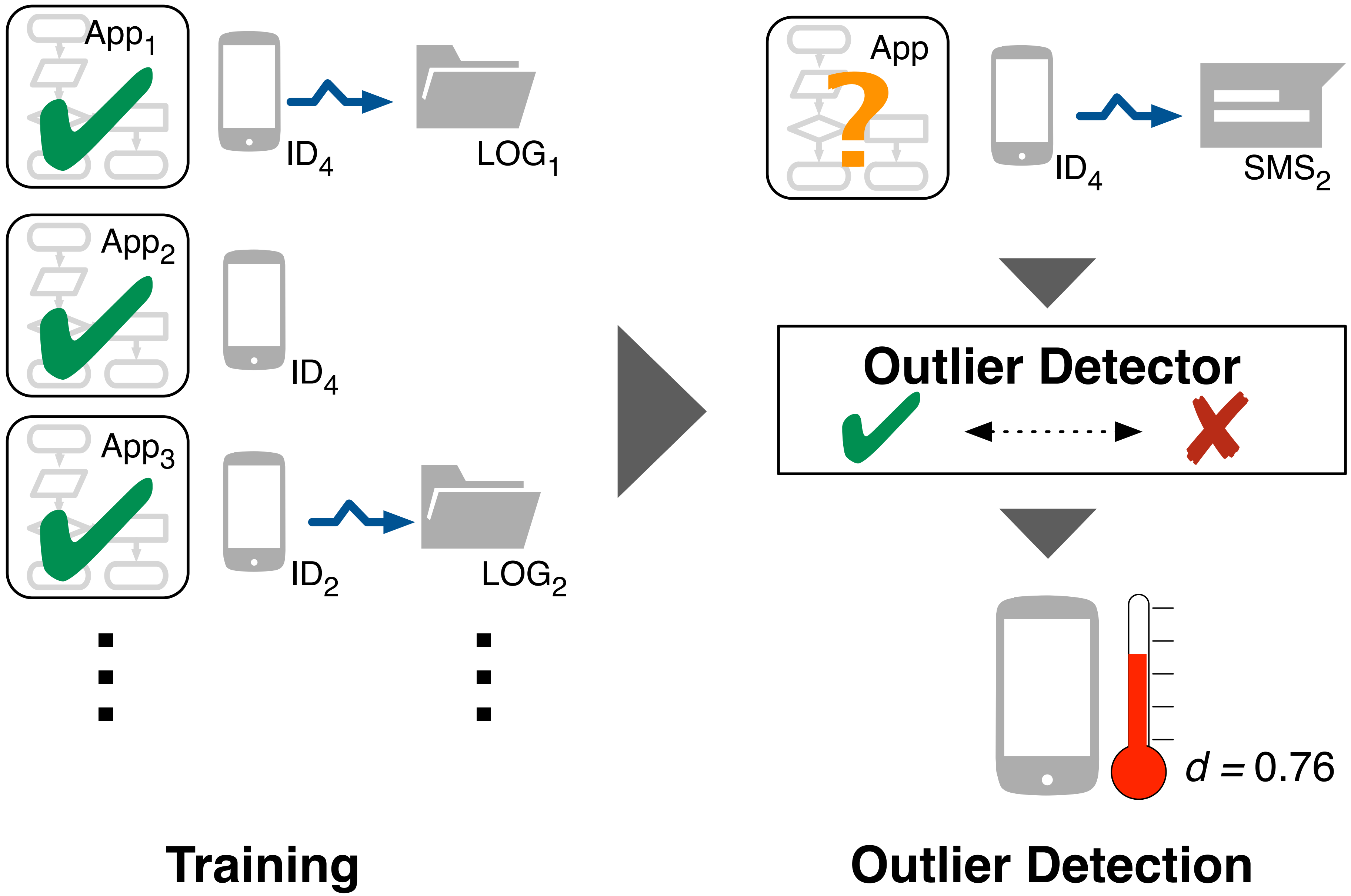
Benign Apps



Malicious Apps



MUDFLOW



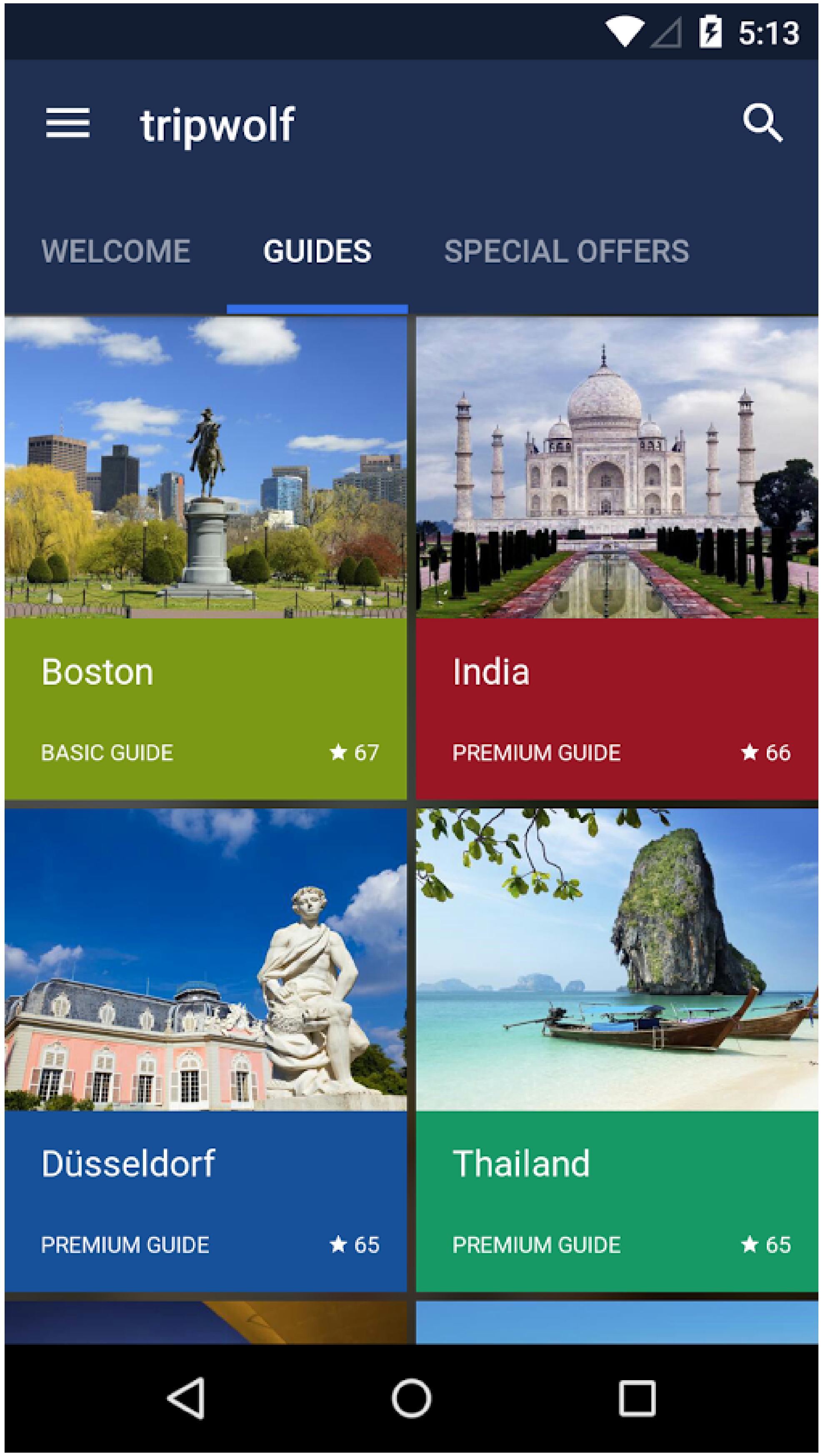
Malware recognition rate >86%


User Interfaces
LOGIN TO YOUR ACCOUNT
Username
Password
LOGIN
Forget username or password
Login with facebook
16 Septemper
PORTSAID - EGYPT
FRI 11/10
28°
LOREM IPSUM DOLOR
LOAD MORE
$299.50
BASIC
10 GB amount of space
10 Users
5 Projects
SIGN UP
PROFESSIONAL
$50
WEDNESDAY
26
ADD EVENT
January
Upload Files
Drag Files Here
BROWSE
UPLOAD
Downloading...
Uploading...
HEAD MENU
Elizabeth Munoz
Notification
Messages
Likes
Explore
Anna Flores
Jackson Santiago
HIRE
Chart
Inquiries
Winners
9,362.74
Losers
6,734.89
Invites
Events
Account Settings
Statistics
Mailbox
Inbox
Starred
Flagged
Sent
Deleted
More
OS AUDIENCE STATS
JUNE
2013
iOS
21%
Mac
48%
Linux
9%
Win
32%
Week
Month
Year
JAN 2013
+ 25%
+ 17%
+ 35%
Fresh apricot with huge sale
$450 - 10 Days
SAT
SUN
MON
TUE
28°
20°
30°
35°

Outliers

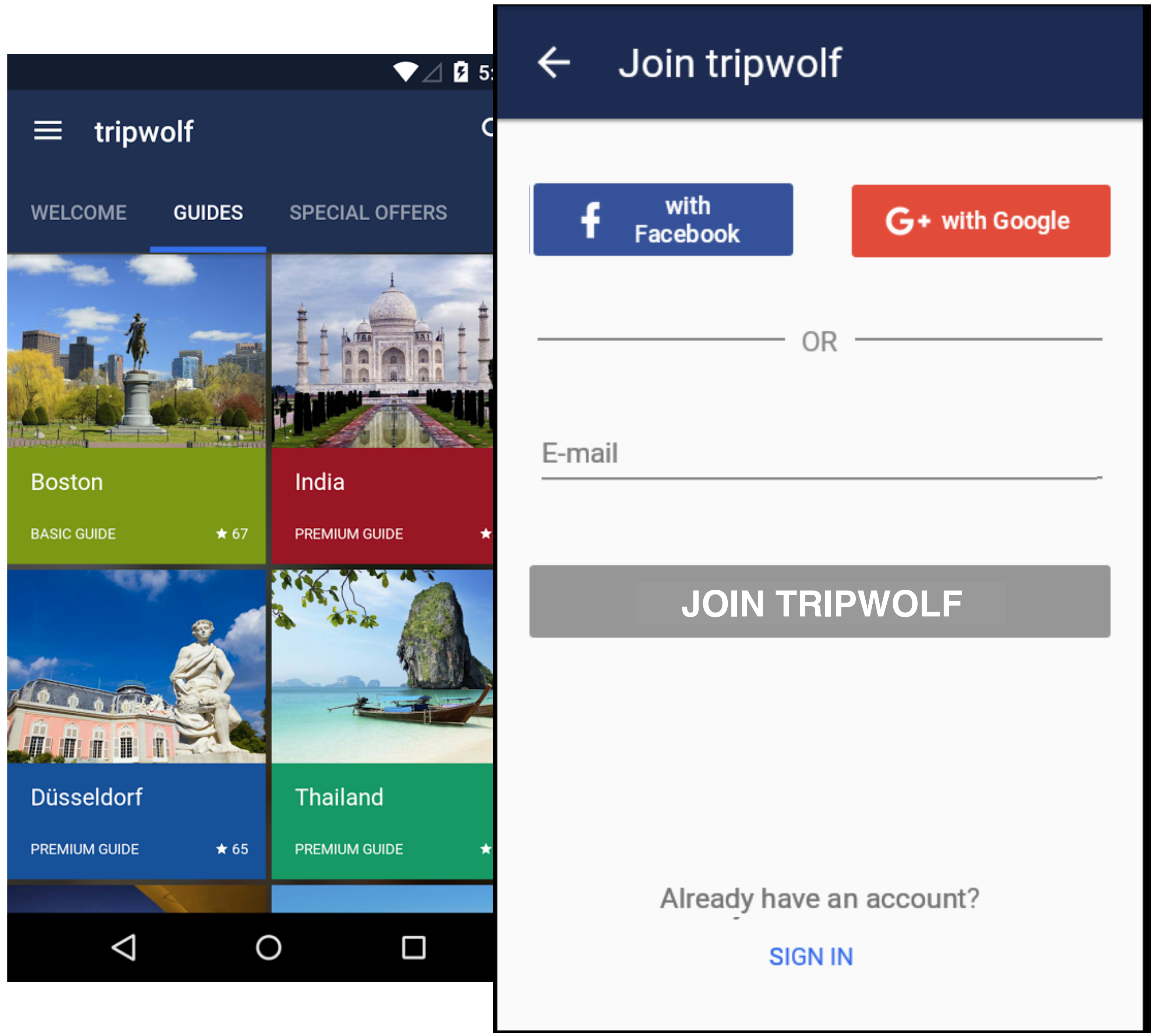
Detect Outliers for UI Elements!

Tripwolf

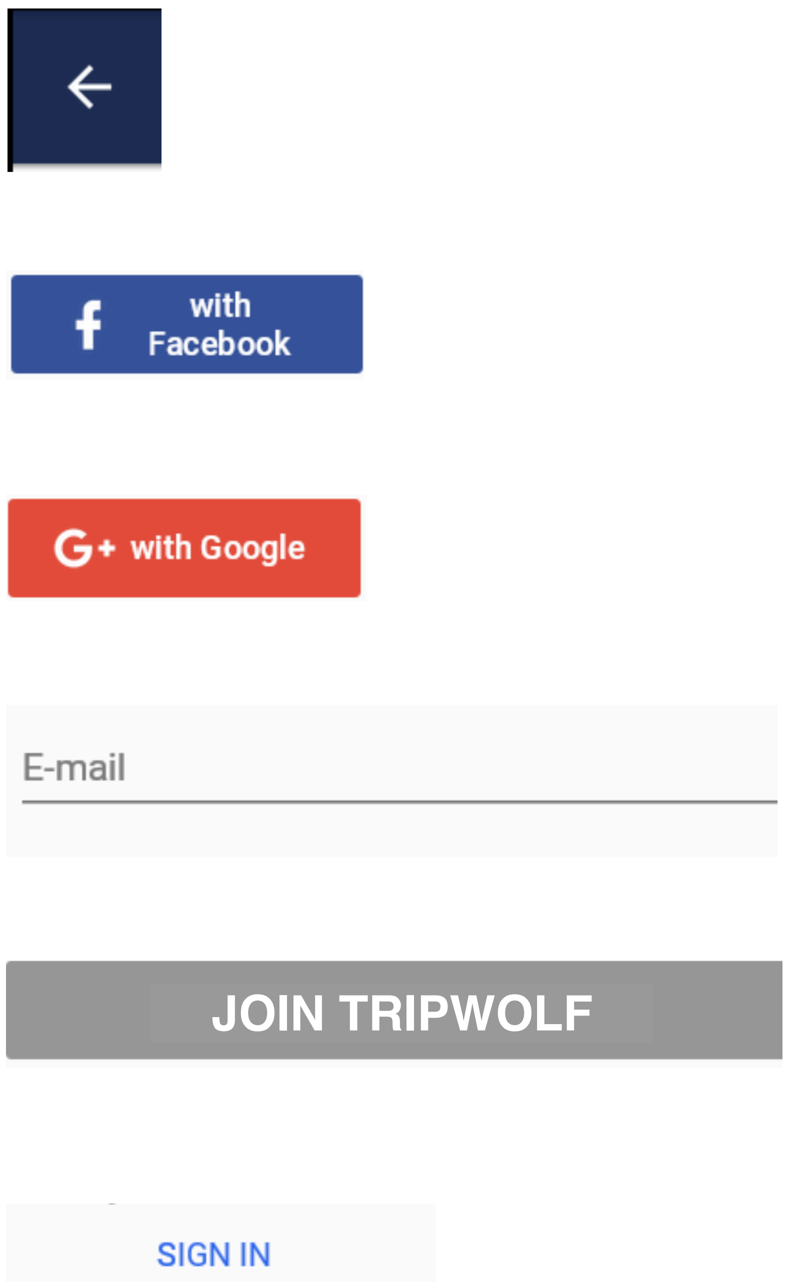


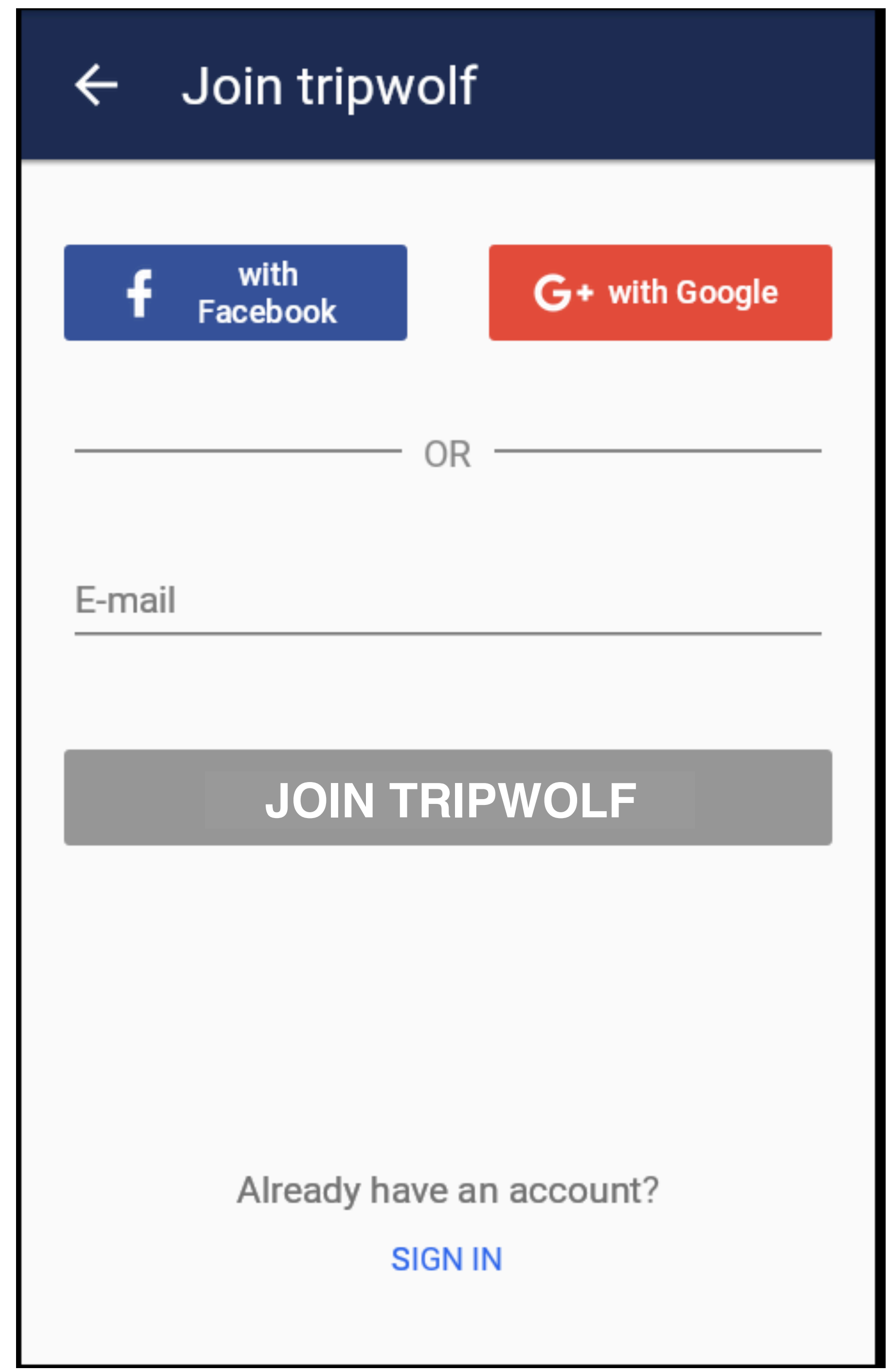
A serious travel app

Registration

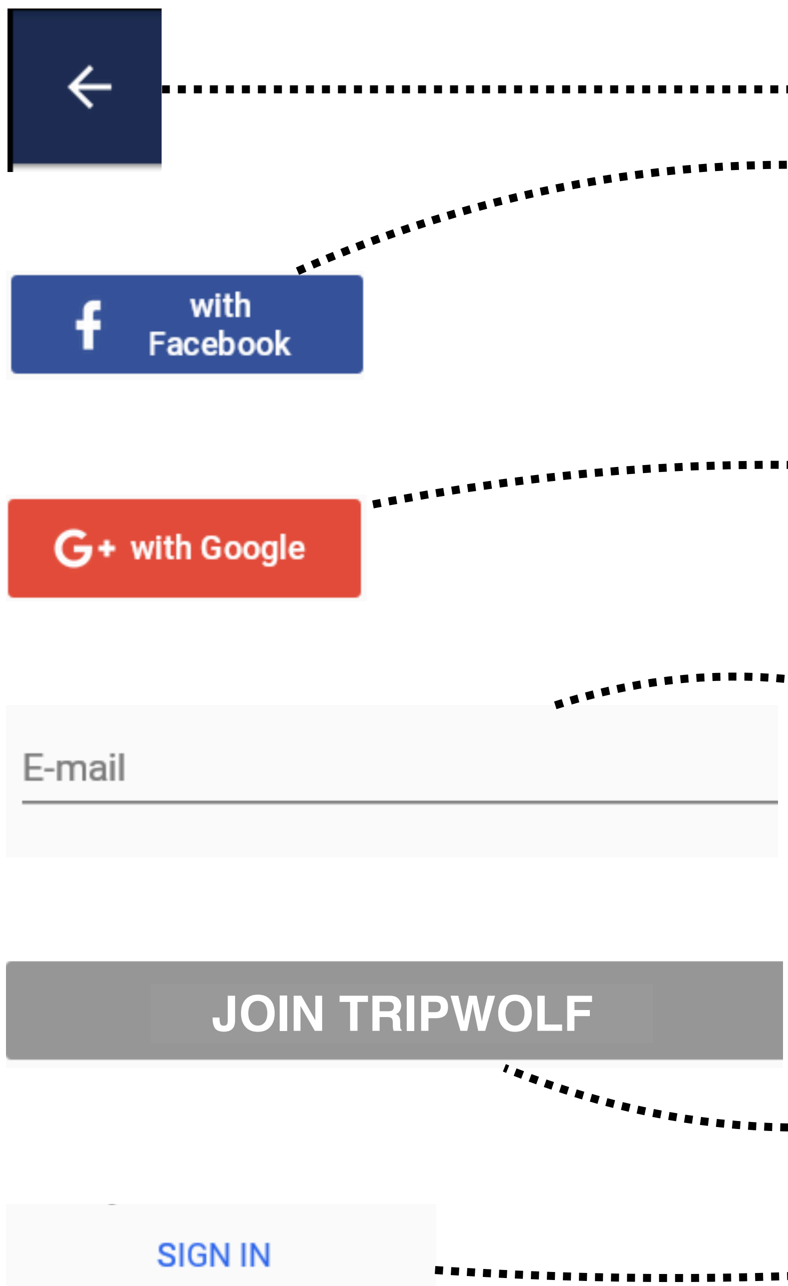


UI Elements


with
Facebook
with Google
E-mail
JOIN TRIPWOLF
SIGN IN


Join tripwolf
with
Facebook
with Google
OR
E-mail
JOIN TRIPWOLF
Already have an account?
SIGN IN

APIs used



startActivity

Handler
Bundle
startActivity

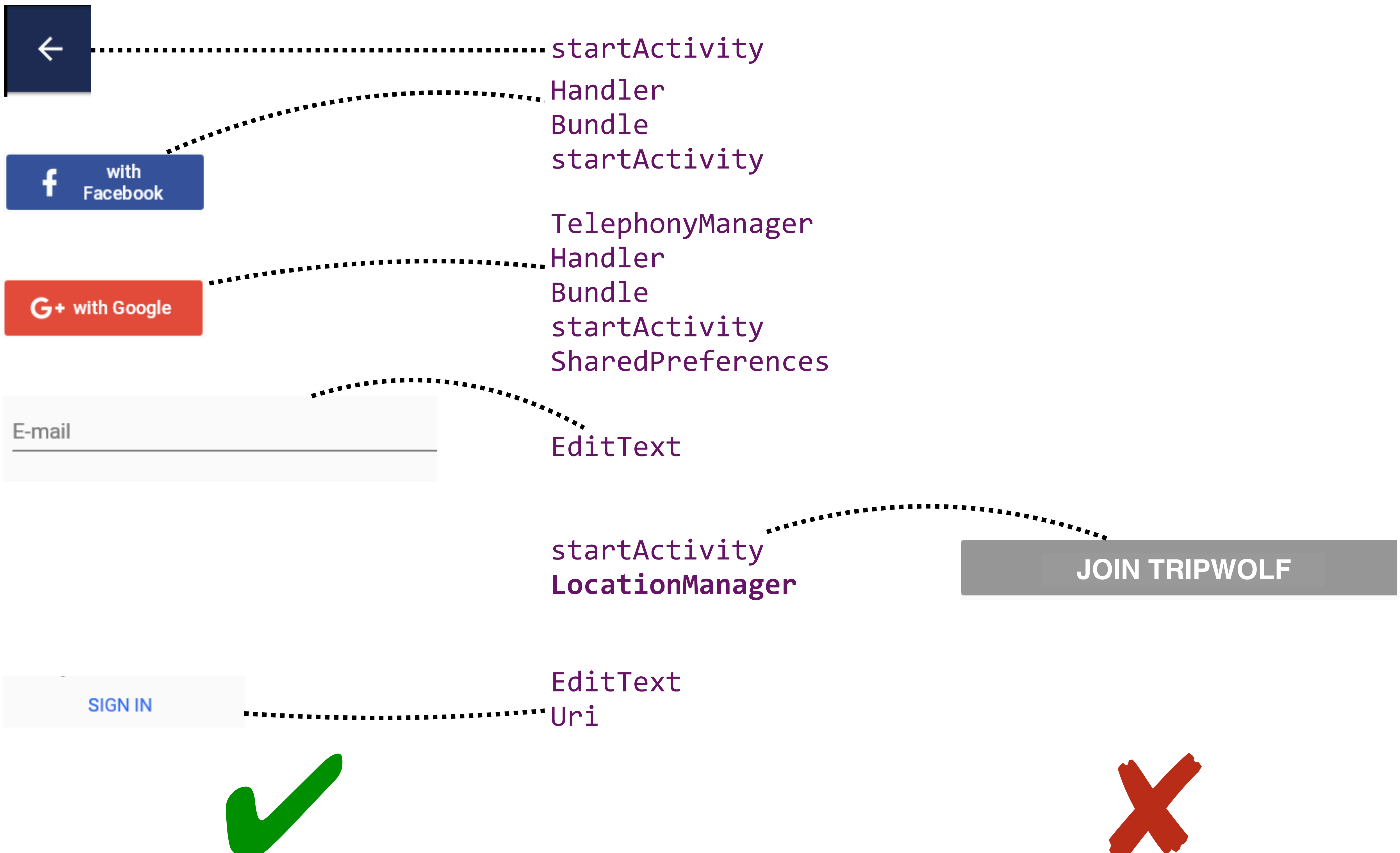
TelephonyManager
Handler
Bundle
startActivity
SharedPreferences

EditText

startActivity
LocationManager

EditText
Uri

Outliers



startActivity
with Facebook
Handler
Bundle
startActivity
G+ with Google
TelephonyManager
Handler
Bundle
startActivity
SharedPreferences
E-mail
EditText
startActivity
LocationManager
JOIN TRIPWOLF
SIGN IN
EditText
Uri

Outliers

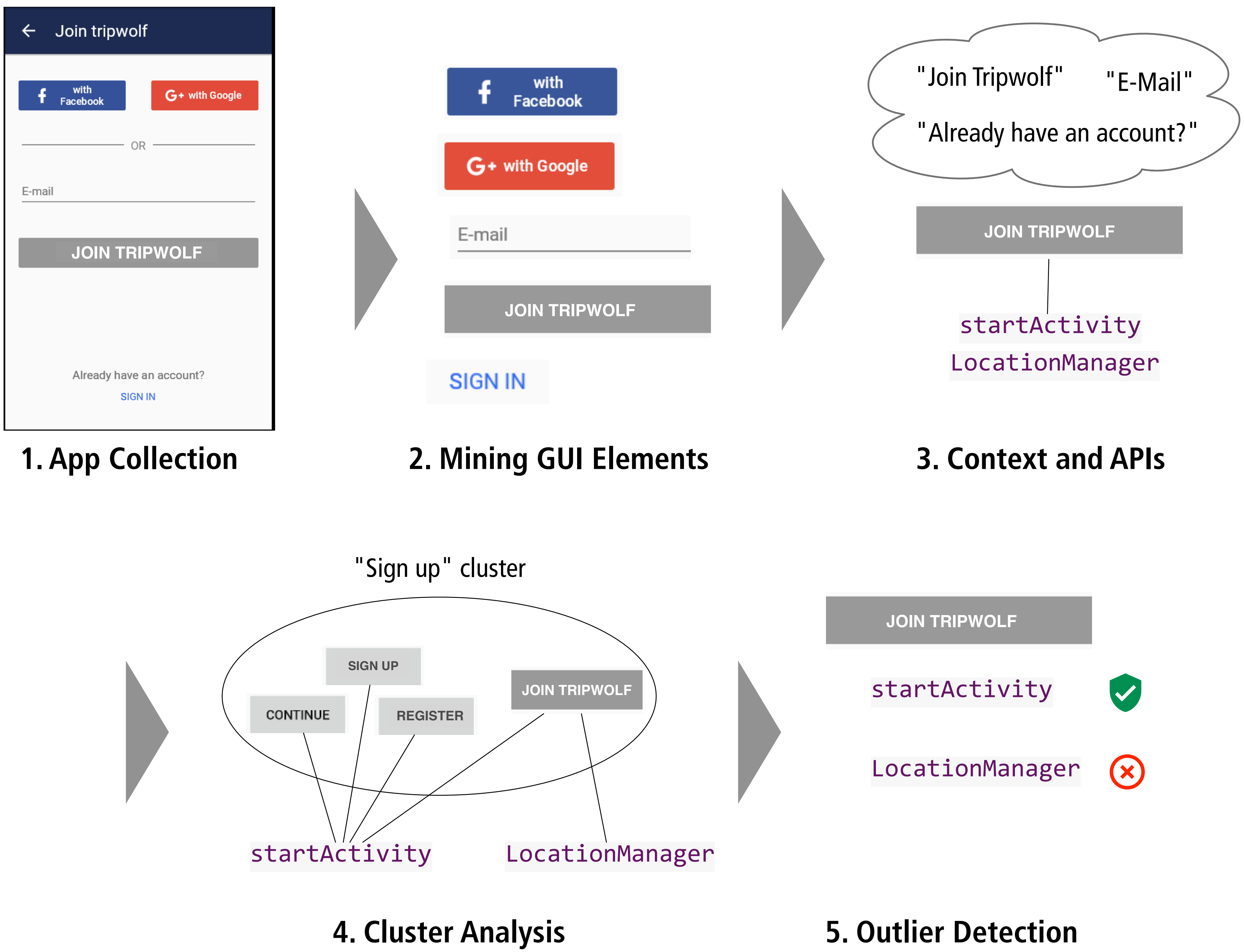
The "Join Tripwolf" button transmits the current location to the Tripwolf servers.

`startActivity`
`LocationManager`

JOIN TRIPWOLF

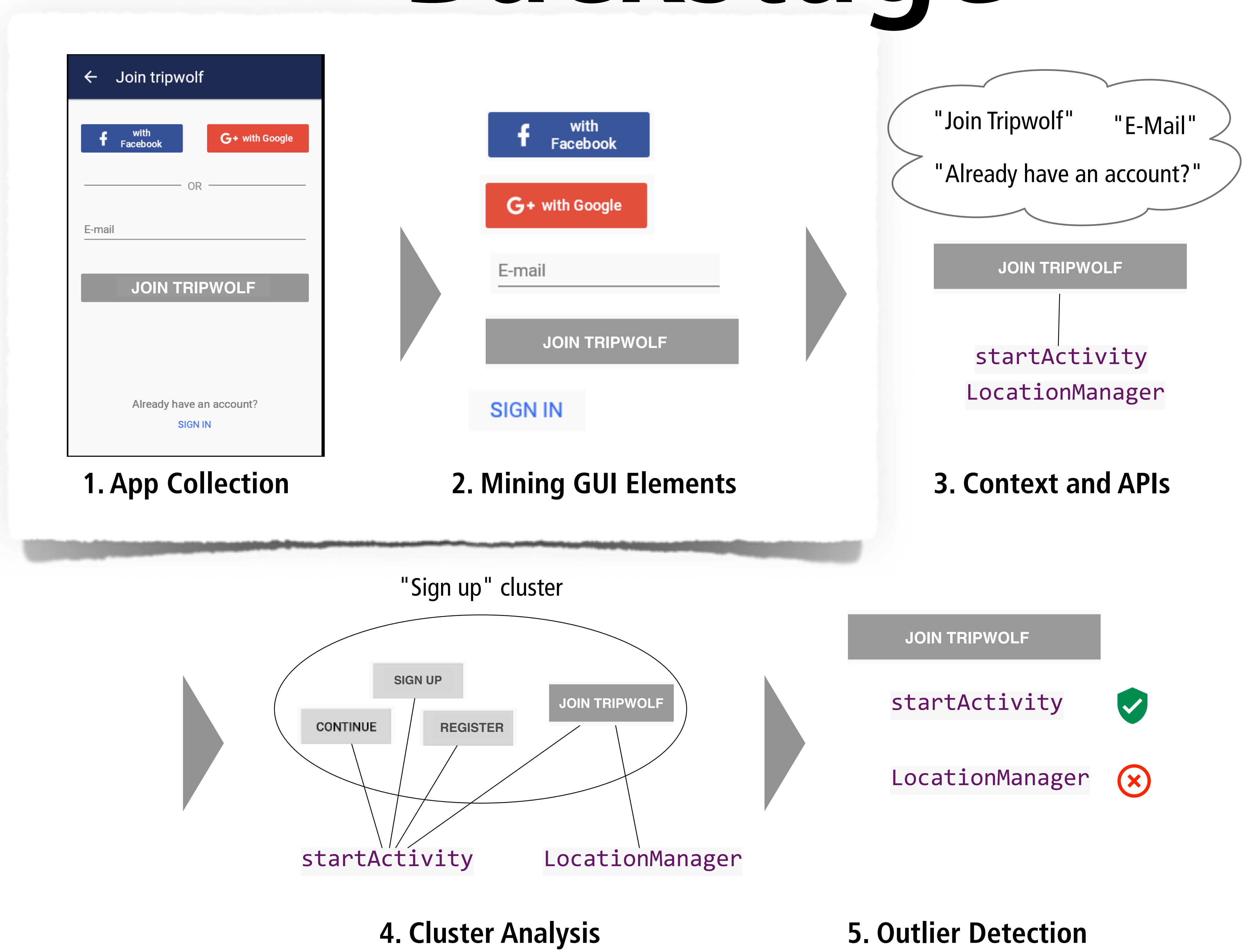
This is unusual for "join" buttons

Backstage



Join tripwolf
with Facebook
with Google
OR
E-mail
JOIN TRIPWOLF
Already have an account?
SIGN IN
1. App Collection
with Facebook
with Google
E-mail
JOIN TRIPWOLF
SIGN IN
2. Mining GUI Elements
"Join Tripwolf"
"E-Mail"
"Already have an account?"
JOIN TRIPWOLF
startActivity
LocationManager
3. Context and APIs
"Sign up" cluster
SIGN UP
CONTINUE
REGISTER
JOIN TRIPWOLF
startActivity
LocationManager
4. Cluster Analysis
JOIN TRIPWOLF
startActivity
LocationManager
5. Outlier Detection

Backstage



Join tripwolf
with Facebook
with Google
OR
E-mail
JOIN TRIPWOLF
Already have an account?
SIGN IN
1. App Collection
with Facebook
with Google
E-mail
JOIN TRIPWOLF
SIGN IN
2. Mining GUI Elements
"Join Tripwolf"
"E-Mail"
"Already have an account?"
JOIN TRIPWOLF
startActivity
LocationManager
3. Context and APIs
"Sign up" cluster
SIGN UP
CONTINUE
REGISTER
JOIN TRIPWOLF
startActivity
LocationManager
4. Cluster Analysis
JOIN TRIPWOLF
startActivity
LocationManager
5. Outlier Detection

Mining GUI Elements

```
<?xml version="1.0" encoding="utf-8"?>
 <LinearLayout android:layout_width="fill_parent"
  android:layout_height="fill_parent">
  <fragment android:id="@+id/fragment"
   class="uinomaly.fragmentclass".../>
  <Button android:id="@+id/buttonOK"
          android:text="@string/buttonOK"
          android:onClick="xmlDefinedOnClick"
          style="@style/okButtonStyle"/>
  <ImageButton android:id="@+id/imageButtonPrint" ...
          android:src="@drawable/print_button"
          android:contentDescription="@string/printText" />
 </LinearLayout>
```

Mining GUI Elements

Text shown

```
<Button android:id="@+id/buttonOK"
        android:text="@string/buttonOK"
        android:onClick="xmlDefinedOnClick"
        style="@style/okButtonStyle"/>
```

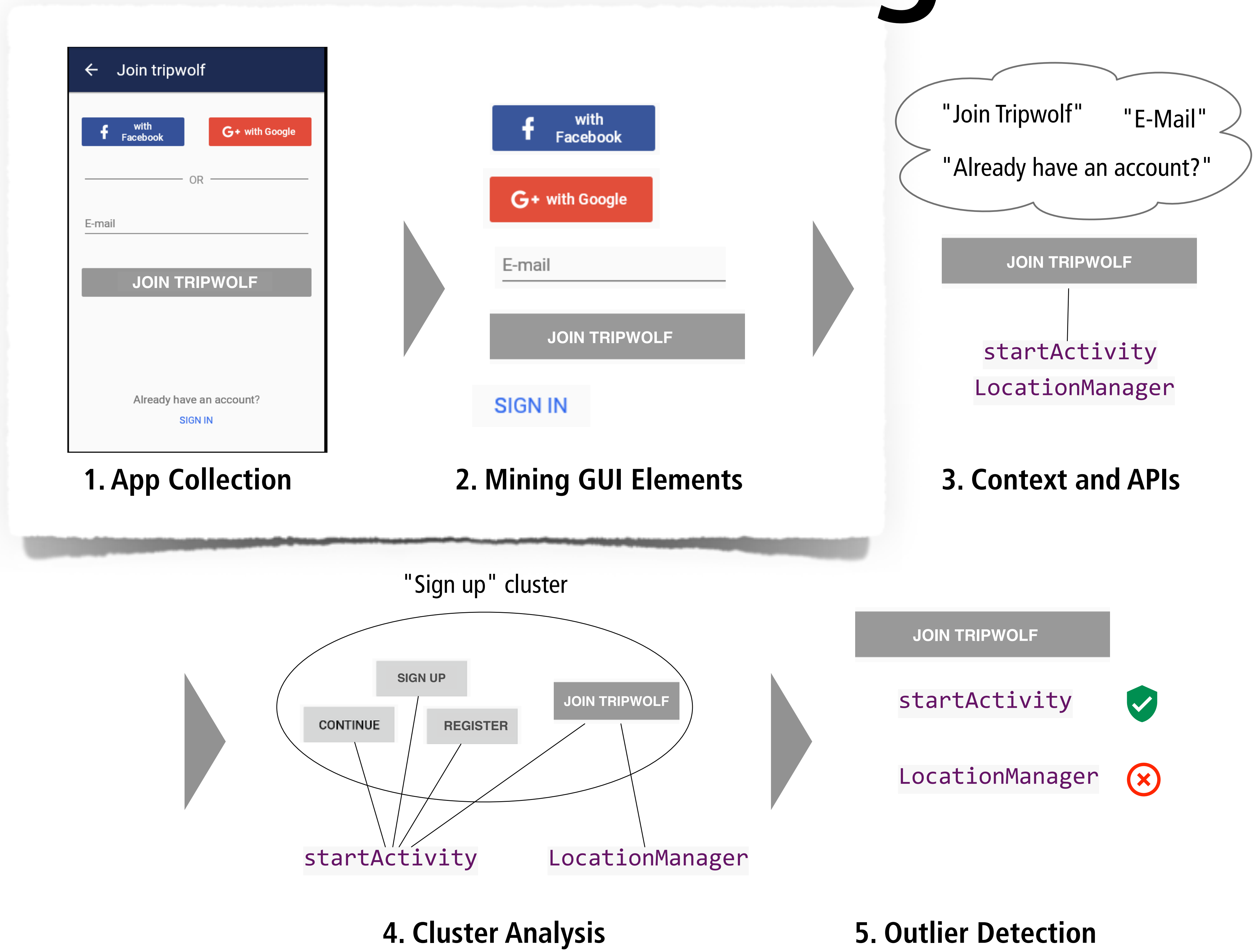
Callback

```
<ImageButton android:id="@+id/imageButtonPrint" ...
        android:src="@drawable/print_button"
        android:contentDescription="@string/printText" />
```

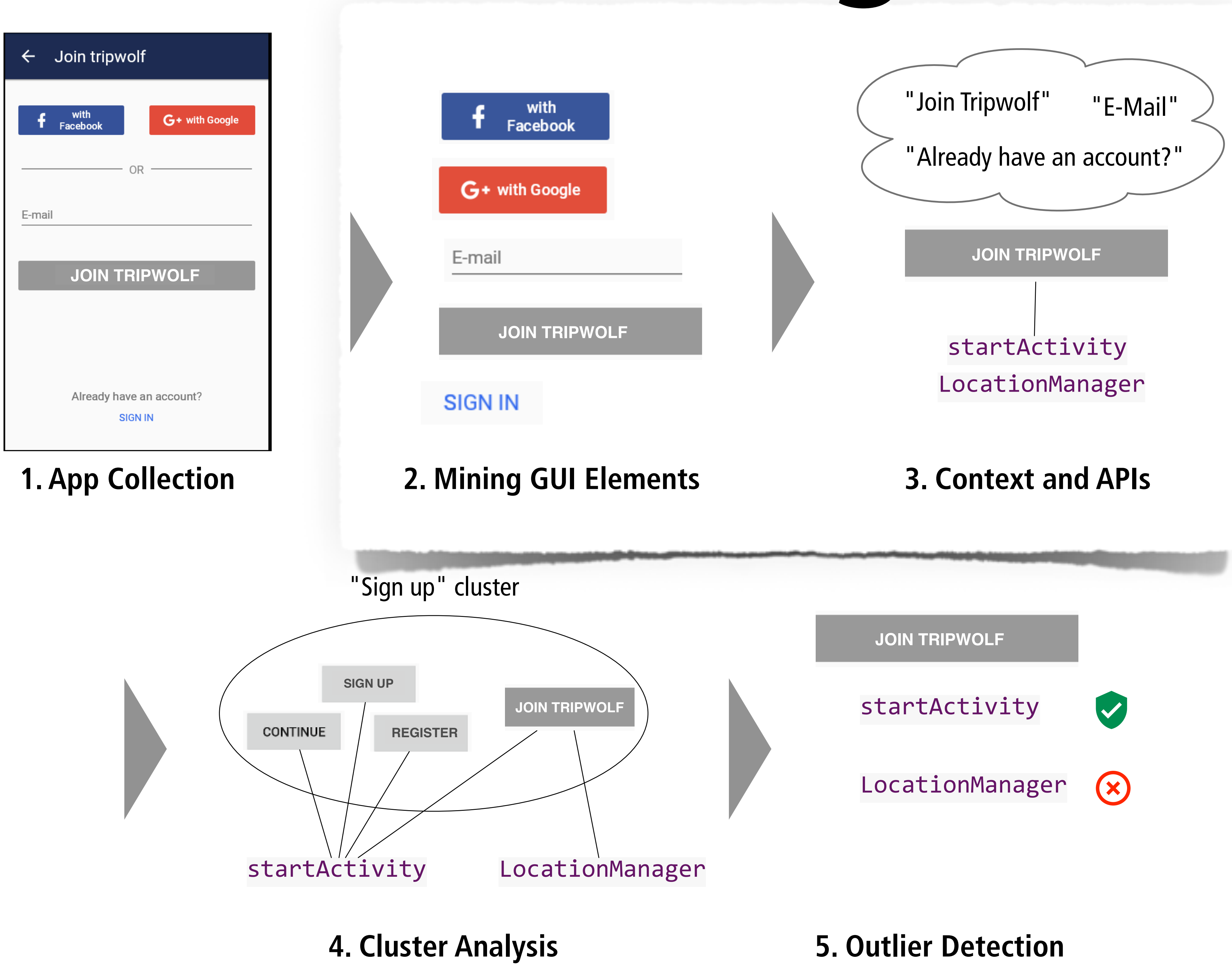
Icon file

Alternate Text

Backstage


Join tripwolf
with Facebook
with Google
OR
E-mail
JOIN TRIPWOLF
Already have an account?
SIGN IN
1. App Collection
with Facebook
with Google
E-mail
JOIN TRIPWOLF
SIGN IN
2. Mining GUI Elements
"Join Tripwolf"
"E-Mail"
"Already have an account?"
JOIN TRIPWOLF
startActivity
LocationManager
3. Context and APIs
"Sign up" cluster
SIGN UP
CONTINUE
REGISTER
JOIN TRIPWOLF
startActivity
LocationManager
4. Cluster Analysis
JOIN TRIPWOLF
startActivity
LocationManager
5. Outlier Detection

Backstage



Join tripwolf
with Facebook
with Google
OR
E-mail
JOIN TRIPWOLF
Already have an account?
SIGN IN
1. App Collection
with Facebook
with Google
E-mail
JOIN TRIPWOLF
SIGN IN
2. Mining GUI Elements
"Join Tripwolf"
"E-Mail"
"Already have an account?"
JOIN TRIPWOLF
startActivity
LocationManager
3. Context and APIs
"Sign up" cluster
SIGN UP
CONTINUE
REGISTER
JOIN TRIPWOLF
startActivity
LocationManager
4. Cluster Analysis
JOIN TRIPWOLF
startActivity
LocationManager
5. Outlier Detection

Mining APIs

- To identify APIs called, Backstage uses a static analysis built on top of Soot
- Builds a *call graph* starting with APIs defined in layout file
- Collects all reachable Android APIs

Context Sensitivity

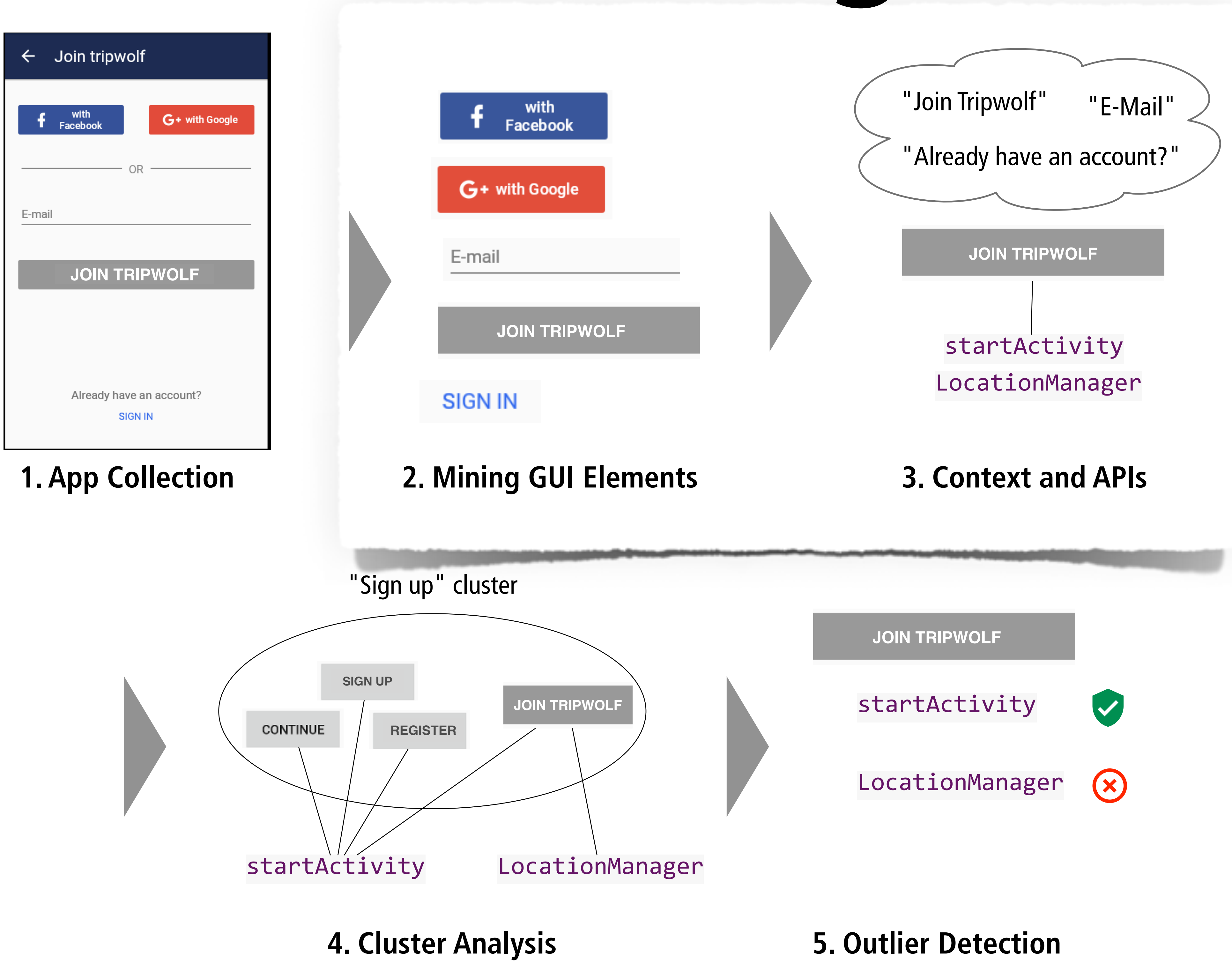
```
View.OnClickListener myClick = new View.OnClickListener() {
    public void onClick(View v) {
     switch (v.getId()) {
      case R.id.ok_button:
       // action if button is the okButton
       break;

      case R.id.cancel_button:
       // action if button is the cancelButton
       break;
     }
    }
  };
```

Two Buttons, Same Callback

Mining Text

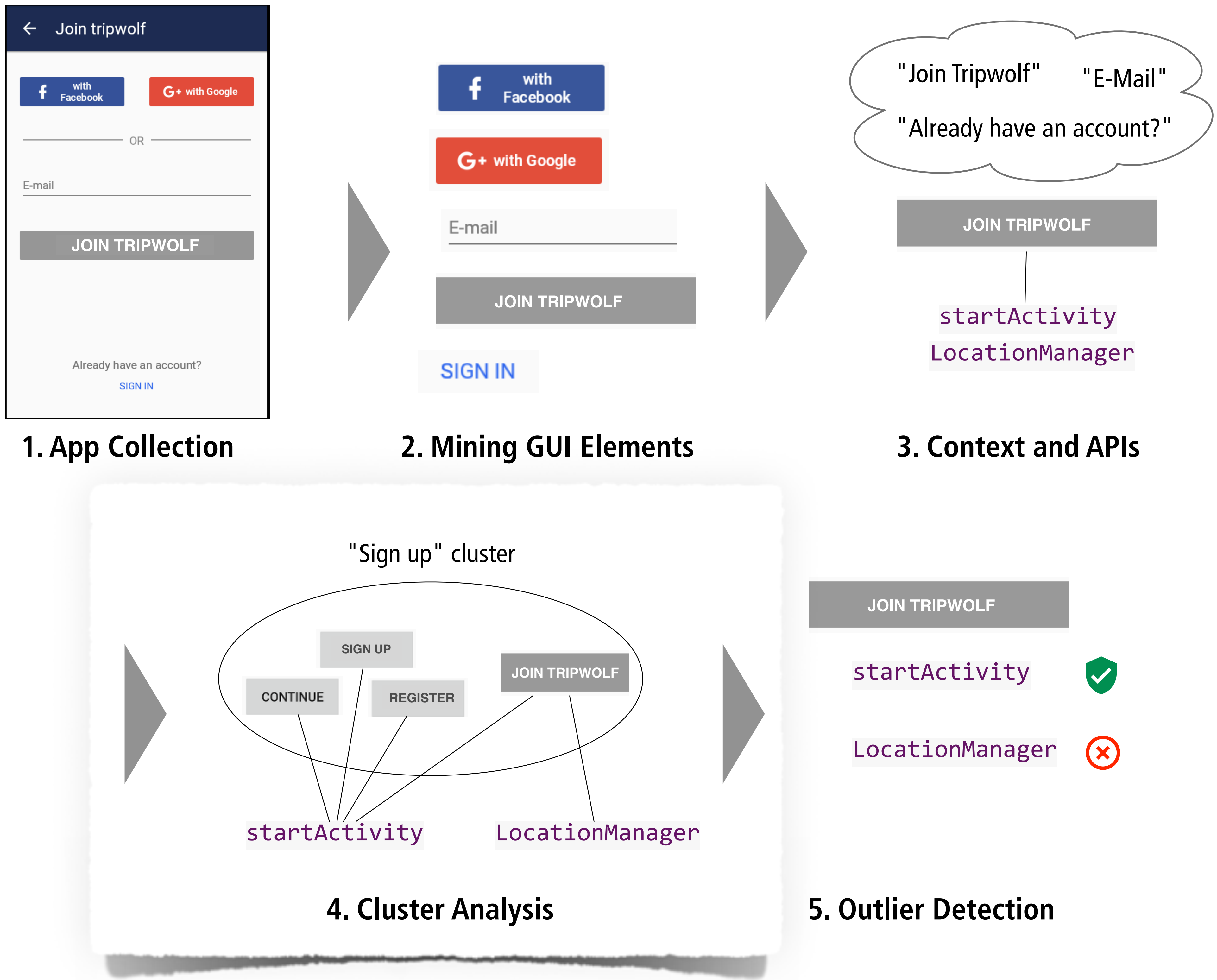
- Extracted all labels from all UI elements
- Static analysis includes labels set dynamically
- Extracted all text from surrounding screens (activities)

Backstage



Join tripwolf
with Facebook
with Google
OR
E-mail
JOIN TRIPWOLF
Already have an account?
SIGN IN
1. App Collection
with Facebook
with Google
E-mail
JOIN TRIPWOLF
SIGN IN
2. Mining GUI Elements
"Join Tripwolf"
"E-Mail"
"Already have an account?"
JOIN TRIPWOLF
startActivity
LocationManager
3. Context and APIs
"Sign up" cluster
SIGN UP
CONTINUE
REGISTER
JOIN TRIPWOLF
startActivity
LocationManager
4. Cluster Analysis
JOIN TRIPWOLF
startActivity
LocationManager
5. Outlier Detection

Backstage



Join tripwolf
with Facebook
with Google
OR
E-mail
JOIN TRIPWOLF
Already have an account?
SIGN IN
1. App Collection
with Facebook
with Google
E-mail
JOIN TRIPWOLF
SIGN IN
2. Mining GUI Elements
"Join Tripwolf"
"E-Mail"
"Already have an account?"
JOIN TRIPWOLF
startActivity
LocationManager
3. Context and APIs
"Sign up" cluster
SIGN UP
CONTINUE
REGISTER
JOIN TRIPWOLF
startActivity
LocationManager
4. Cluster Analysis
JOIN TRIPWOLF
startActivity
LocationManager
5. Outlier Detection

Semantic Distance

- Clustering by LDA (as in CHABADA) failed: Too little text on mobile UIs
- Took us a year to realize that
- Instead, now use *semantic distance* using Google Words2Vec
- Words2Vec trained from 100 billion words

Clustering

- Used k-means to cluster all labels into 250 concepts with low semantical distance

abort · about · accept · account · account info · achievement · activate · activation · activity · add · add content · add email · add list · add photo · address · admin · agree · agreement · album · alert · alphabet · amazon · amount · answer · app · apply · appointment · apps · architecture · archive · attach · audio · authenticate · authorize · average · baby · back · background · backup · badge · bangalore · barbie · barcode · baseball · bath · beauty · bedroom · begin · birth · block · bluetooth · board · broadcast · build · business · buy · bypass · cache · calculator · calendar · call · calorie · camera · campus · cap · card · cardio · career · celsius · challenge · change · chapter · chart · check · checkout · cheer · choose · city · claim · clean · clear · click · clock · cloud · code · colombia · come · comment · commentary · connect · contact · continue · contribution · coupon · cpu · create · create account · credit · credit card · custom · customer · customize · cycle · data · day · deal · debug · decline · default · delete · demo · departure · deposit · description · desire · destination · detail · device · dictionary · do · download ·

abort · about · accept · account · account info · achievement · activate · activation · activity · add · add content · add email · add list · add photo · address · admin · agree · agreement · album · alert · alphabet · amazon · amount · answer · app · apply · appointment · apps · architecture · archive · attach · audio · authenticate · authorize · average · baby · back · background · backup · badge · bangalore · barbie · barcode · baseball · bath · beauty · bedroom · begin · birth · block · bluetooth · board · broadcast · build · business · buy · bypass · cache · calculator · calendar · call · calorie · camera · campus · cap · card · cardio · career · celsius · challenge · change · chapter · chart · check · checkout · cheer · choose · city · claim · clean · clear · click · clock · cloud · code · colombia · come · comment · commentary · connect · contact · continue · contribution · coupon · cpu · create · create account · credit · credit card · custom · customer · customize · cycle · data · day · deal · debug · decline · default · delete · demo · departure · deposit · description · desire · destination · detail · device · dictionary · do · download · draw · edit · edit account · editor · electron · email · enable · enter · error · examination · execute · export · facebook · fax · feedback · fiction · file · fill · find · folder · follower · friend · gallery · google play · handoff · health · hello · image · import · information · install · instrument · internet · invoice · itinerary · jupiter · keyboard · launch · league · license · list · location · log · login · map · meal · merge · message · mild · mode · news · next · notification · ok · open · order · panorama · password · payment · paypal · people · permission · phone · photo · picture · play · please · power usage · premium · prev · price · privacy · profile · project · projector power · pushup · quiz · redeem · register · reminder · report · reset · retry · roster · rule · save · save account · scan · scanner · search · send · setting · share · shopping · show · shutter · skip · sms · space · stay · store · submit · subscription · sync · taxi · term · test · theme · ticket · tip · title · twitter · unlock · update · upgrade · upload · url · user · vehicle · vehicle · version · view · virus · voice · wallpaper · website · weight · workout · zone

Keyboard · launch · league · licen
erge · message · mild · mode ·
anorama · password · payment ·
icture · play · please · power usa
roject · projector power · pushup
eset · retry · roster · rule · save · s
tting · share · shopping · show · sh
subscription · sync · taxi · term ·
update · upgrade · upload · url · u
ice · wallpaper · website · weight

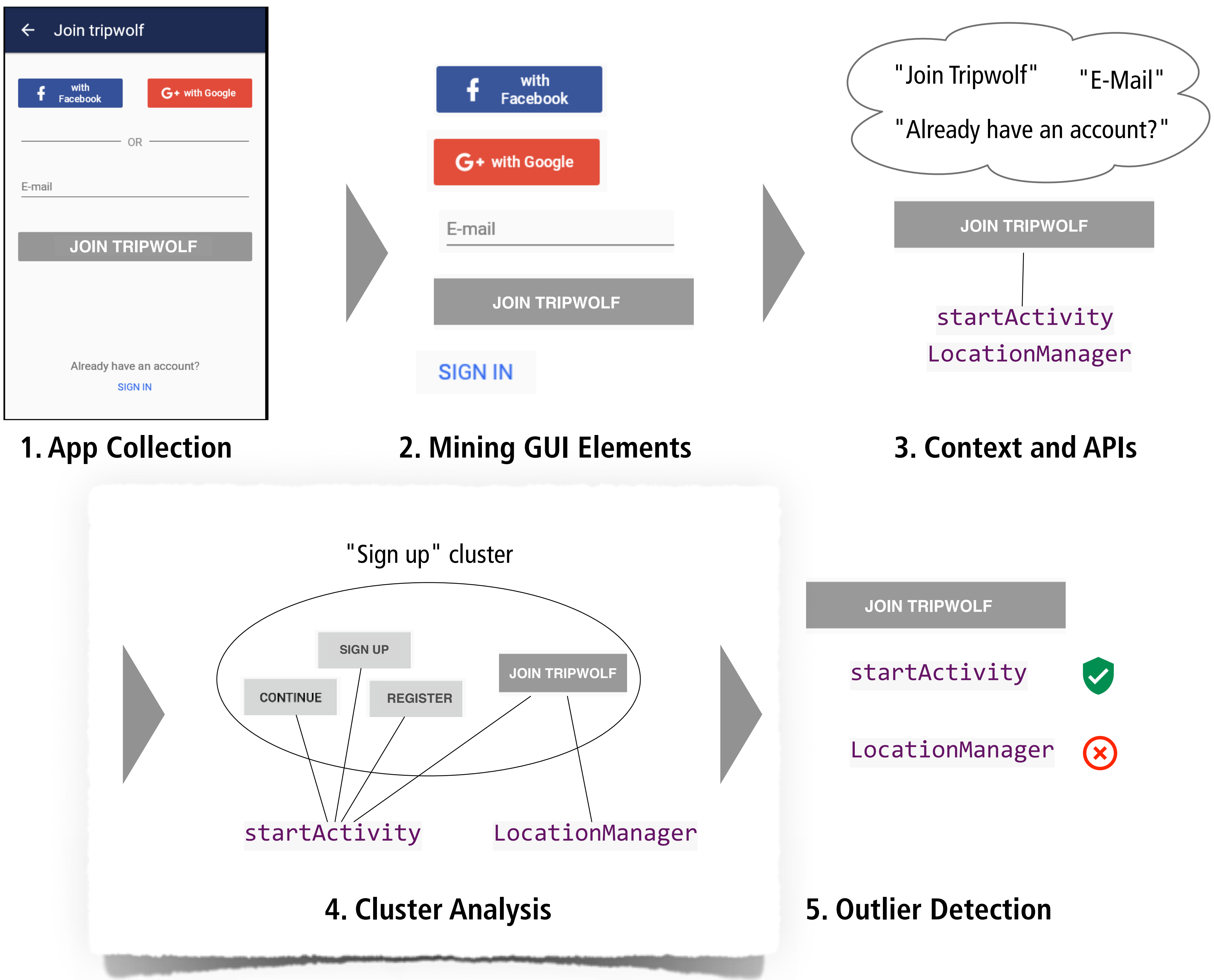
Share


ola
mall
net
kid
moment
get
beer
deal
placemarks
screen
web
invite
setting
post
video
friends
from
favorites
thread
find
reading
clipboard
share
bookmark
saving
number
horoscope
flight
place
hotspot
plan
event
goal
gallery
bible
new
mail
circle
result

Shopping

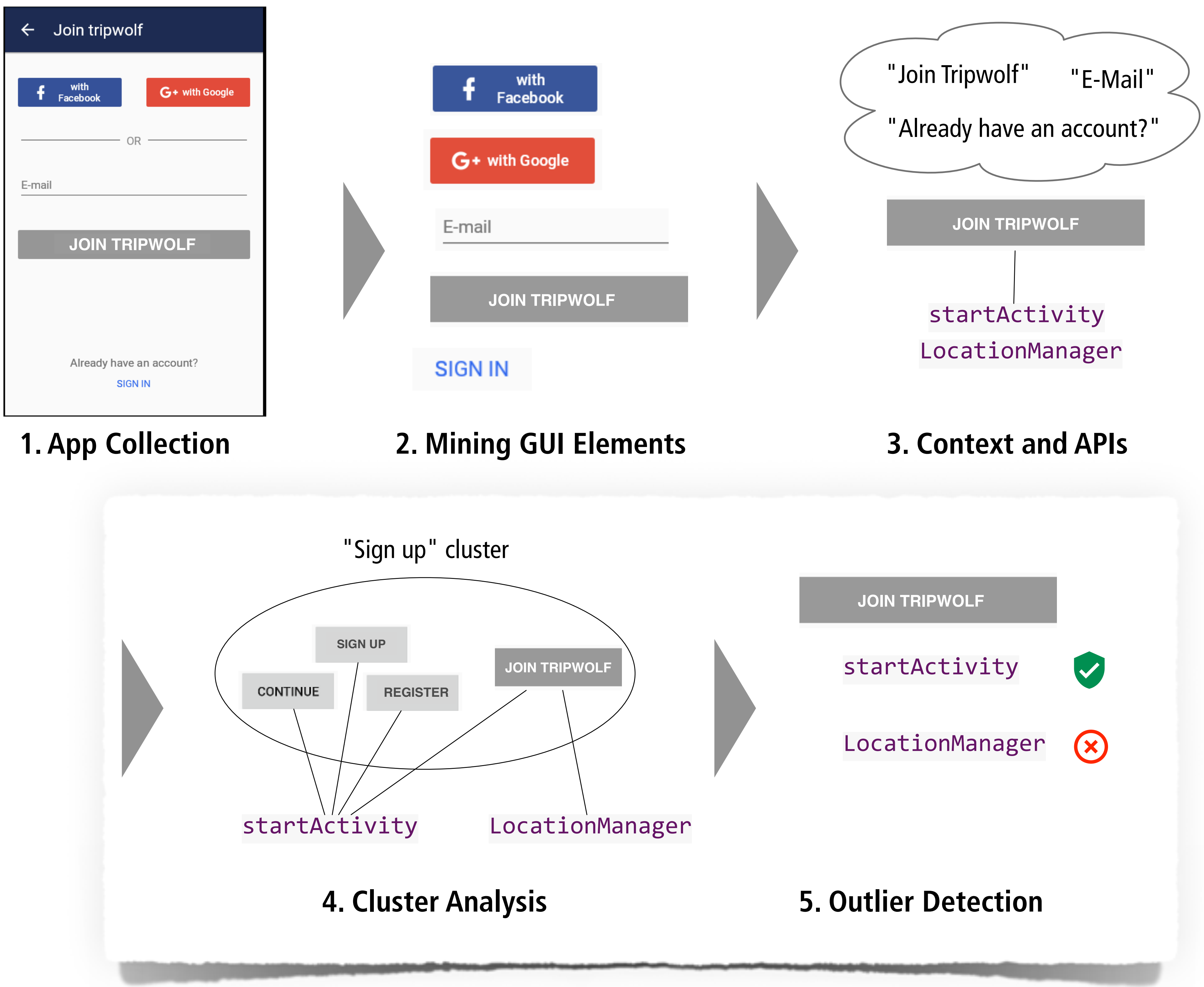
mart
stein
harris
sears
shopping
home
bed
macy
bath
start
office
safeway
gap
hancock
list
keep
save
buy
hhgregg
continue
family
bring
trader
shopper
dollar
general
joe
teeter
yes
about
sprout
market
amazon
stop
shop
target
depot
ace
walmart
beyond
cvs
go
old
basket
winn
kmart
lowes
cart

Backstage

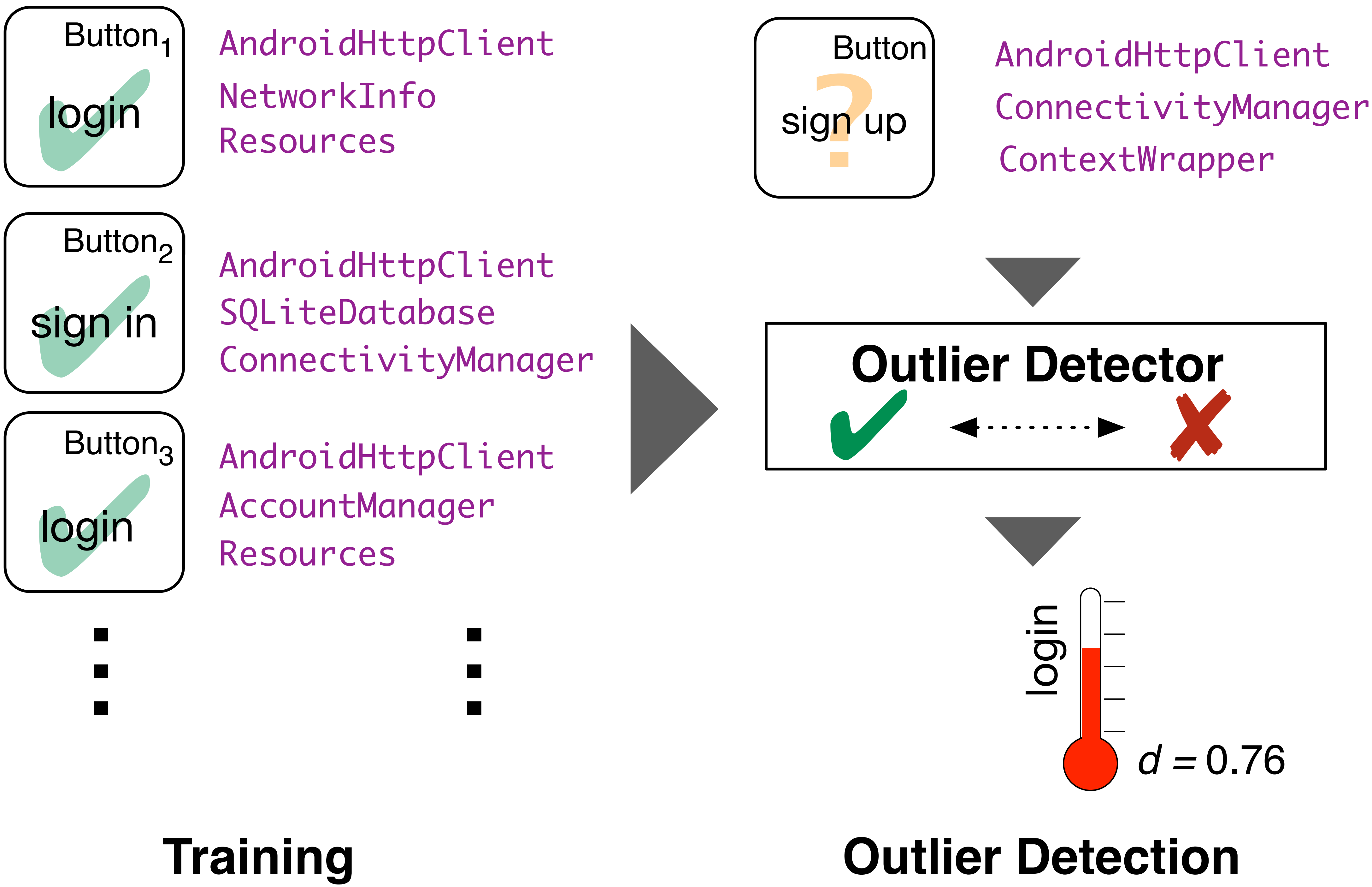


Join tripwolf
with Facebook
with Google
OR
E-mail
JOIN TRIPWOLF
Already have an account?
SIGN IN
1. App Collection
with Facebook
with Google
E-mail
JOIN TRIPWOLF
SIGN IN
2. Mining GUI Elements
"Join Tripwolf"
"E-Mail"
"Already have an account?"
JOIN TRIPWOLF
startActivity
LocationManager
3. Context and APIs
"Sign up" cluster
SIGN UP
CONTINUE
REGISTER
JOIN TRIPWOLF
startActivity
LocationManager
4. Cluster Analysis
JOIN TRIPWOLF
startActivity
LocationManager
5. Outlier Detection

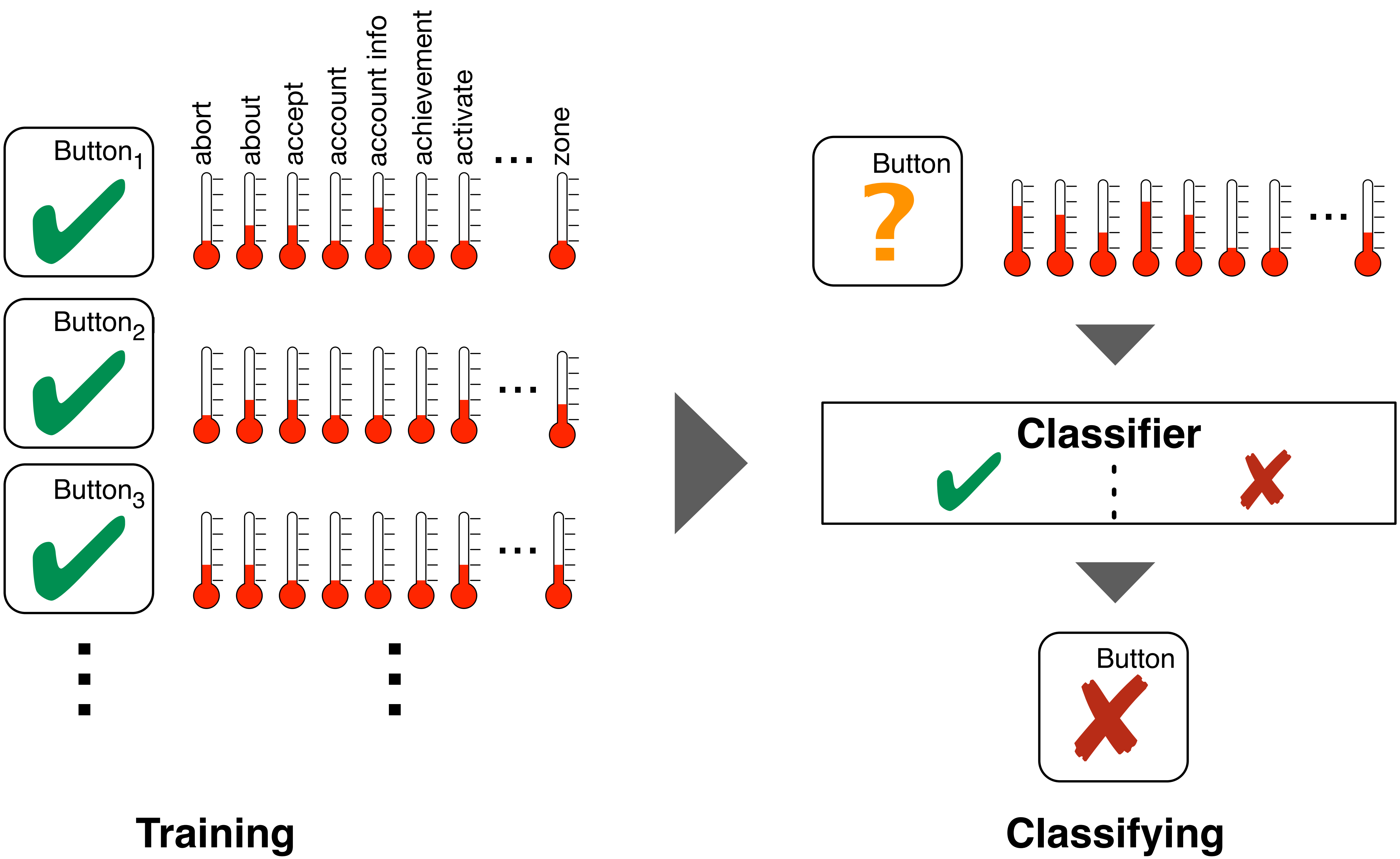
Backstage


Join tripwolf
with Facebook
with Google
OR
E-mail
JOIN TRIPWOLF
Already have an account?
SIGN IN
1. App Collection
with Facebook
with Google
E-mail
JOIN TRIPWOLF
SIGN IN
2. Mining GUI Elements
"Join Tripwolf"
"E-Mail"
"Already have an account?"
JOIN TRIPWOLF
startActivity
LocationManager
3. Context and APIs
"Sign up" cluster
SIGN UP
CONTINUE
REGISTER
JOIN TRIPWOLF
startActivity
LocationManager
4. Cluster Analysis
JOIN TRIPWOLF
startActivity
LocationManager
5. Outlier Detection

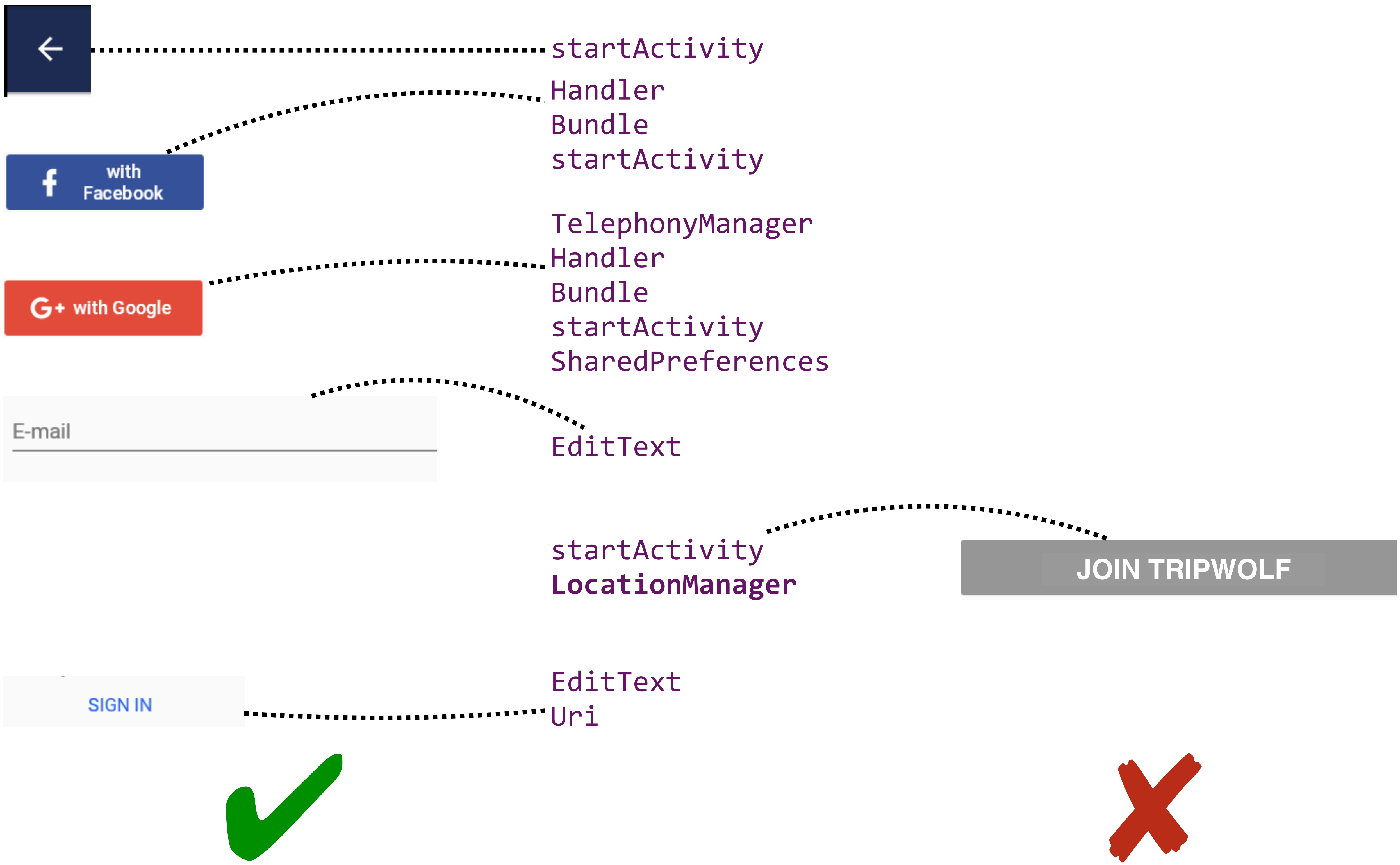
Classifying per Context



Overall Anomalies



Outliers

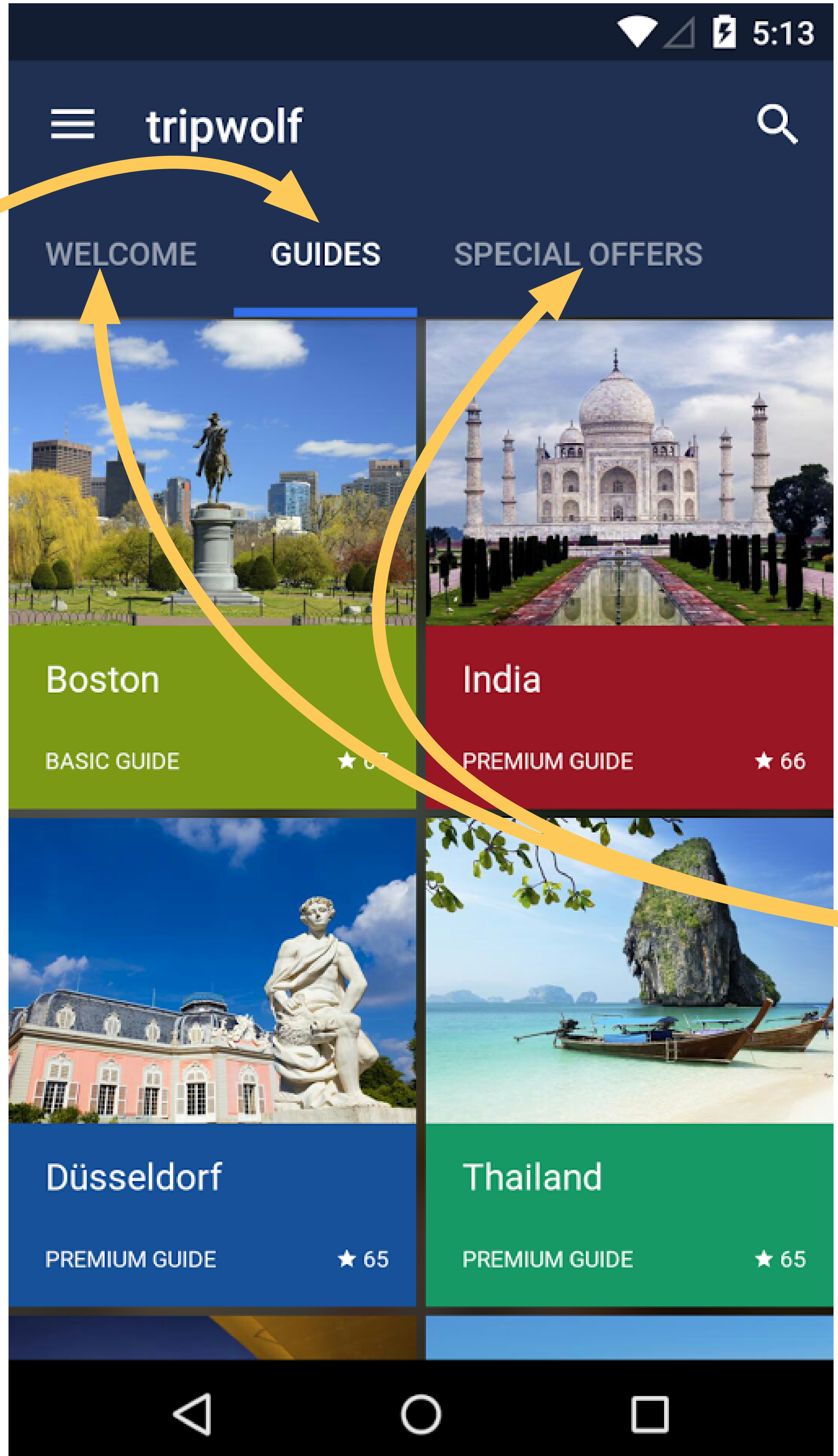


startActivity
Handler
Bundle
startActivity
with Facebook
TelephonyManager
Handler
Bundle
startActivity
SharedPreferences
with Google
E-mail
EditText
startActivity
LocationManager
JOIN TRIPWOLF
EditText
Uri
SIGN IN

Evaluation

How well does Backstage discover UI anomalies?

“Label replace” mutation:
assign a GUI element a different label – e.g. “Guides” is replaced by “Open” or “Print”



“Label crossover” mutation:
swap labels of two GUI elements – e.g. “Welcome” gets “Special Offers” label and vice versa

Results

Label Replace Mutations

	Classified as		
Input	**Abnormal**	**Normal**	**Total**
Mutant	TP = 3369	FN = 1630	4999
Correct	FP = 1100	TN = 4056	5156
Total	4469	5686	10155

Precision = 75%
Recall = 67%
Accuracy = 73%
Specificity = 79%

Label Crossover Mutations

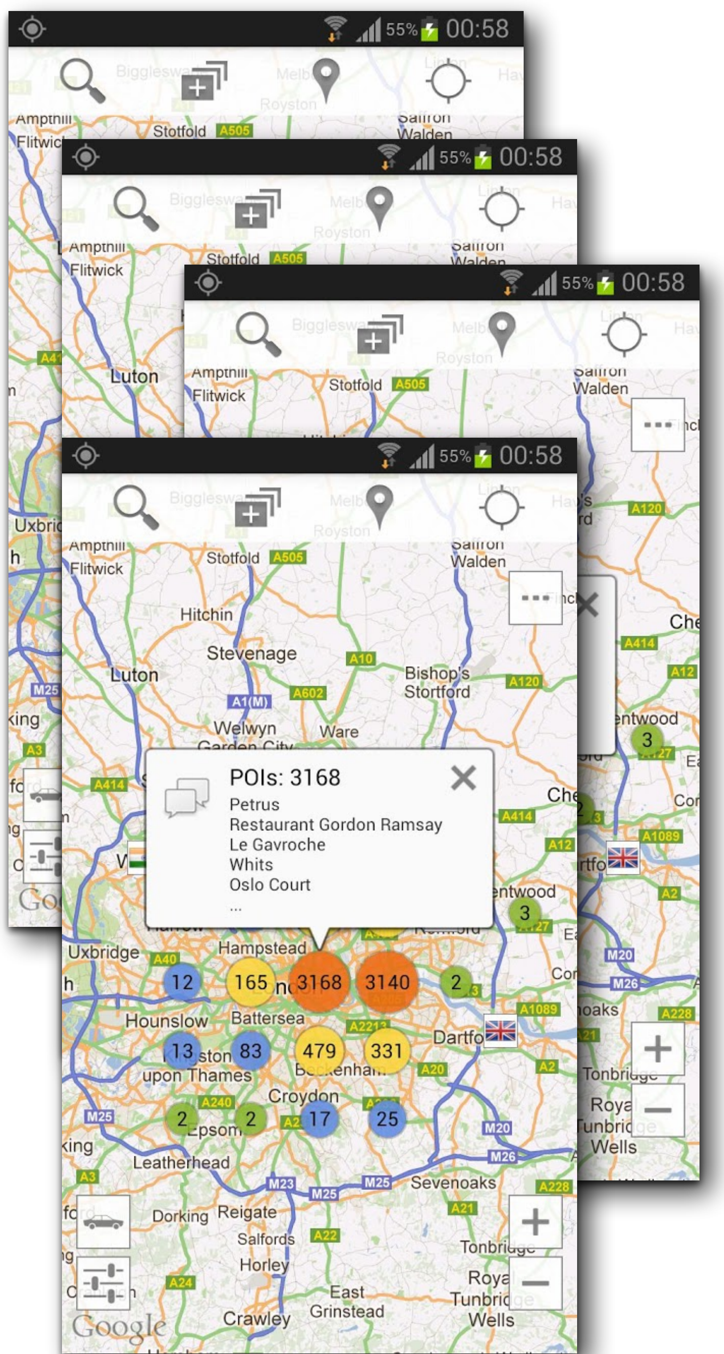
	Classified as		
Input	**Abnormal**	**Normal**	**Total**
Mutant	TP = 2290	FN = 2475	4765
Correct	FP = 1026	TN = 4121	5147
Total	3316	6596	9912

Precision = 69%
Recall = 48%
Accuracy = 65%
Specificity = 80%

Figures

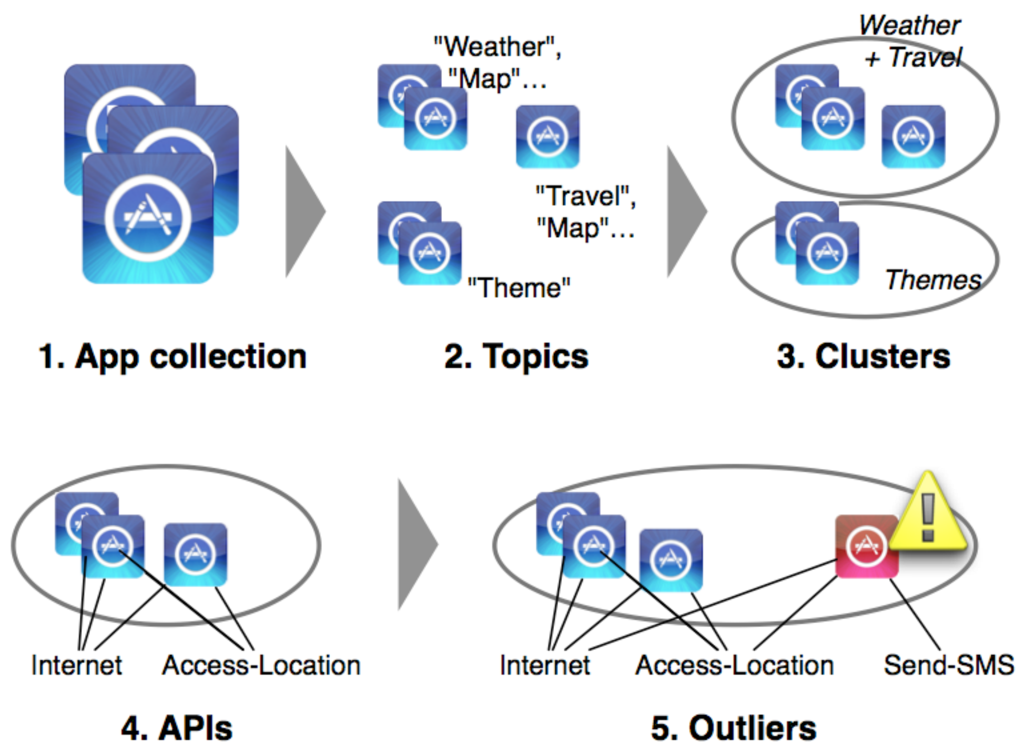
- Backstage detects abnormal UI elements with an accuracy of 73–75%
- First machine learning approach to detect UI anomalies
- Mined 87,100 UI Elements in 12,000 apps
- 5 GB data set publicly available, with UI elements, labels, context, APIs…

App Mining



- For 100,000s of apps:
- Gather *descriptions*
- Gather *metadata*
- Gather *code and UI features*
- Find what is *common* and what is *uncommon*

CHABADA

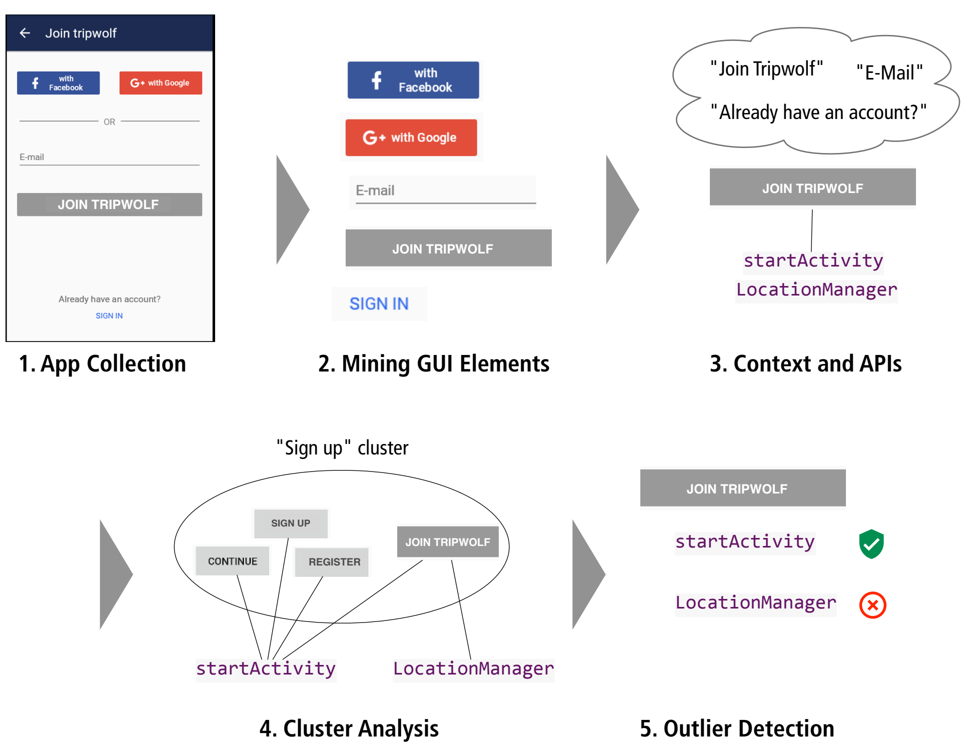


Travel Cluster

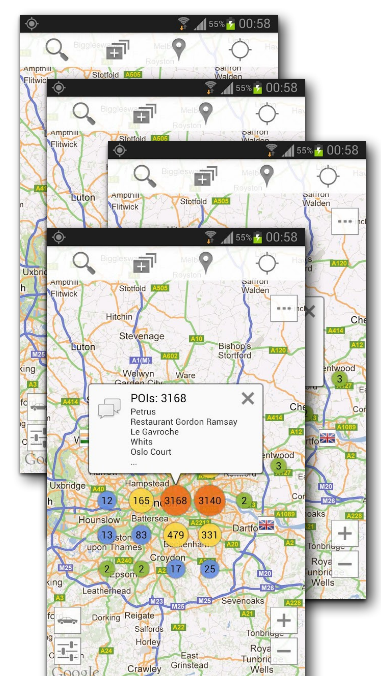
Description

APIs used

Backstage



App Mining



- For 100,000s of apps:
- Gather *descriptions*
- Gather *metadata*
- Gather *code and UI features*
- Find what is *common* and what is *uncommon*

http://www.st.cs.uni-saarland.de/appmining/

erc Google