In Pursuit of a Perfect App Search Engine

From Generic App Search Queries to Intent Discovery and a New Way to Classify Apps

Marcin Rudolf
CTO @ Xyologic
Overview

1. A comparison of apps as object of IR with web pages, books & music

2. Key evidence:
   - Most of users’ queries are very general or just asking for inspiration
   - Apps are related to each other and strongly grouped into classes
Overview

3. How should a perfect app search engine work?

- A new approach: Xyologic uses custom machine learning algorithms to derive hundreds of app classes.

- How to identify the best app in an app class: Rank apps by popularity and quality signals, internationalization of app search.

- „Look Inside The App”
The Essence of an App

- An app is a piece of software that carries on a very specific action in a short time.
- An app is defined by what it does.
- Informational content of the app is not defining what the app is.
- IR systems however work only on information representation of an app.
Apps vs. Books

• Books
  – The essence: information (non-fiction), emotions/pleasure (fiction)
  – Representation: title, author, summary, publisher, table of content, references
  – Social components: reviews, editorials, ratings, citations
  – Distribution: bookstore, library, social
  – Discovery: catalogue, top lists, viral
Apps vs. Music Tracks

• Music
  – The essence: emotions/pleasure, belonging to genre
  – Representation: genre, composer, title, cover art.
  – Social component: reviews, covers, sampling
  – Distribution: copying, broadcast, record stores
  – Discovery: viral, broadcast, top charts
Apps vs. Websites

- Websites (hypertext)
  - The essence: information, linked
  - Representation: text and links
  - Social component: none
  - Distribution: none
  - Discovery: search box
What is Specific About the Apps?

- Apps are defined by what they do. This is indeed a unique new kind of entity!
- An app’s representation is very similar to the one of books. This is mainly because they are distributed in a similar way (app store vs book store).
- Apps are strongly „genreified”, similarly to music.
Apps as Information

• Basic meta data: provided by publishers
  – Title, description, publisher, version info, icon, screenshot, category, video

• Popularity and quality data
  – Downloads, popularity trends, various “flags” (like “price drop”, “spam”)

• Social data
  – Ratings, reviews
Massive Number of Apps Are Installed Monthly

Average smartphone user downloads **8 apps per month.**

- 26.6B downloads (1.45B in March 2012)
- 15.9B downloads (1.7B in March 2012)
- 230M downloads (25M in March 2012)
The Demise of the Leaderboards and Catalogues

- Too many apps for a catalogue. Google Play and Facebook removed the catalogue.
- Google Play does not display apps with zero downloads at all, except in search.
- Top lists lead to a situation where less than 0.1% of the best apps get more than 50% of the downloads.
Leaderboards Do Not Work As Discovery Tool

Top positions in the app stores become increasingly fixed

<table>
<thead>
<tr>
<th>Platform</th>
<th>Feb 2011</th>
<th>Feb 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 25 on Android</td>
<td>8 new apps</td>
<td>1 new apps</td>
</tr>
<tr>
<td>Top 25 on iPhone</td>
<td>4 new app</td>
<td>0 new apps</td>
</tr>
</tbody>
</table>

Age of apps in free top 25, USA. Xyologic Monthly Downloads ranking

Over 55% of apps in Top 100 List are older than half a year!

<table>
<thead>
<tr>
<th>Platform</th>
<th>Age: &lt; 30 days</th>
<th>Age: 30 – 90 days</th>
<th>Age: 90 – 180 days</th>
<th>Age: &gt; 180 days</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 100 Android</td>
<td>9</td>
<td>41</td>
<td>14</td>
<td>78</td>
<td>142</td>
</tr>
<tr>
<td>Top 100 iPhone</td>
<td>31</td>
<td>62</td>
<td>46</td>
<td>170</td>
<td>309</td>
</tr>
</tbody>
</table>

Age of apps in free top 100 list, March 2012, USA. According to Google Play and iTunes ranking
A Long-Tail of Apps That Never Get Discovered

Less than 0.1% of Apps Generates More Than 50% of Monthly Downloads in Dec 2011
A Long-Tail of Apps That Never Get Discovered

What it really looks like...

![Graph showing downloads in Dec 2011 against position in monthly top list for Google Play and Apple Appstore.](graph.png)
How Users Search For Apps

Xyologic conducted a trial with a large mass-market online portal in Europe. Through the integration of Xyologic app search engine 2,000,000 app search queries were generated by the portal users. A sample of 1000 queries was analyzed manually.

### How Users Search For Apps

- **General category queries - 80%**
  - Music
  - Movies
  - Chat

- **„Inspire me” queries - 15%**
  - Games
  - Free apps
  - fun

- **Specific action queries - 5%**
  - Crop photos
  - Block calls
  - View movies
How Users Search For Apps

Top 25 Search Queries Reported by Chomp in June 2011:

2 “Inspire me” queries

23 General category queries

0 Specific action queries

http://chomp.com/etc/chomp-charts/jun-2011
How Users Search The Web - A Learned Behavior

After 12 years of using a search box, people still ask simplest queries: 80% names 20% things

Hot Searches (USA)

Apr 13, 2012 - change date
Updated 53 minutes ago

1. ann romney
2. north korea
3. modesto
4. gawker
5. north korea missile launch
6. selena
7. shea weber
8. pat burrell
9. san jose sharks
10. chicago blackhawks
11. jennifer hudson
12. bulls
13. axl rose
14. the three stooges
15. tax deadline
16. st louis blues
17. phoenix coyotes
18. cabin in the woods
19. esperanza spalding
20. chicago bulls

www.google.com
The Bottom Line

• **Most** users just want to be inspired: they want to find a new cool app or game.
• **Some** users look for general app classes.
• A **small** minority looks for a specific function. Those are power users.

To return satisfactory results, a search engine for apps must be able to deal with vaguely expressed intents and very general queries.
How Apps Are Related

• Similarity metrics: apps are defined by what they do. Two sources of that information: category and description.

• Native grouping in about 30 categories is too general to be useful.

• Xyologic uses custom machine learning algorithms to derive hundreds of app classes that apps are assigned to.
Categories vs. Classes: Music

Download & Play Music
- Music Download
- Mp3 Search and Download
- Spotify
- Shazam
- MP3 Music Hits
- MOG Mobile Music
- Winamp

Ringtones
- Most popular Ringtone
- SMS Ringtones Top50
- Ringtone Maker
- MP3 Ringtone Editor
- Ringtones
- Tune Me Lite
- Ringtone Architect

Play Music & Radio
- Pandora® internet radio
- Slacker Radio
- Pandora® radio for Google TV
- Jango Radio

Virtual Instruments
- My Piano
- Drum Kit
- Su-Preme MPA
- Mobile Metronome Pro
- Electrum Drum Machine/Samp
- G-Stomper - Drum Machine
- RD3 HD - Groovebox

Sound & Effects & DJ
- Music Volume EQ
- Poweramp Music Player (Trial)
- Equalizer
- DJ Studio 3
- myMusicTagger
- Equalizer Unlock Key
- projectM Music Visualizer
- ZPlayer
Categories vs. Classes: Productivity

- **Task Managers**
  - GO TaskManager Widget
  - Advanced Task Manager
  - Advanced Task Killer
  - GO Power Master
  - Easy Uninstaller
  - Android Booster FREE
  - Advanced Task Killer Pro

- **Custom Launchers**
  - GO Locker
  - GO Locker Fourkey Theme
  - GO Locker ICS Theme
  - GO Locker Sense Theme
  - Touch Theme GO Locker
  - GO Locker 3D Worm Theme
  - GO Clock Widget

- **Making Notes**
  - Evernote
  - Sketch
  - ColorNote Notepad Notes
  - Catch Notes
  - S Memo
  - Classic Notes Lite + App Box
  - Handrite Note Free

- **Anti – Vir & Theft**
  - Snap Secure + Free Anti Virus
  - NQ Mobile Security & Antivirus
  - NetQin Security & Anti-virus 5.0
  - Anti-Virus Pro - Tablet

- **Hack & Backup**
  - App Backup & Restore
  - Graph 59
  - ROM Toolbox
  - Speed Boost System Cleaner
  - ICS Launcher
  - Ultimate Backup Pro
  - Android 2.3 Launcher (Home)
Substitutive vs. Complementary Apps

- Substitutive
  - Important for Recommendations
  - Important for Re-Targeting of Apps’ Advertisement

- Complementary
  - App Backup & Restore
  - Snap Secure Free Anti-Virus
  - Advanced Task Manager
  - TeamViewer for Remote Control
  - Dropbox
  - Adobe Reader
Conclusion: Apps Are Segmented in Classes

• How precise is the segmentation?
  – Apps can be assigned to many classes, to each with different weight. However, two most important classes app is assigned to, get per average 75% of the weight.
  – Apps are quite uniformly distributed among classes.
The Starting Point for App Search

• A typical user’s query to handle is *a query with a weakly stated intention leading to query results with many mixed app classes.*

• Our approach is to assume this user behaviour and adapt the app search technology to the *most typical use cases.*
Query Examples: "cars"

**Games**

- **Car Brands**
  - Puzzle Cars 4 Jigsaw
  - Pimp My Reseller
  - Name That Car
  - Cars Memory

- **Racing Games**
  - Drag Racing
  - Speed 3D Racing Car
  - Need Warrior Top
  - Super race cars

- **Arcade Games**
  - Top Truck Free
  - Car Wash
  - Unblock My Car
  - Jet Car Stunts Lite

**Apps**

- **Car Brands**
  - Car Gallery
  - Rally HD Wallpapers
  - Muscle Cars HD Wallpapers
  - Drift HD Wallpapers

- **Car Utilities & Locators**
  - MyCar Locator Free
  - Car Tunes Music Player
  - Edmunds
  - Find My Car

- **Fuel Logs**
  - FuelLog - Car
  - Car Maintenance
  - Gasonomics
  - uCar
A Perfect Search Engine For Apps

The Algorithm

1. Capture initial intent and lead the user to the correct class of apps
2. Rank the apps to find the best app in class using popularity and quality signals
3. Make the choice easy for the user: show her what the app does and how good it is. („Look inside the app“)
Capturing Intent

**Initial Intent**
- General Users Query: “Sleep”

**Intent Narrowed**
- Classify apps in results: “relaxation”, “sleep & sound”, lullabies”, “sleep clocks & util”, “meditation”, “hypnosis”

**Intent Verified**
- User chooses class. We narrow the intent. User explores filtered results that are augmented with the best apps in class

**Repeat**
- Drilldown if further grouping is possible
Can We Give Up The Search Box?

• When search box gets in the way? „Inspire me” queries are 15% of all!
• How to capture user’s initial intent?
  – Search box is still the best way if you know nothing about the user.
  – Already installed apps
  – Interests graphs from social networks
Can We Give Up The Search Box?

- App discovery is complex and not straightforward. This is contrary to hypertext where search box works perfectly.

- Different app classes are consumed in different ways.

- A search box will play a role but will not be central to apps’ discovery.
How Users Find Games?
Example: Social Discovery of Apps
How To Identify Best Apps in Class

**Popularity signals:**
Downloads of apps and publishers: Tells you install base and how fast it grow.

**Quality signals:**
Apps’ ratings.
Sentiments extracted from users’ reviews.
Publisher quality
Popularity Signals

Downloads are important measure in app economy. Free reports at [http://xyologic.com](http://xyologic.com)
Popularity Signals

Downloads Distribution, Paid Apps, Google Play, March 2012

- Best Fitting Zipf’s Distribution
- Real Download Values
Local Trends Are Important

Within many app classes, different apps are popular in different countries

Query: „best local news apps” on iPhone

Great Britain | Germany | Japan | Poland
---|---|---|---
Sün | Bild | GCI | 02 NEWSY
OK! | Welt | Nike | O2
BBC NEWS | Welt | Haiku | 14
BBC NEWS | News Feeds | Haiku | Wprost
Search Internationalization

- Crawl and index app stores in all languages. Apps are a parallel corpora.
- Detecting languages of apps’ text data and language of the queries. Tokenizers and stemmers.
- Access all apps by using any query language.
- Use language-agnostic algorithms.
Quality Signals

- We use customized machine learning algorithms to extract sentiments towards apps that are present in reviews.
- There are many positive and many negative types of sentiments. Love vs. hate.

<table>
<thead>
<tr>
<th>Sentiment</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>a rip off</td>
<td>49%</td>
</tr>
<tr>
<td>very limited</td>
<td>27%</td>
</tr>
<tr>
<td>not worth downloading</td>
<td>24%</td>
</tr>
<tr>
<td>very addictive</td>
<td>43%</td>
</tr>
<tr>
<td>good for kids</td>
<td>17%</td>
</tr>
<tr>
<td>worth the price</td>
<td>11%</td>
</tr>
<tr>
<td>best app ever</td>
<td>11%</td>
</tr>
</tbody>
</table>
Quality Signals

- Detecting junk apps and publishers of spam.
- "Flag signals": price drops, new apps, fast climbers.
- Heuristic algorithms different for each platform.
Look Inside The App

Representation must correspond to reality, otherwise user satisfaction will not be delivered.

Challenge: How to represent an app in UI so user knows what it does and how good it is?
Look Inside The Book

What the book is about at a glance:

• Table of contents
• Index
• Excerpts, Summary
• Search Inside: look for interesting terms in context
Look Inside The Music

Feel the music:
- Quick preview
- Learn what people feel when listening at specific moments
- Share your feelings
- Great for a specific audience
Present What an App Does

• Description. Excerpts
• Class of application
• Screenshots
• Videos
Present Qualitative Information

- Downloads and changes in popularity.
- Handling thousands of reviews: sentiments and in-topic reviews.

- Find what people talk about the app in social media
Outlook: Search and Recommendation Merged in One Experience

What’s typical for recommendations:
- Our engine handles very general and simple queries. Users ask us for inspiration, not for information.
- We narrow down users’ intent by letting them explore various app classes they may be interested in.
- We are able to augment search results with additional relevant apps.

What’s typical for search engine:
- We still maintain multilingual index of apps and retrieve them using keywords and app class information.
- We rank apps using popularity and quality signals.
- We display apps as a “search result snippet”: “Look inside the app”
THINK SIX IMPOSSIBLE APPS BEFORE BREAKFAST

www.xyologic.com