Available Google Play Store crawlers/scrapers report

Dan Plămădeală

February 7, 2020

1 Premise

Given the goal to collect information and download the apk file of the applications on the Google Play store, an obvious step was to first research the available crawlers. To do that, I tried to find as many available crawlers and tools as possible. From the compiled list, I would later choose a crawler that fulfils our requirements or is very close to doing that by requiring only slight changes to its functionality. Some of the resources that I have used for compiling the list of crawlers include GitHub.com, GitLab.com, code.google.com, NPMJS.com, PyPI.org, etc. The main attributes of a crawler, upon which I based my so-called literature review are the programming language used, the last time it was updated, the available functionality, the ease of use, the collected data et cetera.

2 List of crawlers

In Table 1 I have showcased a short summary about the crawlers/scrapers and tools that I was able to find that are related to our scope.

	Programming language	Last updated	Functionality	State	Link
google-play-scraper	Node.js	18 october 2019	Scraper	Working	[1]
google-play-crawler	Java	8 february 2014	Search & Download	Not working	[2]
googleplay-api@1	Python 2	8 december 2012	Search & Download	Not working	[3]
googleplay-api@2	Python 2-3	19 june 2019	Search & Download	Not working	[4]
android-market-api	Java	22 february 2012	Limited info	Not working	[5]
android-marketplace-crawler	Ruby	15 april 2011	Crawler	Not working	[6]
play-scraper	Python 2-3	15 january 2019	Scraper	Not working	[7]
node-google-play	Node.js	20 november 2019	Scrape & Download	Working	[8]

Table 1: Compiled list of crawlers/tools

3 Details

3.1 google-play-scraper

Programming language/Framework: Javscript/Node.js
Last updated: 18 october 2019
Available functionality:

1

- Retrieves the full detail of an application.

- Retrieves a list of applications from one of the collections on Google Play.

- Retrieves a list of apps that results from a search query given a search term.

- Returns a list of all the applications published by a given developer name.

- Given a string, return up to five suggestions to complete a search query term.

- Retrieves a page of reviews for a specific application.

- Returns a list of similar apps to the one specified as an argument.

- Returns the list of permissions an app has access to.

- Retrieve a full list of categories present from the dropdown menu on Google Play.

- Provides the possibility to pause between requests to avoid getting banned by the Google Servers.

Return format: JSON
State: This tool has a rich functionality set, and it appears that it is fully working, considering that it is still maintained and in development. Given our scope, the only thing it is missing is the possibility to download APK files.
Link: `https://github.com/facundoolano/google-play-scraper`

3.2   google-play-crawler

Programming language/Framework: Java
Last updated: 8 february 2014
Available functionality:

- Downloads apk files.

- Retrieves a list of applications from one of the categories on Google Play.

- Retrieves a list of apps that results from a search query given a search term.

- Returns the list of permissions an app has access to.

- Retrieve a full list of categories present from the dropdown menu on Google Play.

- Registers a fake smartphone for accessing the Google Play data.

Return format: Text
State: This tool has a decent functionality set, but it is heavily outdated, and most of the protocol, authentication methods and data retrieval functions are not working because of the changes that were made to the Google Play store in the recent years.
Link: `https://github.com/Akdeniz/google-play-crawler`

3.3   google-play-api@1

Programming language/Framework: Python 2
Last updated: 08 december 2012
Available functionality:

- Downloads apk files.

- Retrieves a list of applications from one of the categories on Google Play.

- Retrieves a list of apps that results from a search query given a search term.

- Returns the list of permissions an app has access to.

- Retrieve a full list of categories present on Google Play.

- Retrieve a full list of subcategories.

- Return the list of apps in a subcategory.

Return format: Text, Protobuf
State: Although the API has good functionality, it has not been maintained for more than 6 years. It would most likely take a lot of effort and time to bring it back to life, so to say.
Link: `https://github.com/egirault/googleplay-api`

3.4 google-play-api@2

Programming language/Framework: Python 2-3
Last updated: 19 june 2019
Available functionality:

- Downloads apk files.

- Retrieves a list of applications from one of the categories on Google Play.

- Retrieves a list of apps that results from a search query given a search term.

- Returns the list of permissions an app has access to.

- Retrieve a full list of categories present on Google Play.

- Retrieve a full list of subcategories.

- Return the list of apps in a subcategory.

- AC2DM authentication with checkin and device info upload.

- Gives the possibility to choose the emulated device you're retrieving the data from.

Return format: Text, Protobuf
State: This is an updated fork of the previous tool (googleplay-api@1) with various improvements and some extra added features. After some minor changes and fixes, I managed to get it working. This might be the only tool that can download the apk file of an app.
Link: `https://github.com/egirault/googleplay-api`

3.5 android-market-api

Programming language/Framework: Java
Last updated: 22 february 2012
Available functionality:

- Provides the ability to browse the market with any given carrier or locale.

- Retrieves a list of apps that results from a search query given a search term or the package name.

- Retrieve the reviews of an app given an app ID.

- Get PNG screenshots and icon.

Return format: JSON
State: This API has a very limited functionality. Moreover, it is not working as it is the oldest tool I have found, thus it has not been updated to correspond to the new formats provided by the Google Play store.
Link: `https://code.google.com/archive/p/android-market-api/`

3.6   android-marketplace-crawler

Programming language/Framework: Ruby on Rails 3
Last updated: 15 april 2011
Available functionality:

- Provides the possibility to crawl apps based on a given category.

- Provides the possibility to crawl apps based on the given names of publishers.

- Provides the possibility to crawl apps based on the given list of package names.

Return format: Database
State: This crawler is based on the previous API (android-market-api). However, since it is based on the previous API, it hasn't been updated for a very long time, thus it isn't working at the moment and would require a complete overhaul to get it working again.
Link: `https://code.google.com/archive/p/android-marketplace-crawler/`

3.7   play-scraper

Programming language/Framework: Python 2-3
Last updated: 15 january 2019
Available functionality:

- Retrieves the full detail of an application.

- Retrieves a list of applications from one of the collections on Google Play.

- Retrieves a list of apps that results from a search query given a search term.

- Returns a list of all the applications published by a given developer name.

- Given a string, return up to five suggestions to complete a search query term.

- Retrieves a page of reviews for a specific application.

- Returns a list of similar apps to the one specified as an argument.

- Returns the list of permissions an app has access to.

- Retrieve a full list of categories present from the dropdown menu on Google Play.

Return format: JSON
State: Based on my research, this is a port of the first scraper (google-play-scraper), from Javascript to Python. It seems to be in an almost working state, with just a small inconvenience caused by Google's authentication methods.
Link: `https://code.google.com/archive/p/android-marketplace-crawler/`

3.8   node-google-play

Programming language/Framework: Javascript/Node.js
Last updated: 20 november 2019
Available functionality:

- Retrieves the full detail of an application.

- Retrieves a list of similar applications.

- Retrieves downloading information.

- Requests complete download info.

Return format: JSON
State: When I first compiled this list, this crawler wasn't working. After a few weeks the developers seems to have fixed it, thus this might be the scraper/crawler and tool that we're going to use for collection information and downloading the apk.
Link: `https://github.com/dweinstein/node-google-play`