



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Lecture 2.2 - Scraping Data



The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

How to Scrape?

Google Play example

Goal: collect the network of similar apps

The screenshot displays the Google Play Store interface for the Shazam app. The main section features the Shazam logo, the app title "Shazam - Discover Music", the developer "Shazam Entertainment Limited", the category "Music & Audio", a 4.5-star rating from 3,102,253 reviews, and a "Teen" age rating. It also indicates "Contains ads" and provides buttons for "Add to Wishlist" and "Install". Below this, three promotional banners are shown: "Name any song in seconds.", "Explore lyrics, videos & playlists", and a preview of the app's interface on a smartphone. To the right, a "Similar" section lists three recommended apps: SoundHound Music, TrackID™ - Music, and Musixmatch Lyrics. Each similar app entry includes its icon, name, developer, a brief description, a star rating, and a "FREE" price tag. A "See more" button is located at the top right of the similar apps section.

Shazam - Discover Music
Shazam Entertainment Limited Music & Audio ★★★★★ 3,102,253
Teen
Contains ads
Add to Wishlist Install

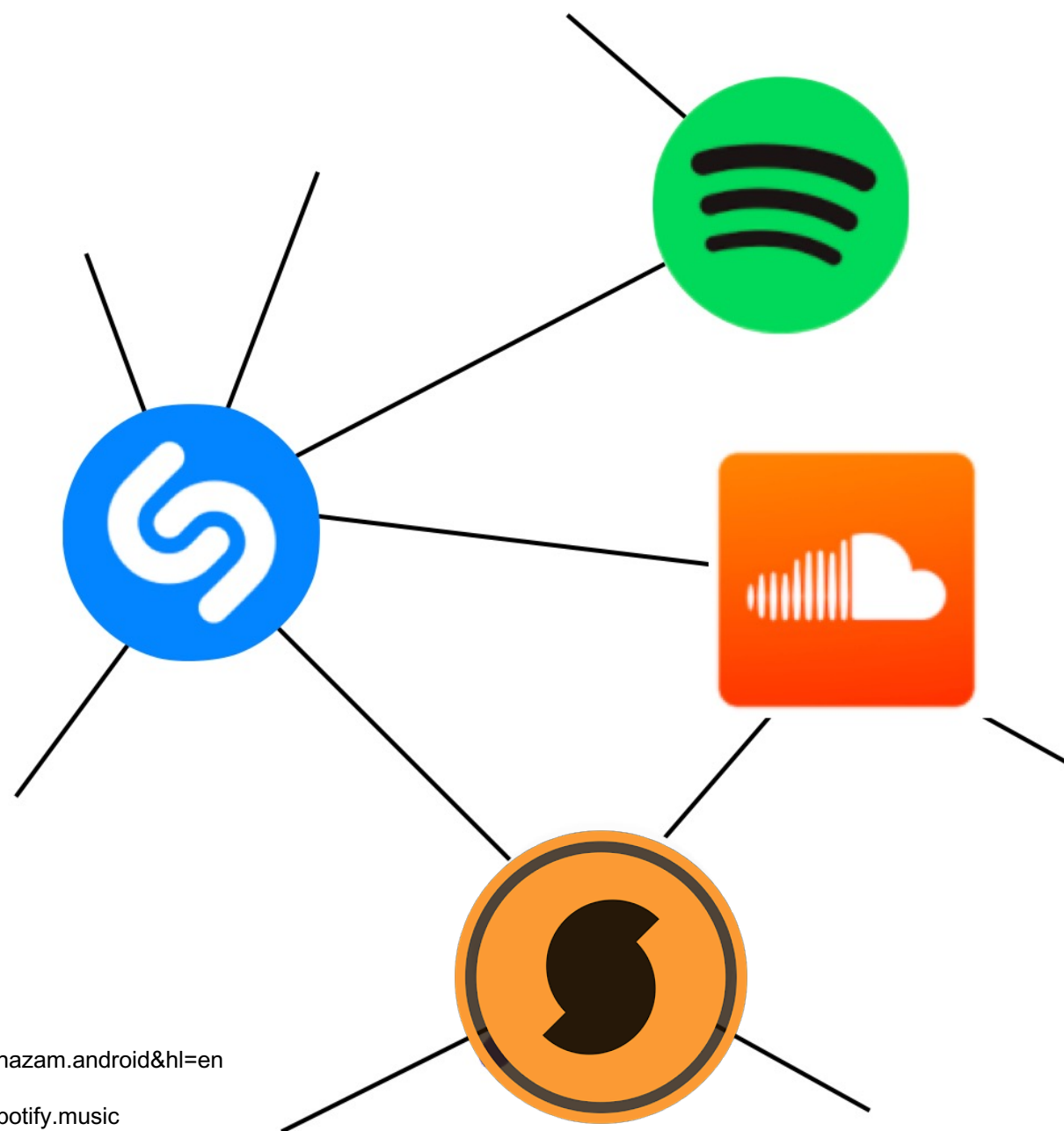
Similar See more

- SoundHound Music**
SoundHound Inc.
The popular music app with 300 million+ downloads globally!
★★★★★ FREE
- TrackID™ - Music**
Sony Mobile Communications Inc.
TrackID™ is the best way to identify the music playing around you.
★★★★★ FREE
- Musixmatch Lyrics**
Musixmatch
Enjoy lyrics for Spotify, Youtube and many other players
★★★★★ FREE

Retrieved: <https://play.google.com/store/apps/details?id=com.shazam.android&hl=en>

How to Scrape?

Goal: Write a **program/algorithm** to scrape Google Play to **collect a million-node network** of similar apps



Each **node** is an app

An **edge** connects two similar apps

Hint: start with some apps (e.g., Shazam), and go from there.

Icons from:

<https://play.google.com/store/apps/details?id=com.shazam.android&hl=en>

<https://play.google.com/store/apps/details?id=com.spotify.music>

<https://play.google.com/store/apps/details?id=com.soundcloud.android>

<https://play.google.com/store/apps/details?id=com.soundcloud.android>

How to Scrape?

Google Play example

Goal: collect the network of similar apps

<https://play.google.com/store/apps/details?id=com.shazam.android>



User browser's "developer Tools"
to extract data

<https://play.google.com/store/apps/details?id=com.spotify.music>

My apps

Shop

Games

Family

Editors' Choice

Account

Redeem

Buy gift card

My wishlist

My Play activity

Parent Guide



Shazam

Shazam Entertainment Limited Music & Audio

★★★★☆ 3,142,995

Teen

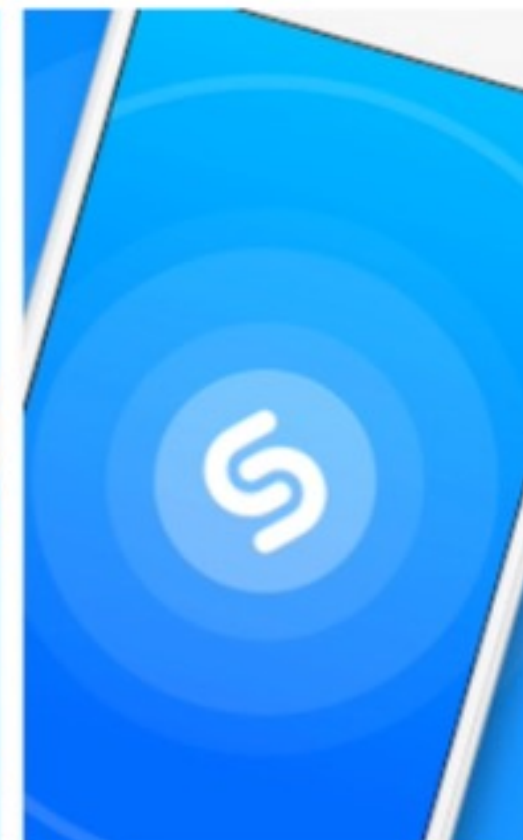
Contains ads

Add to Wishlist

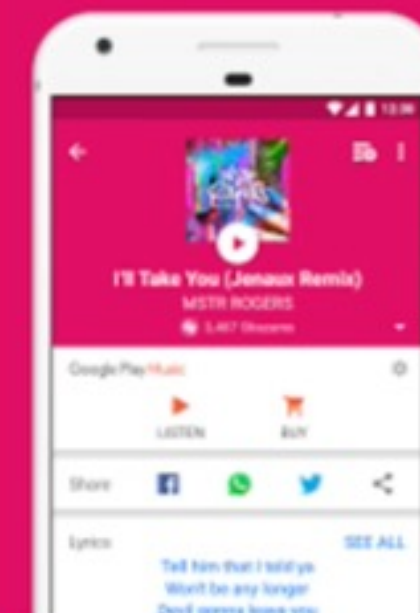
Install

Name
any
song in
seconds.

The world's most popular
way to discover music



Explore lyrics,
videos & playlists



Similar

See more



SoundHound - Music
Search
SoundHound Inc.

★★★★☆ FREE



Musixmatch Lyrics
Musixmatch

★★★★☆ FREE



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Thank You