DLI Accelerated Data Science Teaching Kit

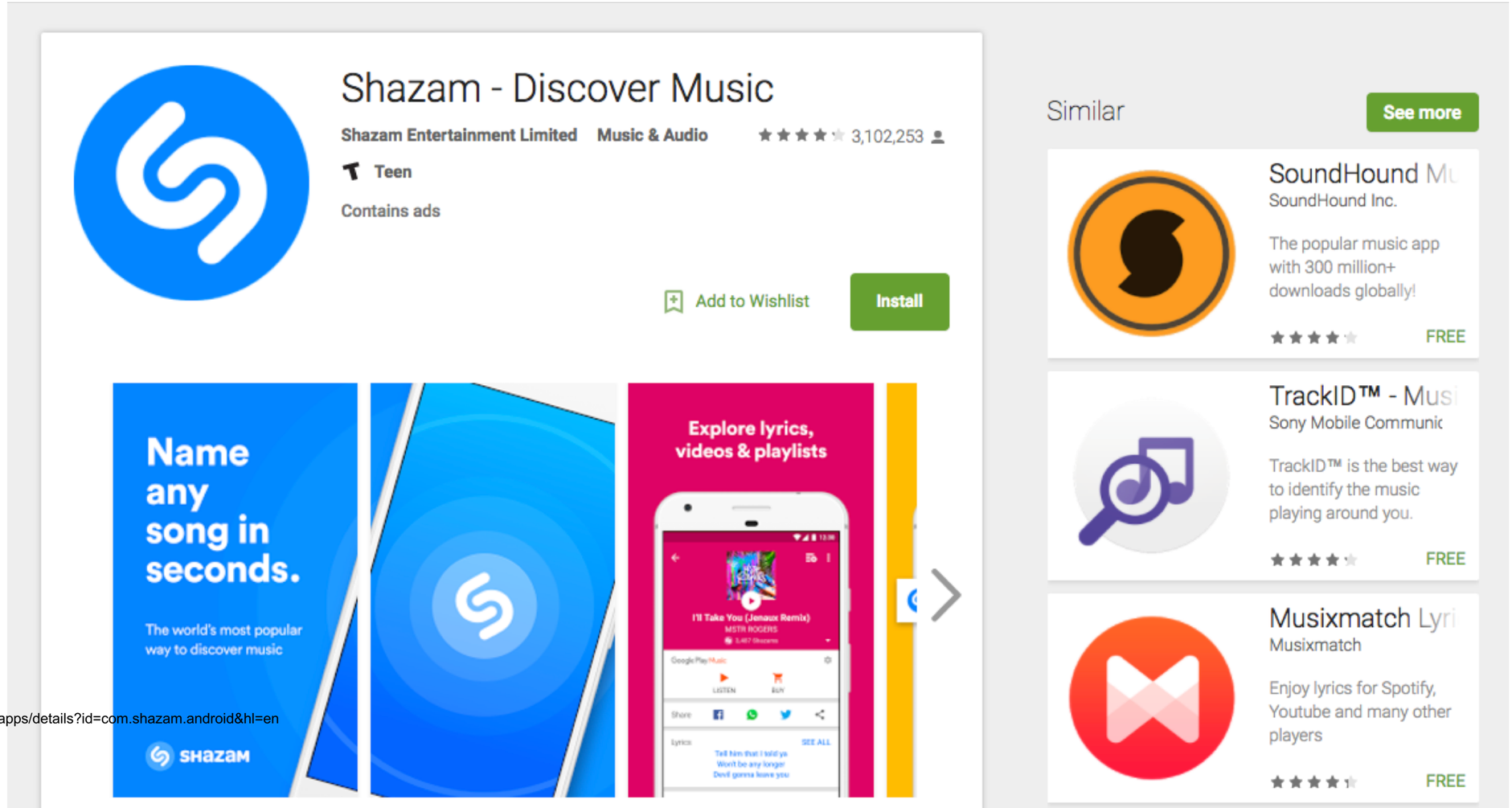
# Lecture 2.2 - Scraping Data

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.

# How to Scrape?

## Google Play example

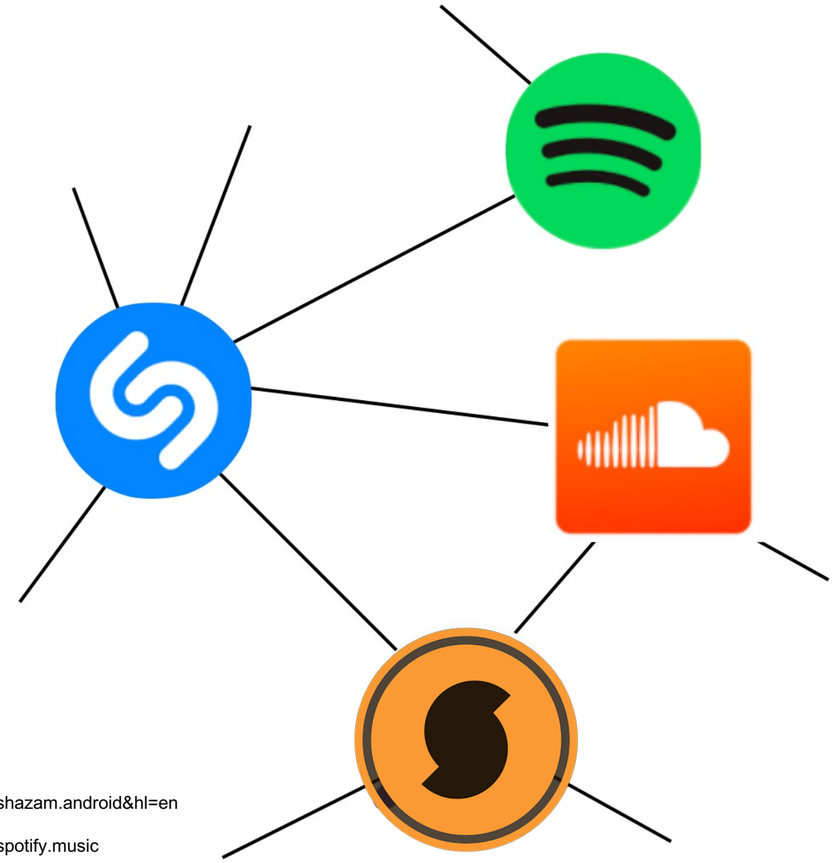
*Goal: collect the network of similar apps*




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



NVIDIA DEEP LEARNING INSTITUTE | Georgia Tech | PRAIRIE VIEW A&M UNIVERSITY

# How to Scrape?

## Google Play example

*Goal: collect the network of similar apps*

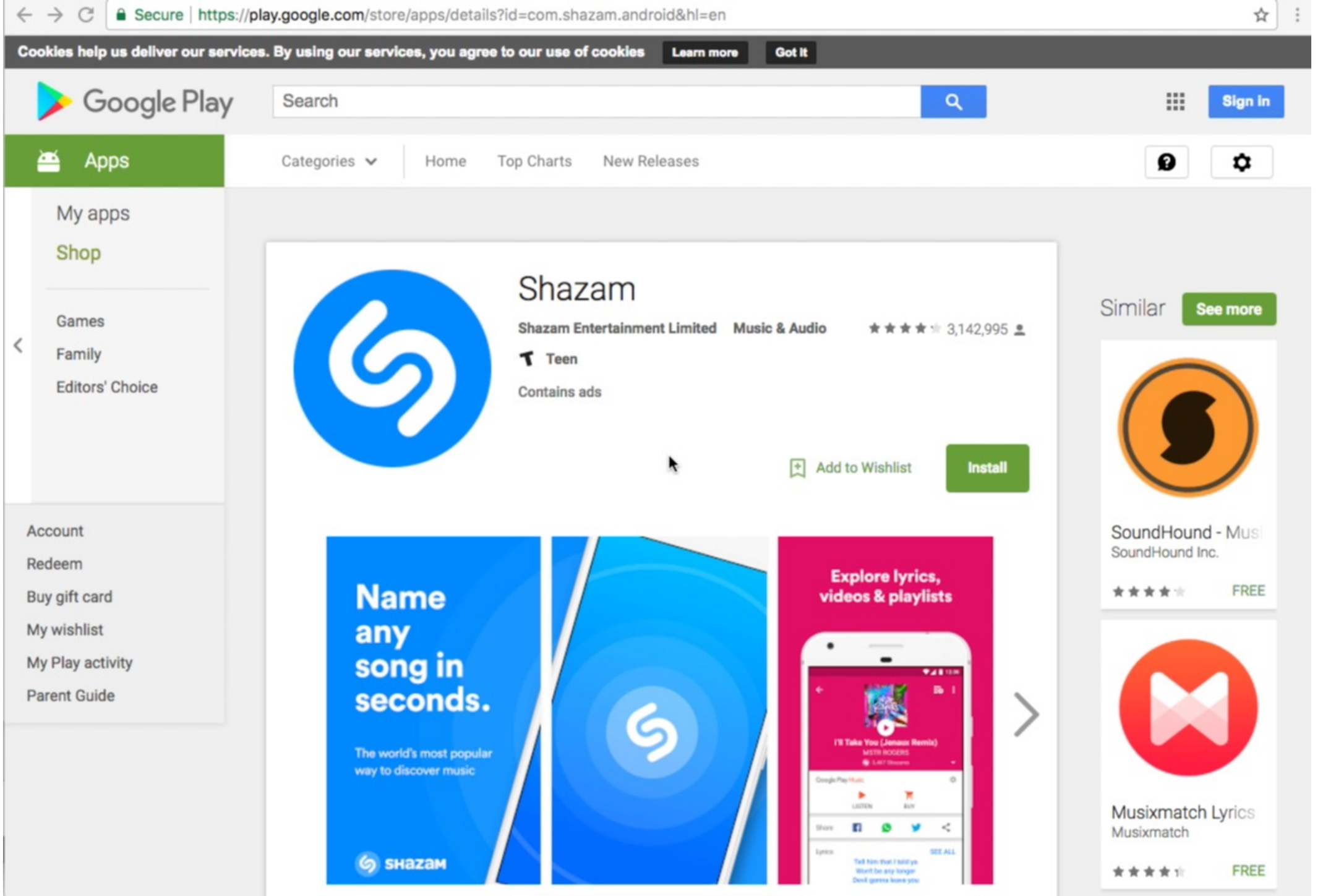
https://play.google.com/store/apps/details?id=**com.shazam.android**

User browser's "developer Tools" to extract data

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



NVIDIA DEEP LEARNING INSTITUTE | Georgia Tech | PRAIRIE VIEW A&M UNIVERSITY


Secure | https://play.google.com/store/apps/details?id=com.shazam.android&hl=en
Cookies help us deliver our services. By using our services, you agree to our use of cookies
Learn more
Got it
Google Play
Search
Sign in
Apps
Categories
Home
Top Charts
New Releases
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
SHAZAM
Explore lyrics, videos & playlists
Similar
See more
SoundHound - Mus
SoundHound Inc.
FREE
Musixmatch Lyrics
Musixmatch
FREE

DEEP
LEARNING
INSTITUTE

PRAIRIE VIEW
A&M UNIVERSITY

DLI Accelerated Data Science Teaching Kit

# Thank You