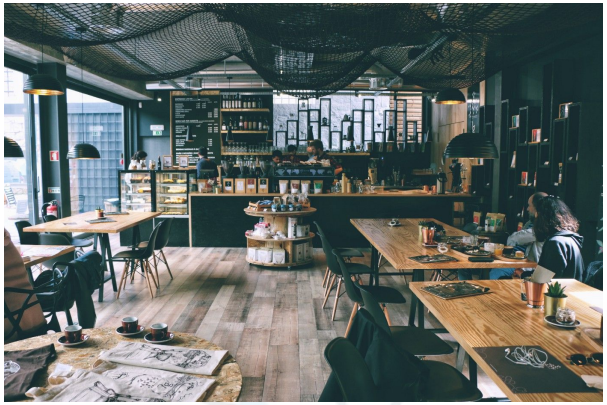
A Decade of Song Covers Knowledge Graph

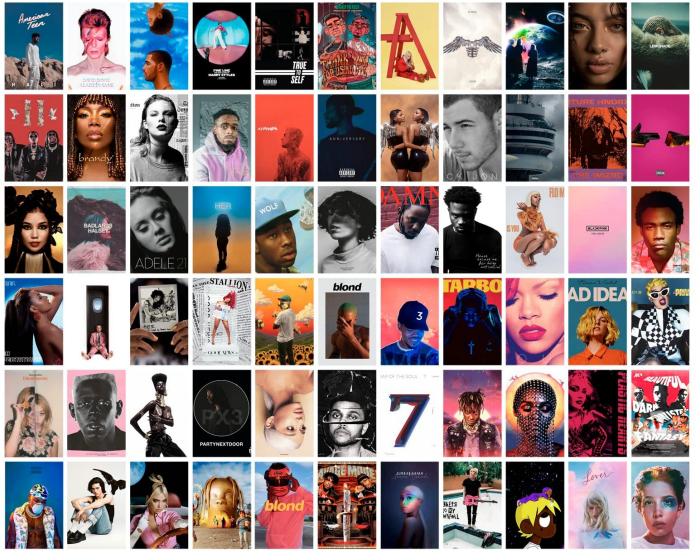
By Duyen Nguyen and Naicih Liou

Objective

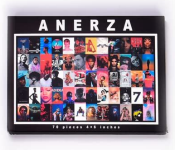


You heard a really good song cover but couldn't tell who the artist is.

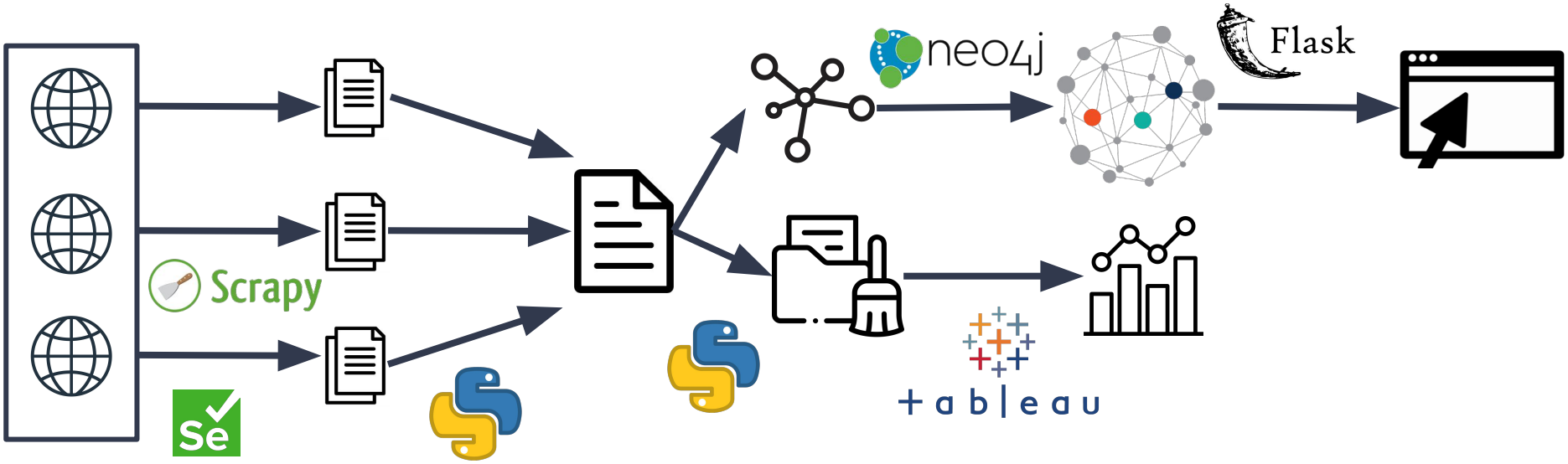
You are curious if your favorite artist has covered other artists' work.



Building a knowledge graph of song covers from 2012-2022.



Methodology



Scrape

Entity Resolution

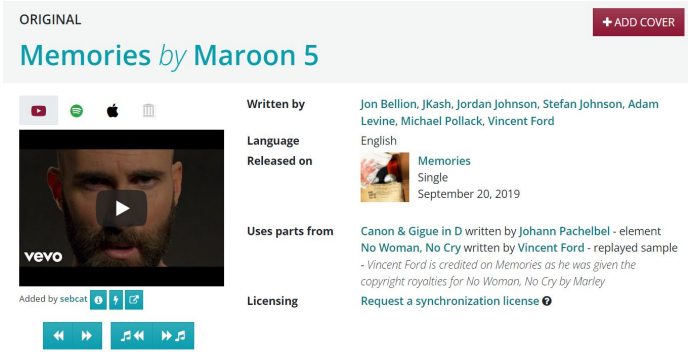
Data Preprocessing

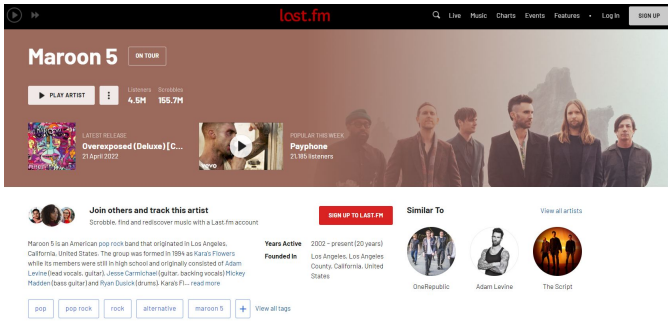
Data Analysis

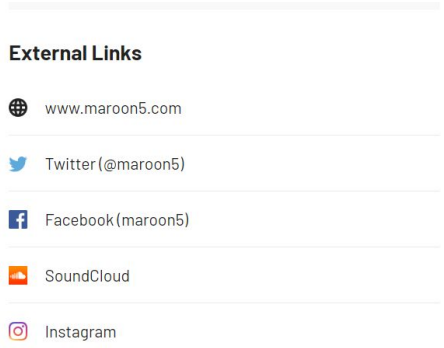
Build KG

Build WebAPP

Data Acquisition







Title	Performer	Release date	Info
Memories	Maroon 5	September 20, 2019	First release
Memories	Jonah Baker	October 4, 2019	Acoustic
Memories	Ben Woodward	October 11, 2019	Acoustic
Memories	Ali Brustofski	October 23, 2019	Acoustic
Memories	Landon Austin	October 25, 2019	Acoustic
Memories	Zita feat. Ben Woodward	November 5, 2019	
Memories	Chad Graham	November 12, 2019	
Memories	Joseph Vincent & Jason Chen	November 15, 2019	
Memories	Conkarah	November 26, 2019	
Memories	Ray Lorraine	December 5, 2019	Acoustic

WIKIDATA

https://secondhandsongs.com
Original songs and cover songs info

https://www.last.fm/
Singers info

https://www.wikidata.org/
Songs with awards won/nominated

Data Acquisition

More than **15 hours** spent for data crawling.

https://secondhandsongs.com
- 10 000+ original songs
- 100 000+ cover songs

(title, release date, original/cover song, writers, singers, number of covers,..)

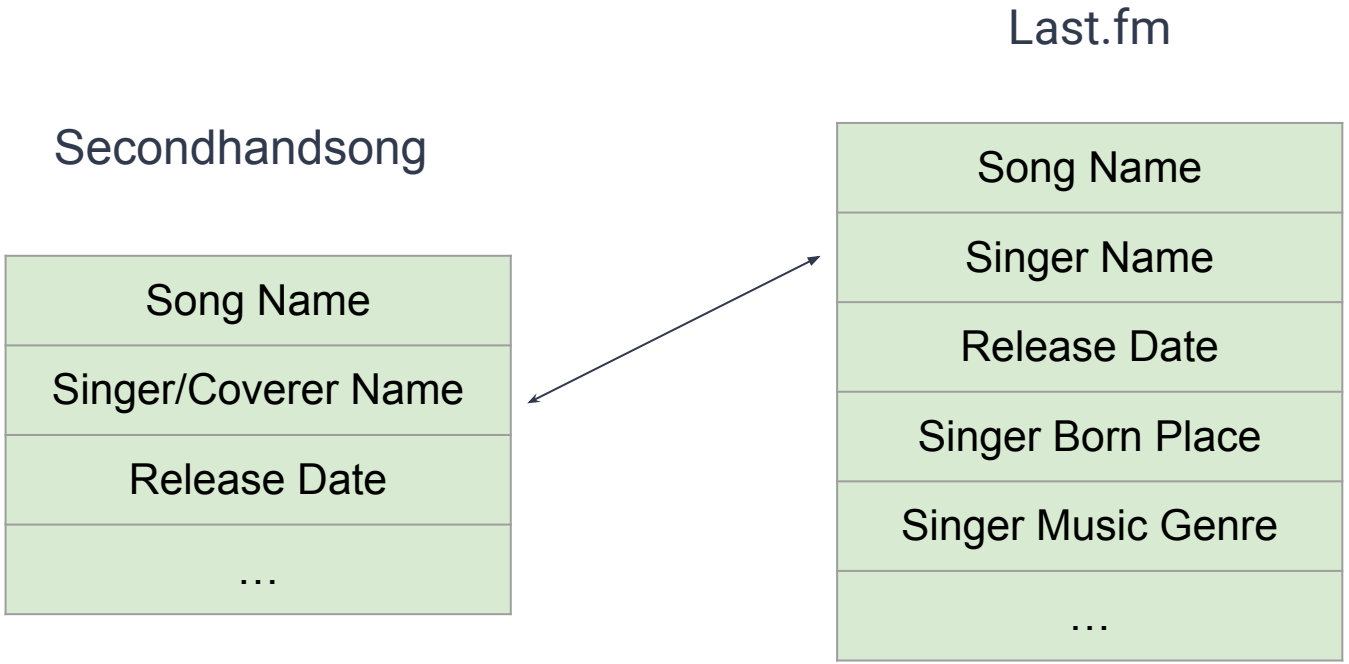
https://www.last.fm/
- Around 10 000 singers

(title, birthplace, birthdate, homepage url, twitter url, facebook url, instagram url)

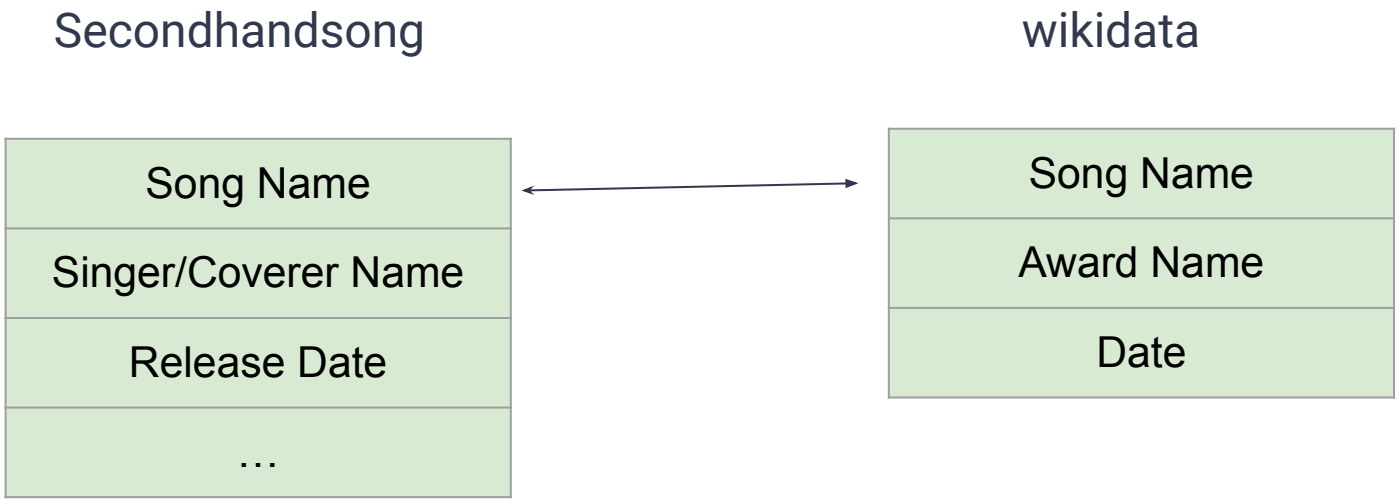
https://www.wikidata.org/
- Around 300 songs with awards

(title, award year)

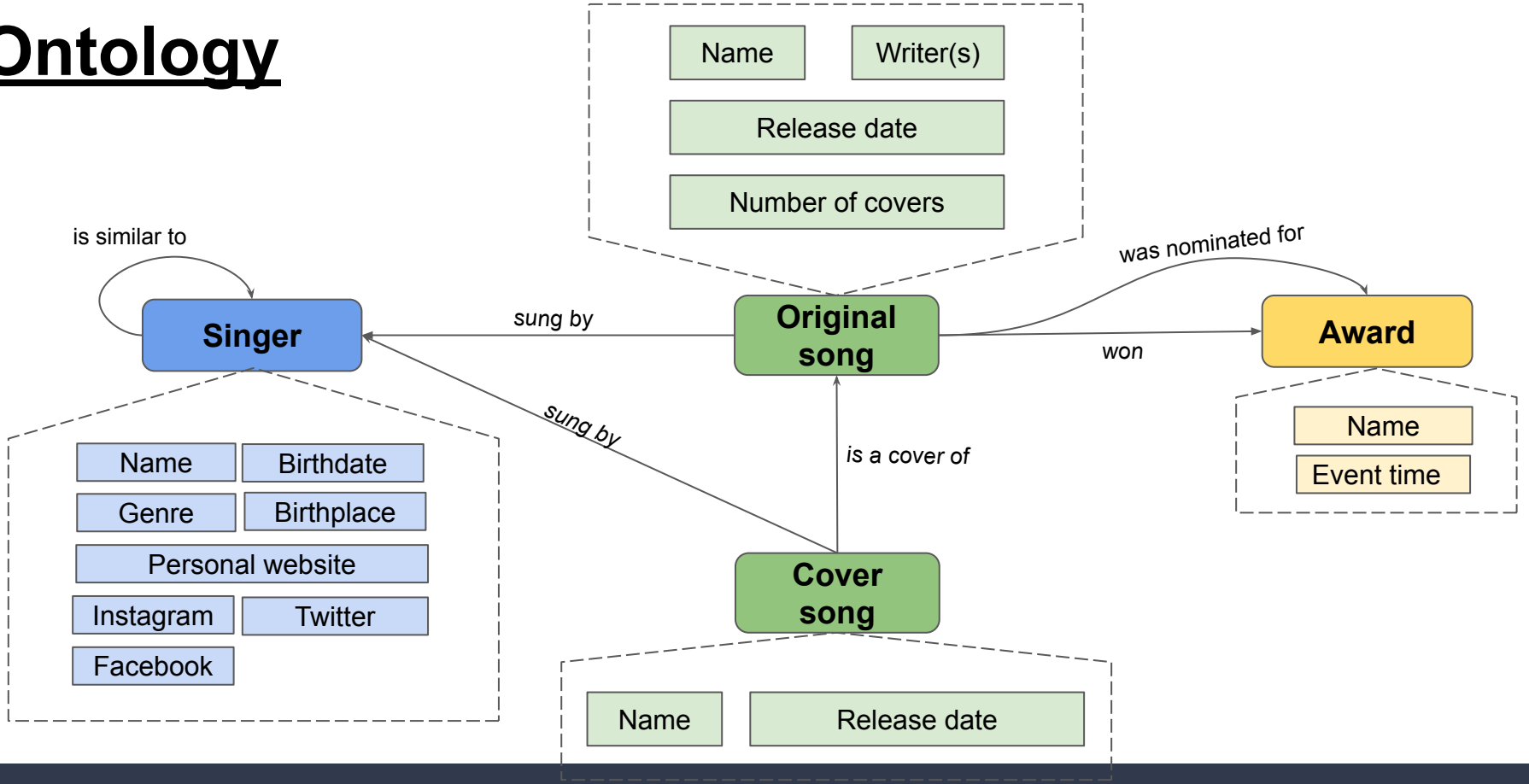
Entity Resolution



Entity Resolution



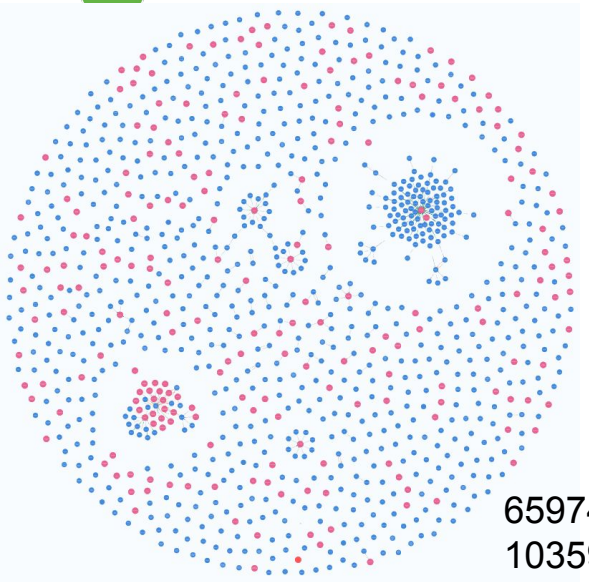
Ontology


Name
Writer(s)
Release date
Number of covers
is similar to
Singer
sung by
Original song
was nominated for
won
Award
Name
Birthdate
Genre
Birthplace
Personal website
Instagram
Twitter
Facebook
sung by
is a cover of
Name
Event time
Cover song
Name
Release date

KG Construction

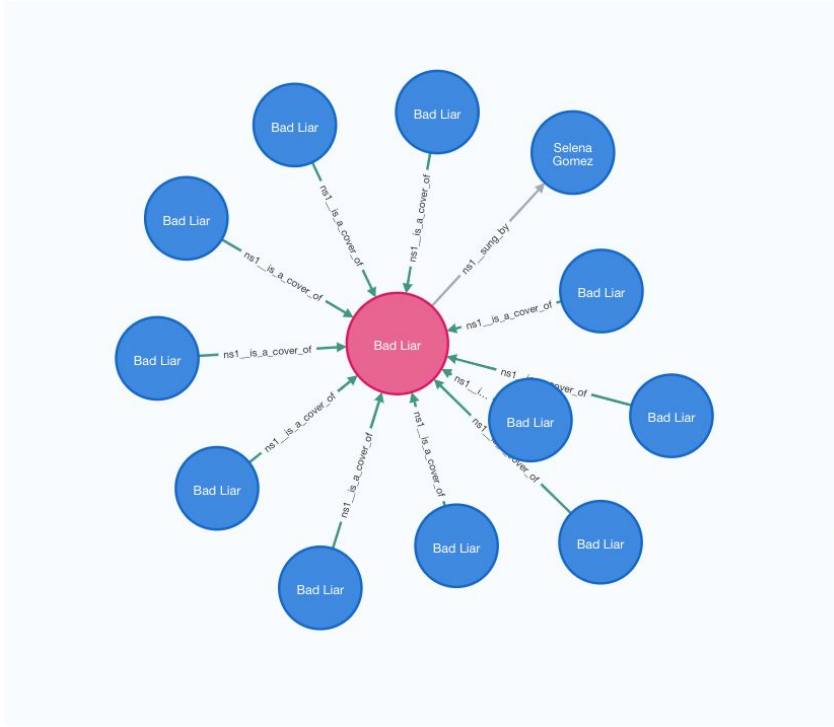
neosemantics

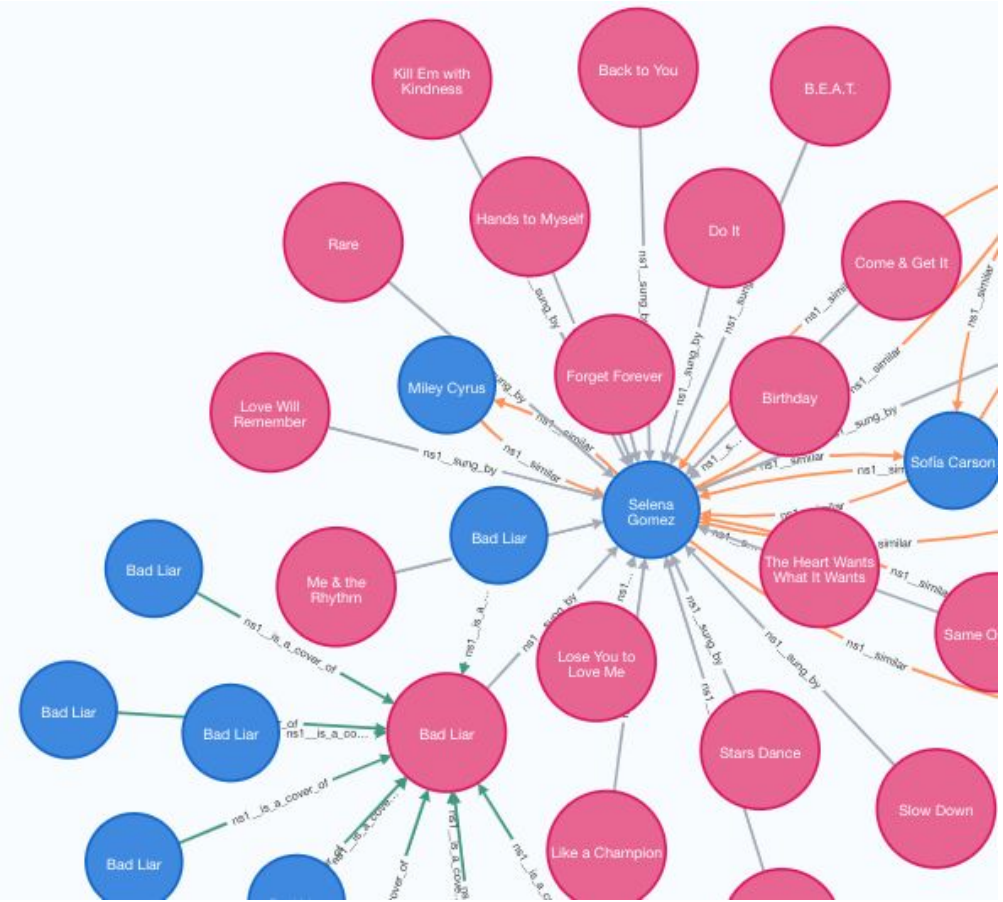

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ex: <http://dsci558.org/myprojectnamespace#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <https://schema.org/> .
@prefix wiki: <https://wikidata.org/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

ex:A_Happy_Thought8441_original a schema:original_song ;
    dc:sung_by ex:Kurt_Elling_featuring_Stu_Mindeman ;
    rdfs:covers 0 ;
    rdfs:has_writers "Stu Mindeman"@en ;
    rdfs:release_date "2018-03-23"^^xsd:date ;
    rdfs:title "A Happy Thought"@en .
ex:A_Million_Dreams4503 a schema:cover_song ;
    dc:is_a_cover_of ex:A_Million_Dreams128_original ;
    dc:sung_by ex:Lucy_Thomas ;
    rdfs:release_date "2019-08-16"^^xsd:date ;
    rdfs:title "A Million Dreams"@en .
```



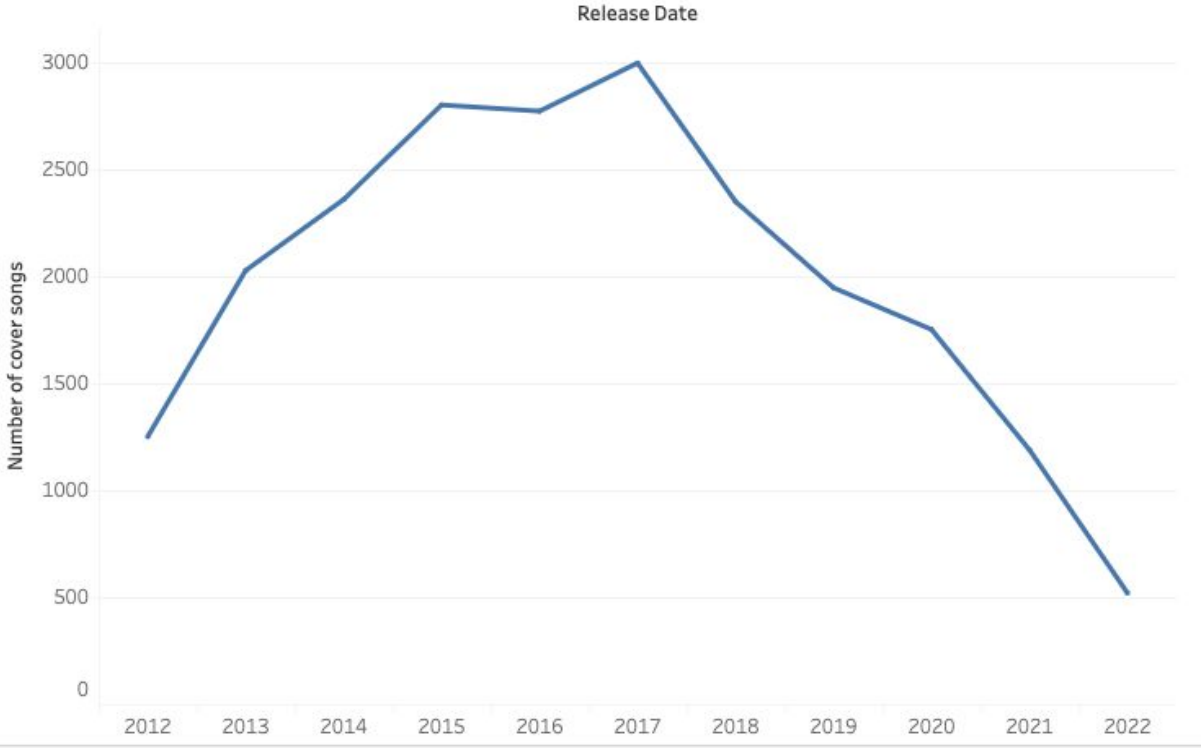
65974 nodes
103591 relationships

KG Construction



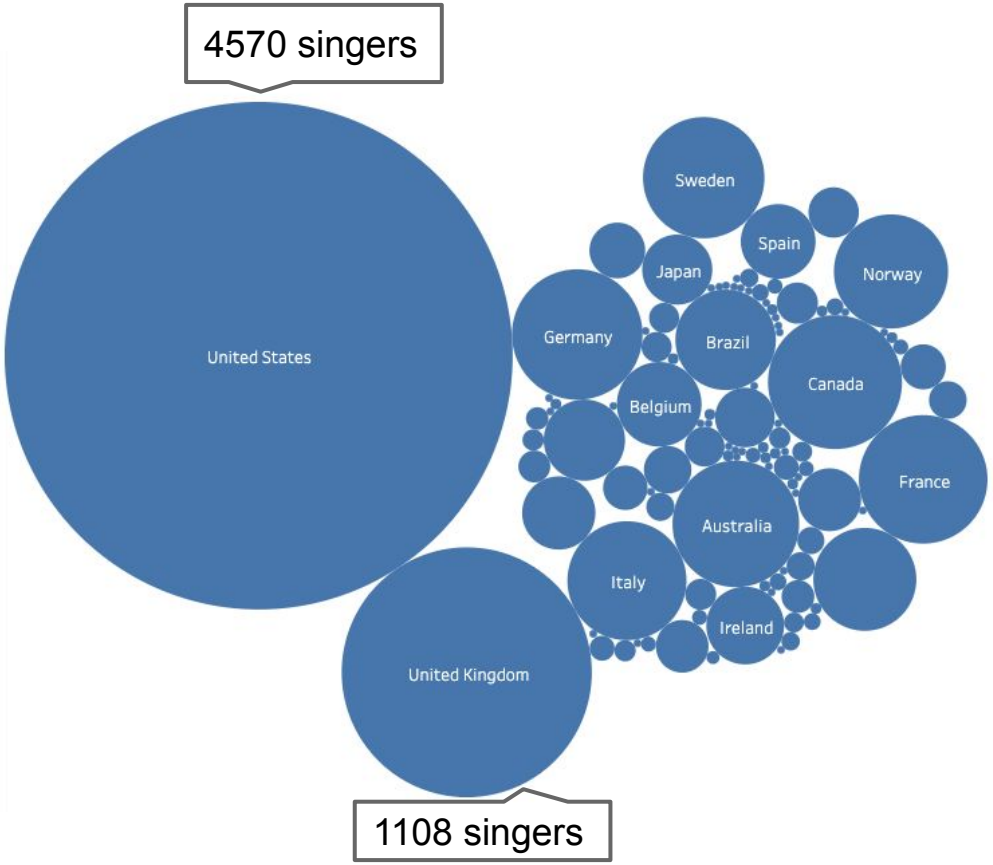


Data Analysis

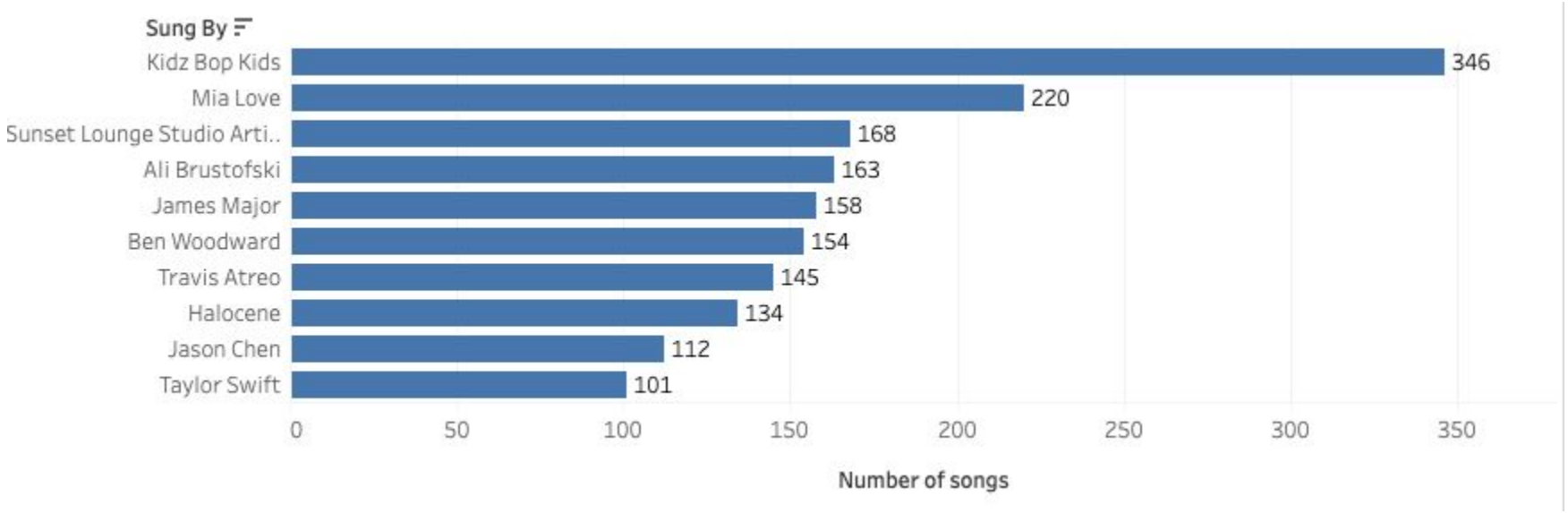


Data Analysis

Among 9000 singers, the number of US singers takes up more than half.



Data Analysis



Technical Challenge - Data Collection

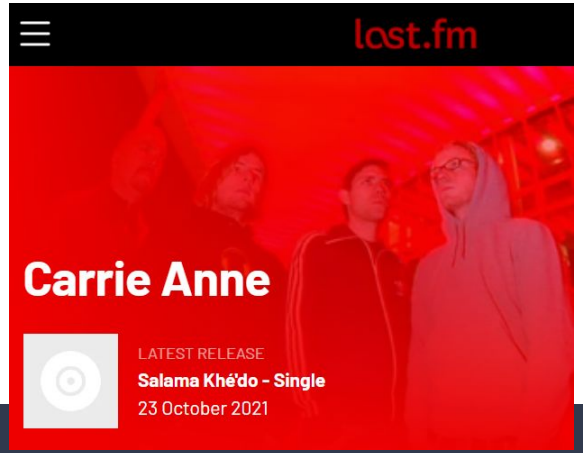
ORIGINAL

Carrie Anne *by* The Hollies

Added by Limbabwe

Written by	Allan Clarke, Tony Hicks, Graham Nash
Language	English
Released on	Carrie Anne Single May 26, 1967
Comments	#3 in the UK and #9 in the US
Licensing	Request a synchronization license

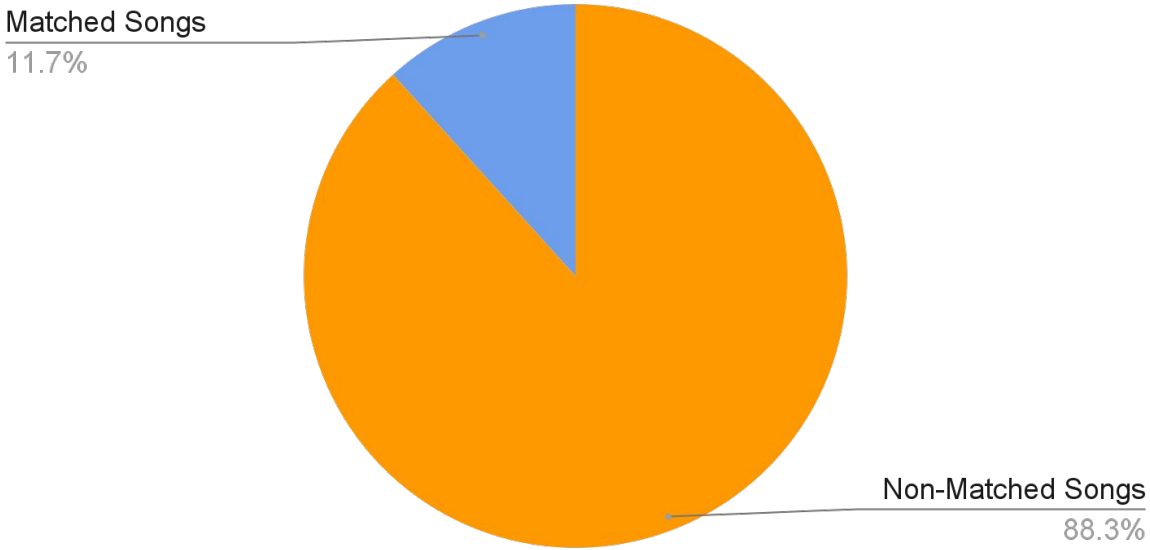
Data Consistency



Dynamic Web Scraping

Technical Challenge - Entity Resolution

- Among 145 nominated songs in wikidata, 17 songs from secondhandsong were matched

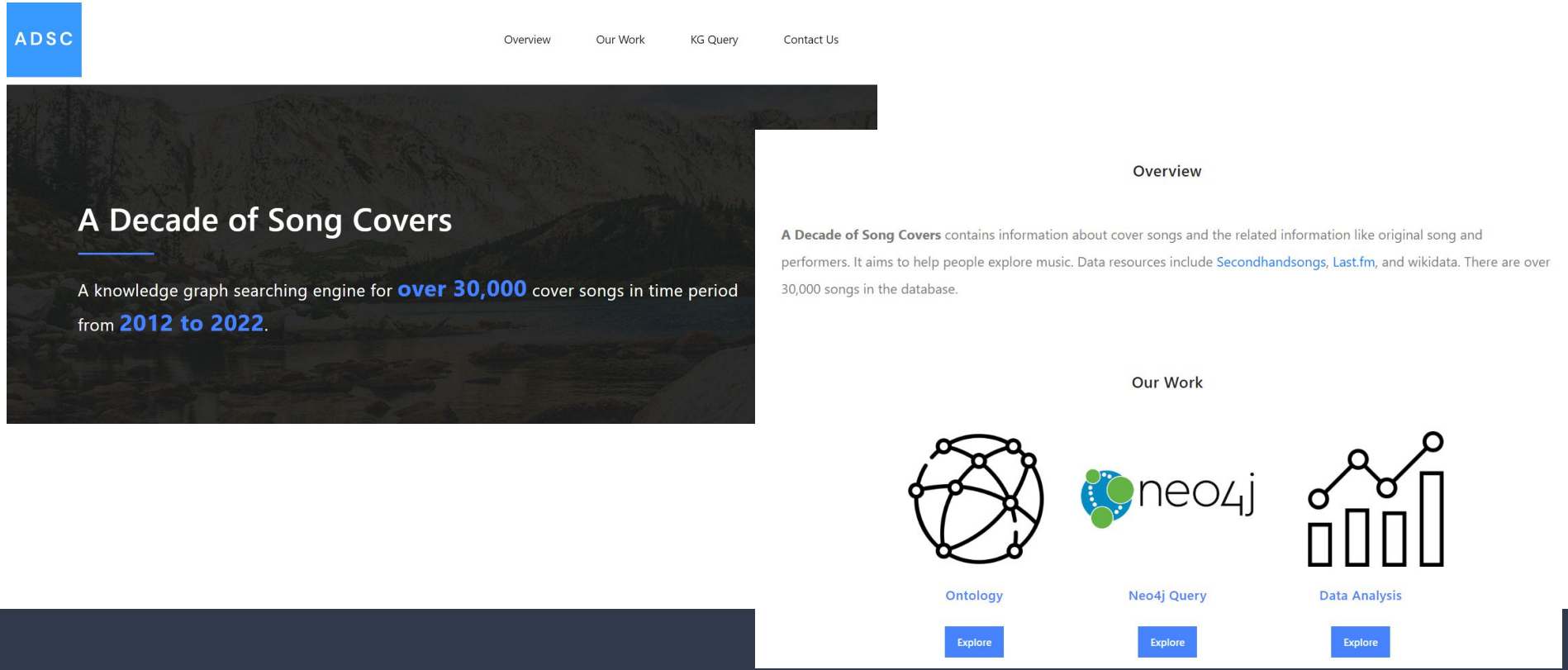


Technical Challenge - Webapp Building

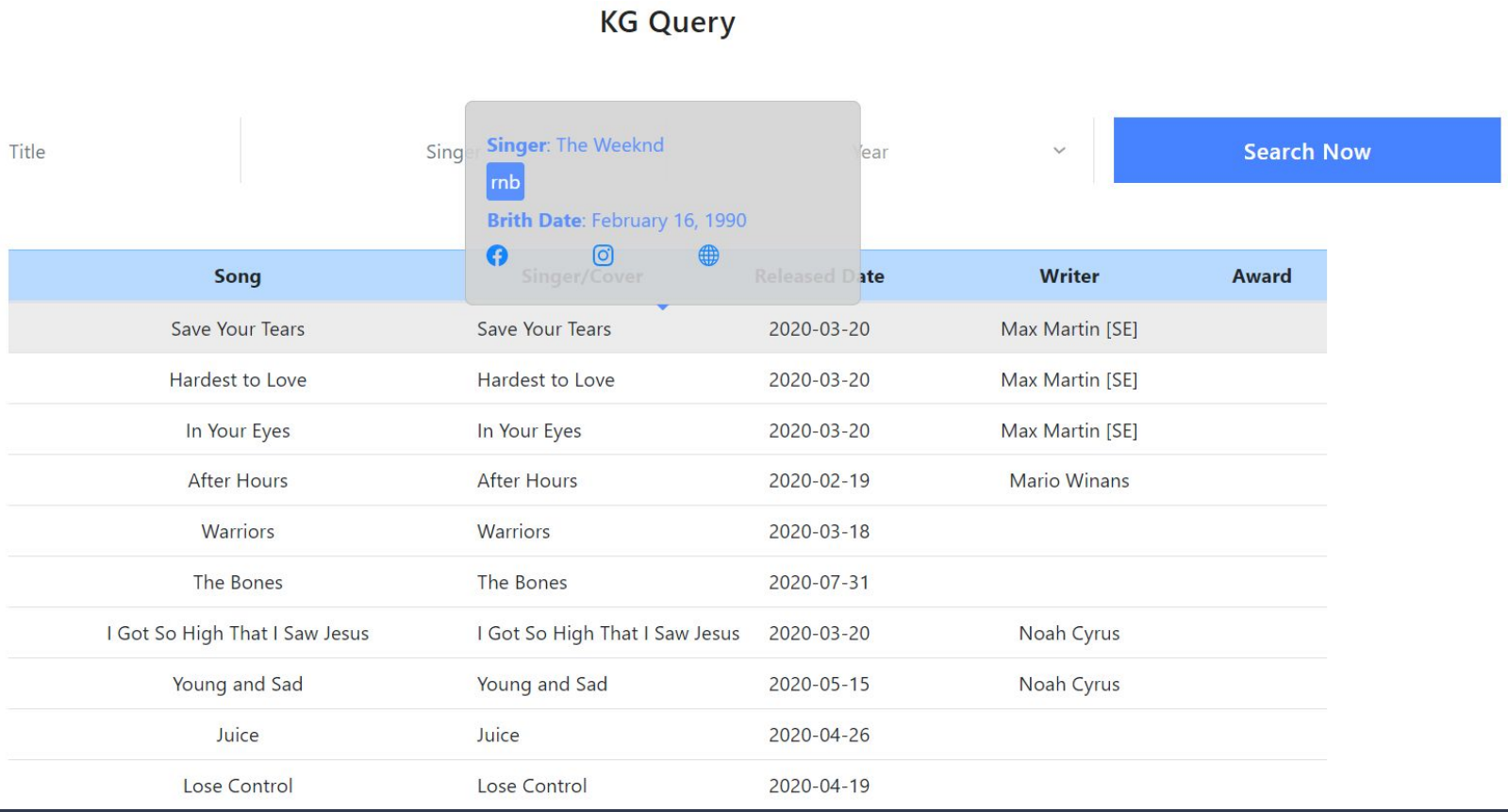
Validation

- We invited 15 users to test our website to confirm the utility.
- Users were asked to search songs by title, singer, and year.
- The feedback was positive and able to satisfy the need of song searching.

Website Demo



Website Demo



Song title	type	Release date	singer	writer	award
	orignal				
	cover				

THANK YOU FOR LISTENING!

Some transition fillers: moving on, next,

- Title ROLEPLAY!! - 1
- Motivation -1
- Methodology - 1
- Crawling/Data Aquisition - selenium, how many hours scraping run -2
- V Entity Resolution, EL yield F1, recall, precision (2)
- Ontology (KG Construction) (2)
- Data Analysis (3)
- Technical challenge - selenium, wiki award matching, ontology creation, flask+neo4j (4)
- Validation, right results shown after querying, ask people to test our website and get feedback -1
- Website demo - overall layout, query table - 2
- Project Links ROLEPLAY ENDING - 1

* Future work?
* Add tools used whenever suit
* Automated 20s per slide

Course Presentation Rubric

Were visual aids used throughout the presentation to support the major ideas?
Did the presenters speak clearly?
Did the presenters effectively budget their time?

Did the presentation clearly motivate the problem / KG?
- Clear motivation for the project with examples/images
- Motivation mentioned, but not very clear or explained
- Motivation unclear or barely covered

Did the presentation effectively convey the data used?
- Clear slide(s) showing data with examples/images
- Slide(s) with data listed out, but not supported by examples/images
- Unclear, Data sources scattered throughout presentation

Did the presentation effectively convey the ontological classes/relations, extraction, entity resolution/linking used?
- Clear slide(s) explaining ontology design, extraction approach, and entity linking or resolution
- Slide(s) describe 1-2 of ontology/extraction/ER and lack specificity
- Unclear, not described in enough detail to understand approach

Was the approach to the first technical challenge clearly explained?
- Presentation conveys the problem, examples, solution, and evaluation
- Presentation covers 2-3 of problem, examples, solution, and evaluation
- Presentation covers problem, examples, solution, or evaluation

Was the approach to the second technical challenge clearly explained? (Or re-use previous score if 1st challenge was significant)
- Presentation conveys the problem, examples, solution, and evaluation
- Presentation covers 2-3 of problem, examples, solution, and evaluation
- Presentation covers problem, examples, solution, or evaluation

Did the project assess the quality of the KG in any way?
- Clear explanation of evaluation of KG, rigorous results
- Vague explanation of evaluation, no results
- Did not evaluate

Did the presentation show analytics, visualization, or query of the KG?
- One or more visualizations of KG/query aligned with use case, easy to read/see
- Visualization that was ineffective or not aligned with motivation
- No visualization or query