

Search System for Billboard Songs

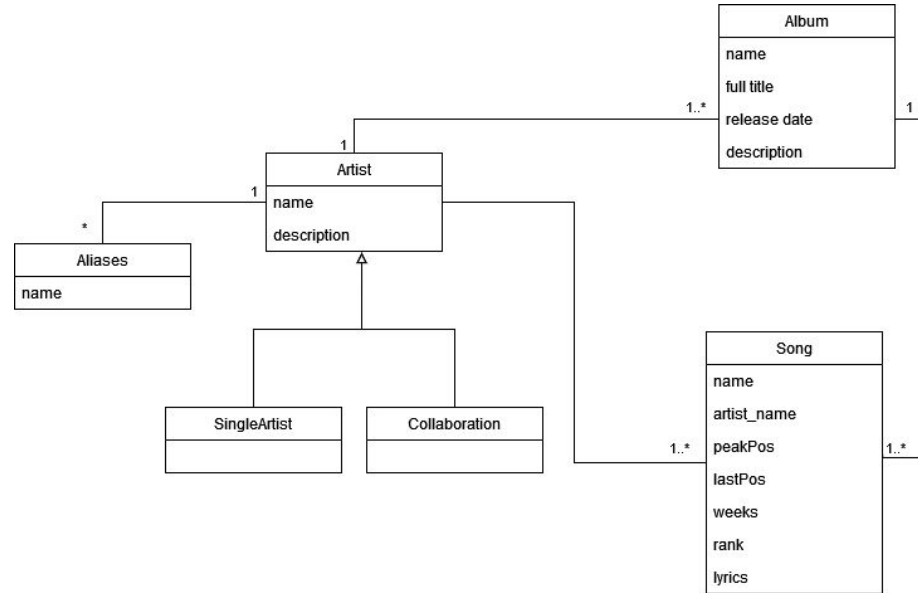
Ricardo Nunes, up201706860

Maria Baía, up201704951

Isla Cassamo, up201808549

Brief (re)introduction

- Our project is about the Billboard top 100. We aimed to develop a search engine that can retrieve any relevant information this



Previous Work

- 1st Milestone :
 - Data Gathering and Pre-processing;
 - Data analysis;
 - Search Topics.
- 2nd Milestone:
 - Collections and Documents;
 - Indexation;
 - System A and B.

System C

- In the previous section, two systems were developed for the information retrieval task: a simple system A, without term boosting, and a system B, with the term boosting.
- This section approaches the construction of a more complex system, a system C, that will incorporate several enhancements to improve the previous metric scores.
- For the system C, a set of tools that Solr provides were taken into account: Faceting , Query Re-Ranking , Synonym Graph Filter and MoreLikeThis.
- However, after analysing the set of tools mentioned above, and taking into account our search engine and the improvement needs for the previously presented queries, we excluded the Faceting tool and the Query Re-Ranking.

Faceting

- Dynamic clustering of search results into categories based on indexed terms
- Useful to implement in checkboxes in an UI
- Applied the Range Faceting from the Faceting tool to get range results in the following numeric fields:
 - **week**, to specify the amount of weeks a song stayed on the char
 - **music_rank**, to specify ranks in the Billboard chart

```
"facet_counts":{  
  "facet_queries":{},  
  "facet_fields":{},  
  "facet_ranges":{  
    "music_rank":{  
      "counts":[  
        "0",39,  
        "10",53,  
        "20",223,  
        "30",467,  
        "40",1933,  
        "50",1315,  
        "60",1210,  
        "70",1641,  
        "80",2706,  
        "90",8948],  
      "gap":10,  
      "start":0,  
      "end":100}},  
    "facet_intervals":{},  
    "facet_heatmaps":{}}}
```

```
"facet_counts":{  
  "facet_queries":{},  
  "facet_fields":{},  
  "facet_ranges":{  
    "weeks":{  
      "counts":[  
        "0",8039,  
        "10",7023,  
        "20",4345,  
        "30",476,  
        "40",151,  
        "50",60,  
        "60",11,  
        "70",2,  
        "80",1,  
        "90",0],  
      "gap":10,  
      "start":0,  
      "end":100}},  
    "facet_intervals":{},  
    "facet_heatmaps":{}}}
```

Synonymous Graph Filter

- Artists names are connected with several aliases, that can hinder the task of search.
- However, this filter was not used for the purpose mentioned above, since none of our queries do a search related with artists names.
- We decided to test this filter for the Information Need 1, where a file of synonymous words was given for the word "friendship"

System C		
Rank	Song	R
1	Say You'll Be There	Y
2	Spice Up Your Life	Y
3	Headlines (...)	Y
4	Buddy	Y
5	Mirror Man	Y
6	Freeze	N
7	Body Talk	N
8	Get Me Bodied	Y
9	Hold It Down	Y
10	Slow And Sexy	N
P@10	0.7	
AP	0.9325	

Table 7. P@10 and AP results for system C, 1st information need

MoreLikeThis

- Used for searching documents which are similar to one document that we provide.
- Used fields:
 - **mlt** for all the queries as true
 - **qf** that depends on the query in question
 - **mlt.mind** and **mlt.mintf** equal to 0 for all the queries, so that no document is excluded
 - the id where specifies the document id by which the search should guide

Results - IF N°1

- Search musics where the primary theme is friendship
- Document similar to album “Wannabe” by Coldplay, based on lyrics field

q={!mlt%20qf=lyrics%20mintf=0%20mindf=0}
71c9a539-d2b5-4ef6-b470-66a940017bb4

System C		
Rank	Song	R
1	Say You'll Be There	Y
2	Spice Up Your Life	Y
3	Headlines (...)	Y
4	2 Become 1	N
5	Killa	N
6	Freeze	N
7	Body Talk	N
8	Get Me Bodied	Y
9	Na Na Na	N
10	Slow And Sexy	N
P@10	0.4	
AP	0.875	

Table 8. P@10 and AP results for system C, 1st information need

Results - IF N°2

- Search for debut albums of artists in 2000
- Document similar to album “Parachutes” by Coldplay, based on album_description field

q={!mlt%20qf=album_description%20mintf=0%20mindf=0}
b3077236-e1bf-4166-aea6-bf9ad065a6b6

System C		
Rank	Album Name	R
1	Madonna by Madonna	N
2	Pretty. Odd. by Panic! at the Disco	N
3	Nevermind by Nirvana	N
4	Try This by P!nk	N
5	Born to Do It by Craig David	Y
6	Corinne Bailey Rae by Corinne Bailey Rae	N
7	Discovery by Electric Light Orchestra	N
8	A Fever You Can't Sweat Out by Panic! at the Disco	N
9	Kylie by Kylie Minogue	N
10	Screaming for Vengeance by Judas Priest	N
P@10	0.1	
AP	0.2	

Table 9. P@10 and AP results for system C, 2nd information need

Results - IF N°3

- Search for New York born artists with albums in the billboard 200 record chart
- Document similar to artist “Sharissa”, based on artist_description field

```
q={!mlt qf=artist_description mintf=0 mindf=0}  
57cc7f34-3e36-441c-9631-575927e48236
```

System C		
Rank	Song	R
1	Freeway Featuring Peedi Crakk	N
2	Freeway Featuring Jay-Z & Beanie Sigel	N
3	Fugees	N
4	Wyclef Jean Featuring Akon, Lil Wayne & Niia	N
5	Wyclef Jean Featuring Missy Elliott	N
6	Wyclef Jean Featuring Claudette Ortiz	N
7	Scarface	N
8	LaToya Jackson	N
9	Zhane	N
10	Alexandra Stan	N
P@10	0	
AP	0	

Table 10. P@10 and AP results for system C, 3rd information need

Results - IF N°4

- Search for albums that won the Grammy Award for Best R&B Album since the '90s
- Document similar to album "Get Lifted", based on album_description field

System C		
Rank	Song	R
1	Talking Book by Stevie Wonder	N
2	Give Me the Reason by Luther Vandross	N
3	Power of Love by Luther Vandross	N
4	Dance with My Father by Luther Vandross	Y
5	The Night I Fell in Love by Luther Vandross	N
6	Never Too Much by Luther Vandross	Y
7	Toni Braxton by Toni Braxton	N
8	Be by Common	N
9	Secrets by Toni Braxton	Y
10	Never Let Me Go by Luther Vandross	Y
P@10	0.4	
AP	0.3275	

Table 11. P@10 and AP results for system C, 4th information need

Conclusions

- This project required the search of appropriate datasets, determining the tools needed to explore them and make use of those tools to the desired extent.
- Even though System C did not out-performed the previous systems, it was a valid and important experiment to make, since one can't know beforehand the results of a modification to the system.
- However, the Synonymous Graph Filter show us an improvement on the results for the information need 1.