

500 Best Albums Ever

Information Retrieval

Eunice Amorim
Henrique Nunes
Henrique Sousa



Recap



The image features a light blue, cloud-like shape in the center, containing the word "Recap" in a large, bold, dark gray sans-serif font. Surrounding this central element is a decorative network diagram. It consists of several blue circular nodes, each with a white center and a blue outer ring. These nodes are connected by thin gray lines. In the top right, a path of three nodes is shown, with an ellipsis (...) above the final node. In the bottom left, a path of four nodes is shown, with an ellipsis (...) above the first node. The overall design is clean and modern, with a light gray background.

MUSIC

ADVERTISEMENT

The 500 Greatest Albums of All Time

The classics are still the classics, but the canon keeps getting bigger and better

BY ROLLING STONE

Illustration by Sean McCabe for Rolling Stone

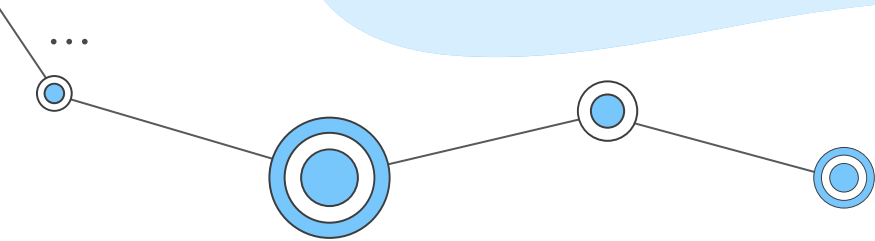
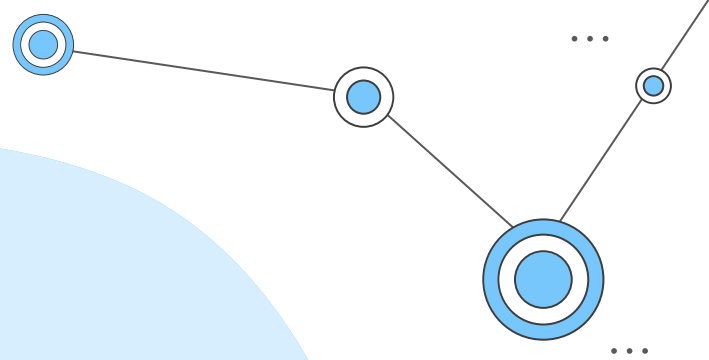
SEP 22, 2020 9:15 AM

TRENDING

Trump Threatened to



Tool



Tool

Solr 



Document



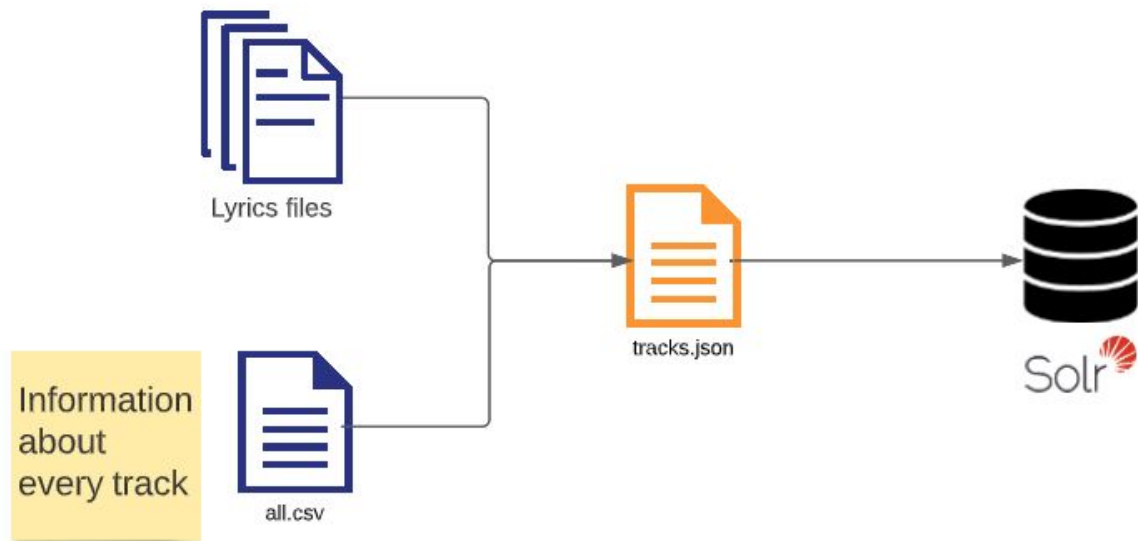
The image features a light blue, cloud-like shape in the center, containing the word "Document" in a bold, dark gray, sans-serif font. Surrounding this central element is a decorative network diagram. This diagram consists of several blue circular nodes, each with a white center and a blue outer ring. These nodes are connected by thin gray lines. In the top right corner, a path of three nodes is shown, with an ellipsis (...) above the final node. In the bottom left corner, a path of four nodes is shown, with an ellipsis (...) above the first node. The overall design is clean and modern, suggesting a digital or networked document.

Document Definition

Track
<pre>name: string album:string album_release_date: int album_ranking: int n_tracks: int track: string track_duration: int lyrics: string</pre>

...

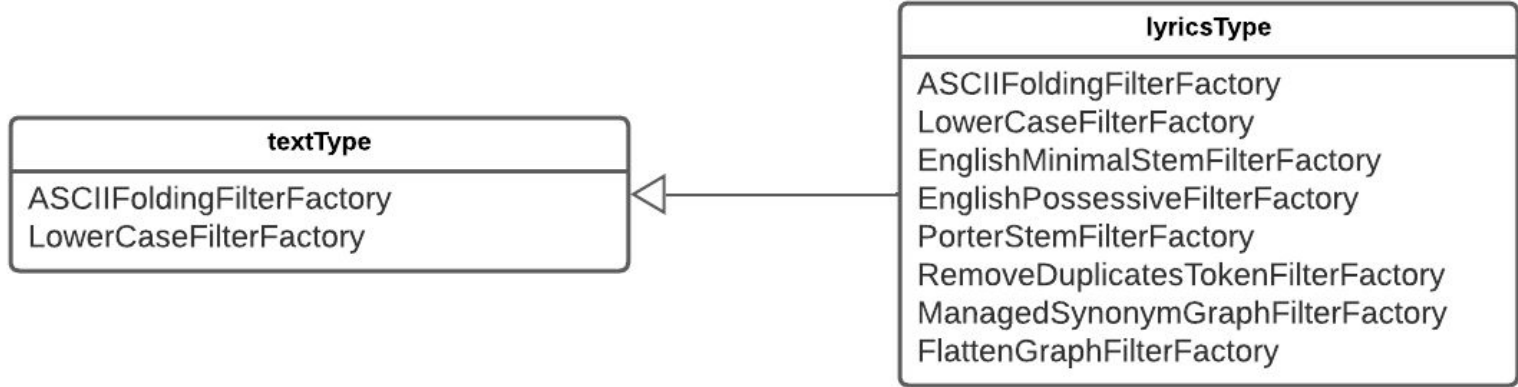
Uploading Data



Indexing

A decorative network diagram consisting of several blue circular nodes connected by thin grey lines. The nodes are arranged in a non-linear fashion, with some having concentric circles. Ellipses (...) are used to indicate that the network continues beyond the visible nodes.

Indexing



Indexing

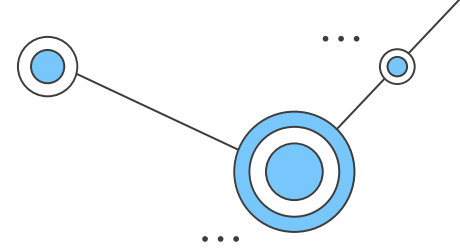
Track
name: textType
album: textType
album_release_date: pint
album_ranking: pint
n_tracks: pint
track: textType
track_duration: pint
lyrics: lyricsType

...

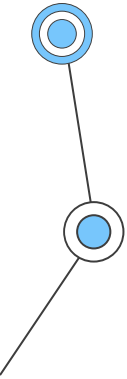


Information Needs

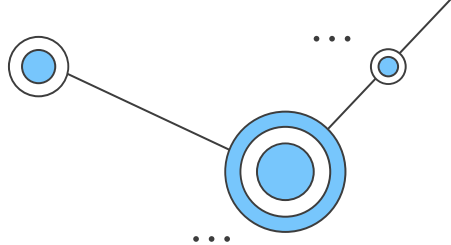
Information needs



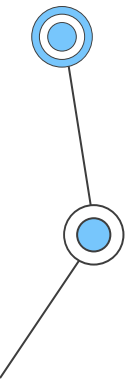
1. I want 1991 Nirvana songs having between 2 and 3 minutes
2. I want songs with a regretting tone
3. I want an underrated relaxing song
4. I want a song that speaks of love in a depressing way
5. I want a song that talks of life and love
6. I want songs about surprise and happiness
7. I want a song with a sentence like "I like her"
8. I want the very best songs



Information needs to queries



- Challenges
 - Converting a information need to a natural language query
 - Choosing simply and few words, and other attributes
- Methods and tools
 - Field boosts - 2
 - Term boosts - 4
 - Independent boosts - 3
 - Phrase match w/ slop - 7
 - Wildcards - 6
 - Proximity searches - 5





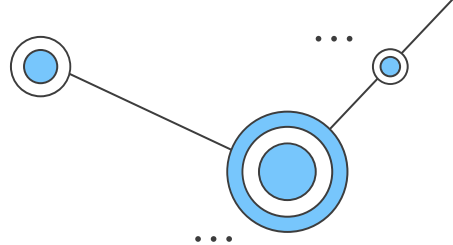
Demo



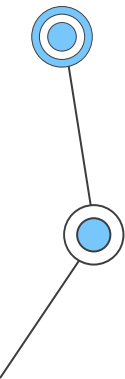
Evaluation



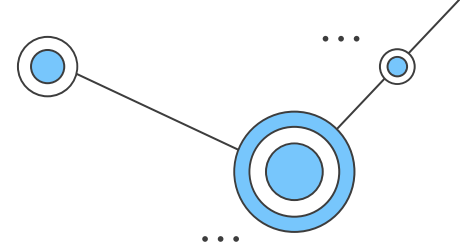
Evaluation



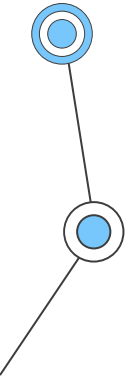
- Metrics
 - By query
 - Average Precision
 - P@10
 - P-R Curve
 - For the system
 - Mean Average Precision
- Manual evaluation
 - Determine relevant documents, up to the 30th retrieved document
- Comparison with other schemas



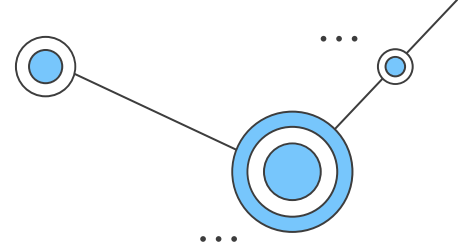
Evaluation



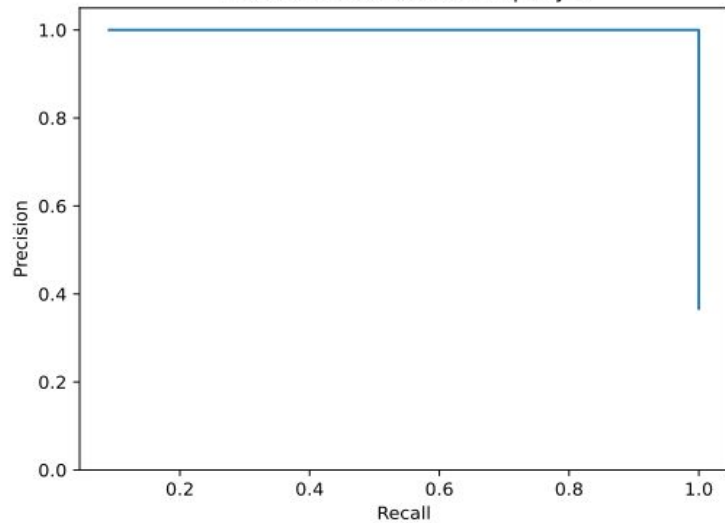
- Mean Average Precision
 - Initial Schema: 0.544416
 - Intermediate Schema: 0.638584
 - Final Schema: 0.656935



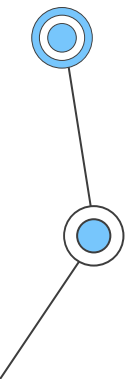
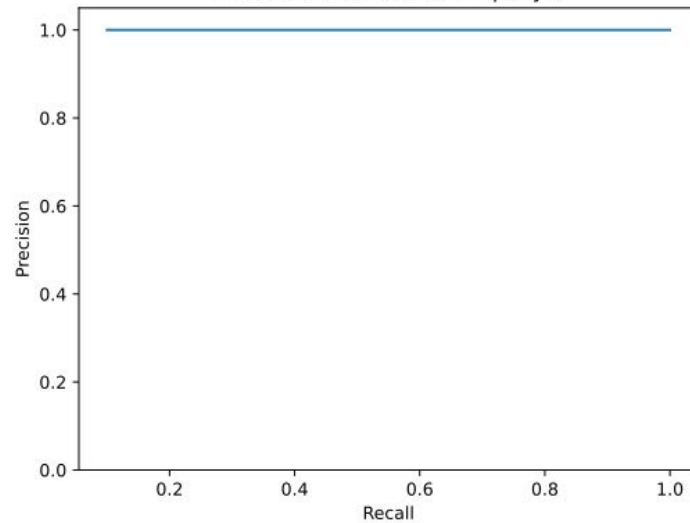
Findings



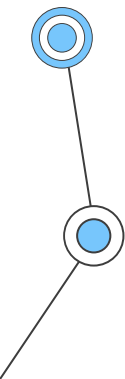
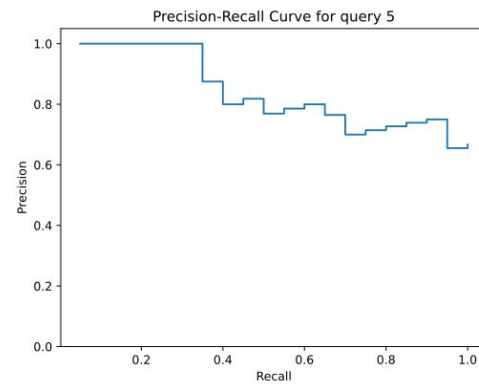
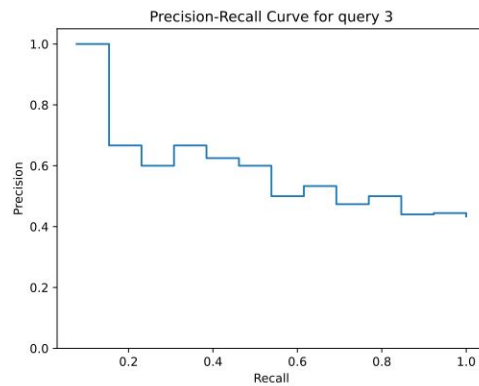
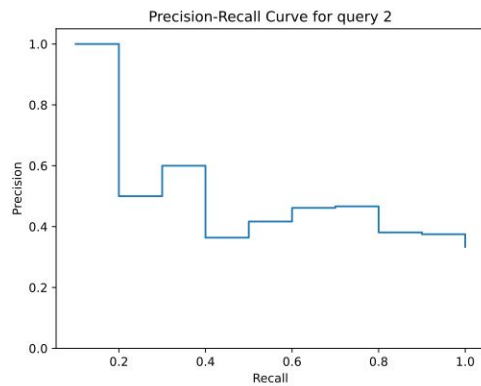
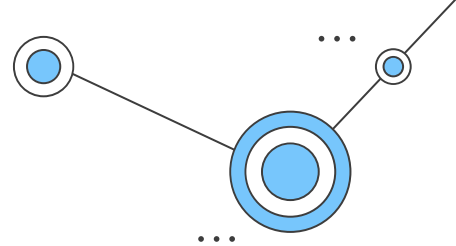
Precision-Recall Curve for query 1



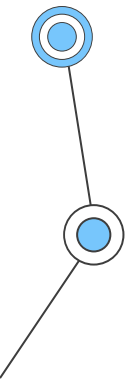
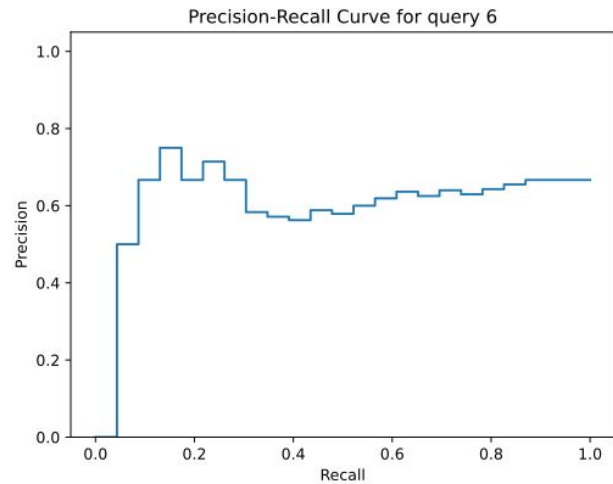
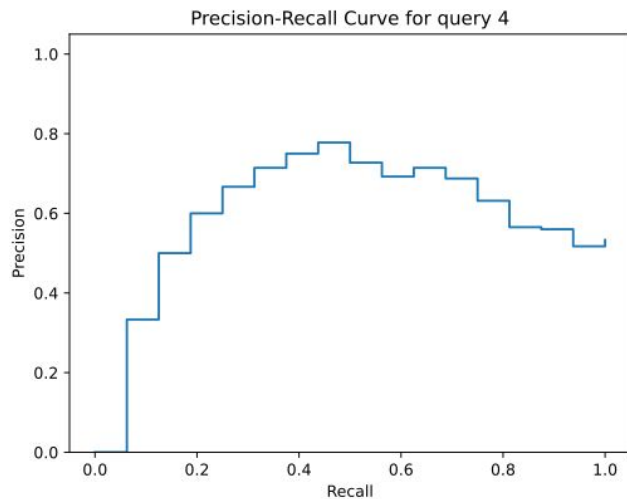
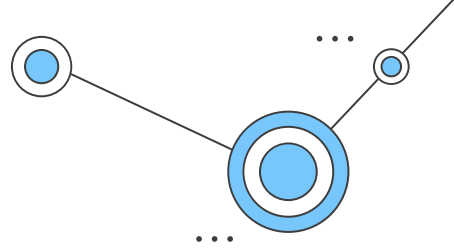
Precision-Recall Curve for query 8



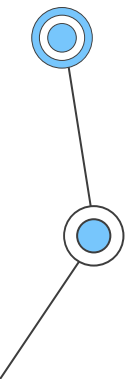
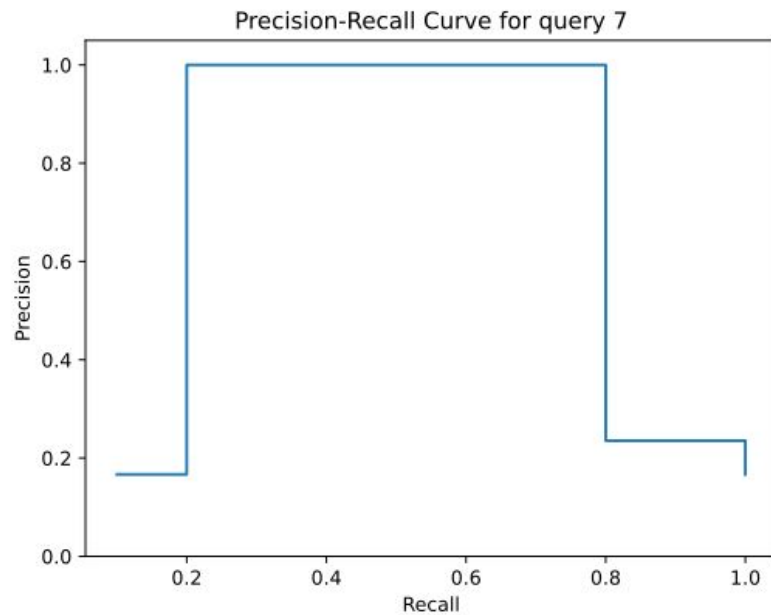
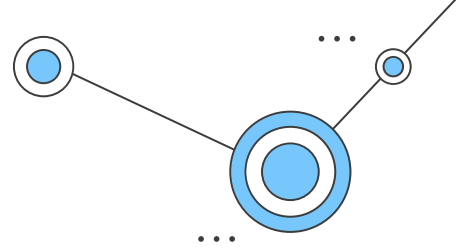
Findings



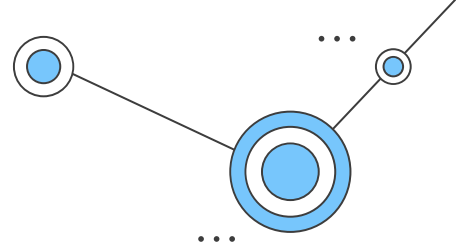
Findings



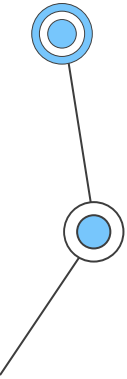
Findings



Discussion



- The metrics used are adequate
- Manual evaluation is toilsome and takes a long time
- Manual evaluation of just 30 documents fails to see a bigger picture of relevant documents that the query didn't gather
- Converting an information need to a query is not trivial without NLP features





Thanks!

Do you have any questions?

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), infographics & images by [Freepik](#) and illustrations by [Stories](#)

