Finding the Right Audio Content for You

Brammert Ottens
On behalf of great teams at Spotify
To unlock the potential of human creativity, by giving a million creative artists the opportunity to live off of their work, and billions of fans the opportunity to enjoy and be inspired by it.
456M Active Users
184 Markets
82M+ Music Tracks
3.6M+ Podcast Titles
4B+ Playlists
The Search Journey

**TYPE/TALK**
User communicates with us

**CONSIDER**
User evaluates what we show them

**DECIDE**
User ends the search session
High level architecture

Client → Search Service
{q: 'rihanna', user: 'user1'} → Ranked Candidate List

Search Service → Retrieval Service
Candidate List → {q: 'rihanna'}

Retrieval Service → Re-Ranking Service
Candidate List + {q: 'rihanna', user: 'user1'} → Ranked Candidate List
Metrics
Listeners come to Spotify with different *mindsets* and *intents*.

**Mindsets**
- **FOCUSED**: One specific thing in mind
- **OPEN**: A seed of an idea in mind
- **EXPLORATORY**: A path to explore

**Intents**
- **LISTEN**: Have a listening session
- **ORGANIZE**: Curate for future listening
- **SHARE**: Connect with friends
- **FACT CHECK**: Find specific info

DECIDE
User ends the search session

CONSIDER
User evaluates search results

TYPE/TALK
User communicates with us
Search Satisfaction ≈

SUCCESS ∆ EFFORT =

DECIDE metrics ∆ (TYPE metrics ⊕ CONSIDER metrics)

Decide
User ends the search session

Consider
User evaluates search results

Type
User communicates with us

Effort metrics associate with the type and consider phases

Success metrics associate with the decide phase

Success Metrics

Listen goal: stream
Organize goal: add to a playlist, save to collection, follow artist/playlist, download
Share goal: share link

Effort Metrics
# character deletions
session duration
# reformations before success in a search session
…
Ranking
Re-ranking

A query

List of candidates to be ranked

Feature Builder

Metadata

user, query and item-based features:
- Item popularity
- whether user has searched for this item before
- edit distance between prefix query and the matched item title
- similarity of the item to the user taste (vector)
- Users recent search behaviour...

fi,1 fi,2 ... fi,k

Learn to Rank model trained on search success logs.
Trained on search sessions that contain a success. All features are logged, to prevent offline/online skew.
We use LambdaLoss with NDCG weights, to optimize for NDCG

Ranked list of candidates
Ranking Architecture
One App
One Search Experience
One App
One 456M Search Experiences
A diverse catalogue
One list to rule them all
We need to rank many different types of content

- We have vectors for some types of content, but not all
- Success for music queries looks different from success for podcast queries
- Not all features are available for all different content types
Short Queries

- Most queries we see are short, and thus not very informative and ambiguous.
- Personalization works very well for *known intent* searches
- Open searches are a challenge for instant search
Differences in behavioural patterns

<table>
<thead>
<tr>
<th>Relative difference in success metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream</td>
</tr>
<tr>
<td>Add to collection</td>
</tr>
<tr>
<td>Add to playlist</td>
</tr>
<tr>
<td>Follow artist</td>
</tr>
<tr>
<td>Download</td>
</tr>
<tr>
<td>Share</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relative difference in effort metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average relative character deletions</td>
</tr>
<tr>
<td>Average query length in characters</td>
</tr>
</tbody>
</table>

* Tian, Mi; CHIIR 2022 industry day presentation*
Differences in behavioural patterns

### Relative difference in success metrics

<table>
<thead>
<tr>
<th>Action</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream</td>
<td>+3%</td>
</tr>
<tr>
<td>Add to collection</td>
<td>+30%</td>
</tr>
<tr>
<td>Add to playlist</td>
<td>-59%</td>
</tr>
<tr>
<td>Follow artist</td>
<td>-92%</td>
</tr>
<tr>
<td>Download</td>
<td>+593%</td>
</tr>
<tr>
<td>Share</td>
<td>+44%</td>
</tr>
</tbody>
</table>

### Relative difference in effort metrics

- Average relative character deletions: +53%
- Average query length in characters: +13%

Users may have different goals searching for different content types. Searching for podcast on Spotify requires much higher effort than music.

* Tian, Mi; CHIIR 2022 industry day presentation
Term-based search does not cut it
- Are electric cars better for the environment?
- Environmental impact of electric vehicles?
- Electric cars and ecology
Data for fine tuning

- Query/Episode pairs from **Search logs**
  - successful podcast searches
  - query reformulations
- **Synthetic** queries from popular episode titles and descriptions
- **Manually** written queries for popular episodes
Elasticsearch

Natural Language Search (Vespa)

Other sources...

Final-stage ranking

- electric cars climate impact
- Are electric cars better for the environment?
- Environment impact of electric vehicles
- Electric cars and ecology
Semantic Search does cut it
How can you help?
Datasets

Dive into datasets for everything from podcasts to music recommendation

https://research.atspotify.com/datasets/

The Million Playlist Dataset: Learning from Music Playlists
Oct 05, 2020
Dataset for music recommendation and automatic music playlist continuation. Contains 1,000,000 playlists, including playlist- and track-level metadata.

Spotify Podcasts Dataset: 100,000 episodes with text and audio
Apr 15, 2020
Dataset for podcast research. Contains 100,000 episodes from thousands of different shows on Spotify, including audio files and speech transcriptions.

WSDM Cup: The Music Streaming Sessions Dataset
Nov 15, 2018
Dataset for researching how to model user listening and interaction behavior in music streaming. Also includes data for music information retrieval and session-based sequential recommendations.

OpenMic: Audio and Crowd-Sourced Instrument Labels
Sep 23, 2018
Dataset for researching multi-instrument recognition in polyphonic recordings, a fundamental problem in music information retrieval.
Research Areas

Turning in-depth audio knowledge into rewarding audio experiences

- Algorithmic Responsibility
- Audio Intelligence
- Human-Computer Interaction
- Language Technologies
- Machine Learning
- Search & Recommendations
- User Modeling
- Music Creation