On the Challenges of Podcast Search at Spotify

Mi Tian
Spotify
mtian@spotify.com

Claudia Hauff
Spotify
claudiah@spotify.com

Praveen Chandar
Spotify
praveenr@spotify.com

1 TALK ABSTRACT

Online music streaming is enjoying ever-growing popularity in the last decades, enabled by the abundance of music content in digital format and online streaming services. In the recent years, *podcasts*, as a talk-focused media format, have witnessed a rapid growth among listeners. Podcasts come in many forms and sizes. They range from 20-minute daily meditation sessions, weekly recaps of global news, interviews with celebrities, and hosts bantering with each other for hours.

More and more streaming services are now expanding their catalogs to support both music and podcasts *on the same platform*, such as Amazon Music, Pandora and Spotify. This setup requires an effective *aggregated search system* to assemble information from heterogeneous information sources and content types (e.g., artist profiles, playlists, podcast shows, etc.) into one result interface in order to support diverse information needs [6].

In this talk, we will discuss the challenges with *podcast search*, as they present themselves for Spotify. We describe the general challenges with specific examples from our search system, and outline our ongoing works to address them.

Instant Search Challenge. Instant search is a search paradigm that renders the *search results pages* (SERPs) for each keystroke event in an attempt to reduce user search effort [7]. With this reduction in user search effort however come drawbacks, as no explicit signal of a complete query (as in the case of web search for instance) is received. An example of the difference between complete and incomplete query results are shown in Figure 1.

Content Mixing Challenge. Consider Figure 1 one more time. Both SERPs mix different types of content: songs, artists, playlists, and podcasts shows and episodes. Qualitative studies have found that users adopt various search strategies (such as query-based search, directed and undirected browsing) to find podcasts, and that users' goals and search strategies are strongly influenced by their perceptions of available tools, with a lack of tools notably perceived for online audio search [1]. A recent study based on large-scale query logs from Spotify has identified fundamental differences in
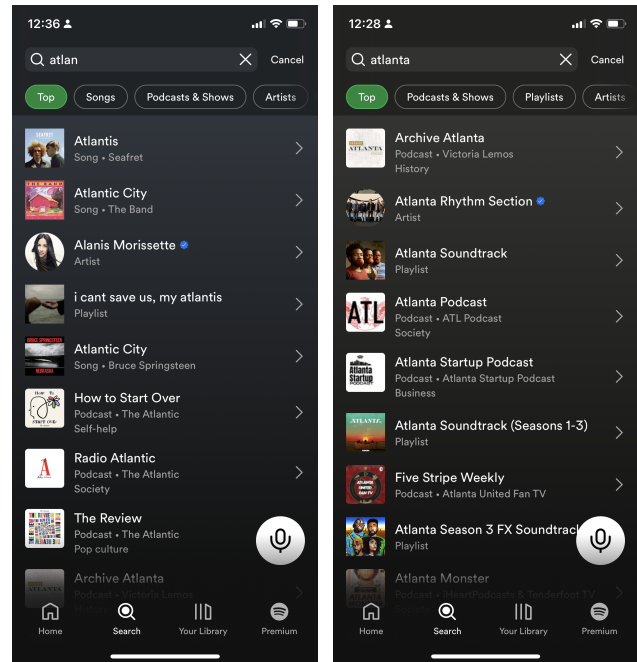
Figure 1: Consider the information need *What to do in Atlanta, Georgia, USA?* which a user may "translate" into the query `"atlanta"`. Shown here is Spotify's mobile SERP for *incomplete* query `"atlan"` and *complete* query `"atlanta"`, retrieved for a US user on July 18, 2022.

user goals and behaviors for music and podcast search [4]. Advanced machine learning approaches (e.g. multi-task learning [4]) are required to take such differences into account when computing a single ranking of mixed content for the user.

Query Understanding Challenge. Users often have a good idea about the desired music content and are able to express their information needs succinctly, with queries commonly formulated to describe the *entity name* (eg. artist or album name), *genre* or *mood* ("funk"), *language* ("hindi songs") and *listening context* (eg. "shower playlist"). However, users may struggle to translate their information needs for talk content into "good" queries (i.e., queries that bring about relevant content). Here, query understanding and query support (e.g., via query auto-completion or related searches) are especially important to help users find relevant content.

Metrics Challenge. Established metrics from the IR community have been deployed to quantify short-term *success* and *effort* for instant search in online streaming platforms [3]. To what extent these metrics are suitable to measure long-term user satisfaction in podcast search (in contrast to music search), however, remains to be further investigated. From a business perspective, it is also

a challenging task to establish the causal relationship between podcast search metrics and composite business metrics such as *lifetime value* (LTV), when podcast is one of the various audio formats users can stream on the platform.

Offline Evaluation Challenge. While *randomized controlled experiments* with real users, or *A/B tests*, are the norm in large-scale production systems, offline evaluation is an important first step to save A/B test bandwidth and reduce potentially harmful user impact [5]. Creating a podcast corpus with suitable information needs and high-quality judgments is however not an easy task for various reasons. To name but a few, podcast queries are still under-represented compared to music queries in our search logs, rankings from the logs are personalized and queries are in a prefix form, relevance judgments are noisy, identifying representative information needs for a large part of the user base is hard. Improving podcast offline evaluation remains a challenging yet vital task for effective product innovation and design making.

UI Challenge. What and how much information is required for a user to make an informed decision about which podcast to listen to? How can this information be extracted at scale and personalized for each user? Compared to a web search engine, the composition of the UI of audio-focused information access systems remains an under-studied area. Qualitative research suggests that users evaluate music and podcast search results in different ways, which posts special challenges to the UI design of systems hosting both content types. For the latter, *language, recency* and *topical information*, as well as a meaningful *presentation* of the information to enable a frictionless browsing experience, are highly valued by users when evaluating search results [2].

Misinformation Challenge. While fulfilling users' search goals in information seeking and streaming, the search system should proactively identify misinformation and unfairness, and safeguard users' wellness when mediating their interaction with information. This though is challenging when new content is produced at all times and cannot all be thoroughly audited.

2 SUMMARY

In this talk, we discuss domain-specific podcast search challenges in the context of the Spotify app. We believe that our challenges are general enough to not be unique to Spotify. They are broadly faced by multi-modal information access systems with heterogeneous information sources and vast catalogs. The challenges we present will be interesting to a wide range of IR researchers as they cover a variety of topics including evaluation, query intent understanding and search ranking.

3 SPEAKER BIOS

Mi Tian is a Staff Data Scientist at Spotify, based in Berlin. She is interested in user satisfaction understanding, evaluation of Search and IR systems, metrics and experimentation from an applied perspective. Mi has a PhD (2016) in audio based Music Information Retrieval and have a passion for connecting research insights with consumer products.

Claudia Hauff is a Staff Research Scientist at Spotify, based in the Netherlands. She has a PhD (2010) in Information Retrieval and spent more than 10 years in academia, supervising a team of junior researchers on a wide range of IR topics including conversational search, collaborative search, search as learning and axiomatic approaches to neural IR.

Praveen Chandar is a Staff Research Scientist at Spotify working on experimentation and evaluation of search & recommender systems. His research interests are in information retrieval, and recommendation systems with a focus on experimentation and evaluation. Praveen received his PhD (2014) from the University of Delaware, working on novelty and diversity aspects of search evaluation. He has published papers at top conferences including SIGIR, KDD, WSDM, WWW, CIKM, CHI, and UAI. He has also served as a program committee member at top conferences such as SIGIR, KDD, WSDM, etc., and has co-taught tutorials at NeurIPS 2020 & KDD 2021 on mixed method metric development.

REFERENCES

[1] Jana Besser, Martha Larson, and Katja Hofmann. 2010. Podcast search: User goals and retrieval technologies. *Online information review* (2010).
[2] Ben Carterette, Rosie Jones, Gareth F. Jones, Maria Eskevich, Sravana Reddy, Ann Clifton, Yongze Yu, Jussi Karlgren, and Ian Soboroff. 2021. Podcast Metadata and Content: Episode Relevance and Attractiveness in Ad Hoc Search. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 2247–2251. https://doi.org/10.1145/3404835.3463101
[3] Praveen Chandar, Jean Garcia-Gathright, Christine Hosey, Brian St. Thomas, and Jennifer Thom. 2019. Developing evaluation metrics for instant search using mixed methods methods. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval.* 925–928.
[4] Helia Hashemi, Aasish Pappu, Mi Tian, Praveen Chandar, Mounia Lalmas, and Benjamin Carterette. 2021. Neural Instant Search for Music and Podcast. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining.* 2984–2992.
[5] Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M. Henne. 2009. Controlled Experiments on the Web: Survey and Practical Guide. *Data Ming and Knowledge Discovery.* 18, 1 (feb 2009), 140–181. https://doi.org/10.1007/s10618-008-0114-1
[6] Mounia Lalmas. 2011. Aggregated Search. In *Advanced Topics in Information Retrieval*, Massimo Melucci and Ricardo Baeza-Yates (Eds.). The Information Retrieval Series, Vol. 33. Springer, 109–123. https://doi.org/10.1007/978-3-642-20946-8_5
[7] Ganesh Venkataraman, Abhimanyu Lad, Viet Ha-Thuc, and Dhruv Arya. 2016. Instant Search: A Hands-on Tutorial. In *SIGIR '16* (Pisa, Italy). ACM, 1211–1214.