MILLION SONG SEARCH: WEB INTERFACE FOR SEMANTIC MUSIC SEARCH USING MUSICAL WORD EMBEDDING

SeungHeon Doh[1] **Jongpil Lee**[2] **Juhan Nam**[1,2]
[1] Graduate School of Culture Technology, KAIST, South Korea
[2] Neutune Research, Seoul, South Korea
{seungheondoh, juhan.nam}@kaist.ac.kr, jongpillee@neutune.com

ABSTRACT

We present a web interface for large-scale semantic search using a musically customized word embedding in the backend. The musical word embedding represents artist entities, track entities, tags, and ordinary words in a single vector space. It is learned based on the affinity between the words and the entities using a wide spectrum of text data including Wikipedia, music review, and music tags. The system can allows users to type a query within 9.8M vocabulary words in the musical word embedding. It also supports a multi-query blending function using a semantic averaging of the queries to provide more refined search.

1. INTRODUCTION

Music search is an essential feature in streaming music services where users can readily access to large-scale music tracks. However, current music services mainly provide query-by-metadata which retrieves songs by artist, album, or track information. In order to broaden the scope of search query, semantic music search based on query-by-tag has been actively studied in the music information retrieval (MIR) community [1]. Semantic music search allows users to find songs using music related terms such as genre or mood. However, the majority of the previous works tackled it as a classification problem which allows only a limited number of tags.

Word embedding is a natural language processing (NLP) technique that represents a large set of vocabulary in a vector space such that semantically similar words are located closely and dissimilar ones are far apart. This opened up the possibility of associating audio with word through the mapping of their embedding spaces [2–4]. Recently, word embedding has been customized to include more music context using music tags, music review along with Wikipedia [5]. This approach extended the type of query to ordinary words in various music listening contexts.

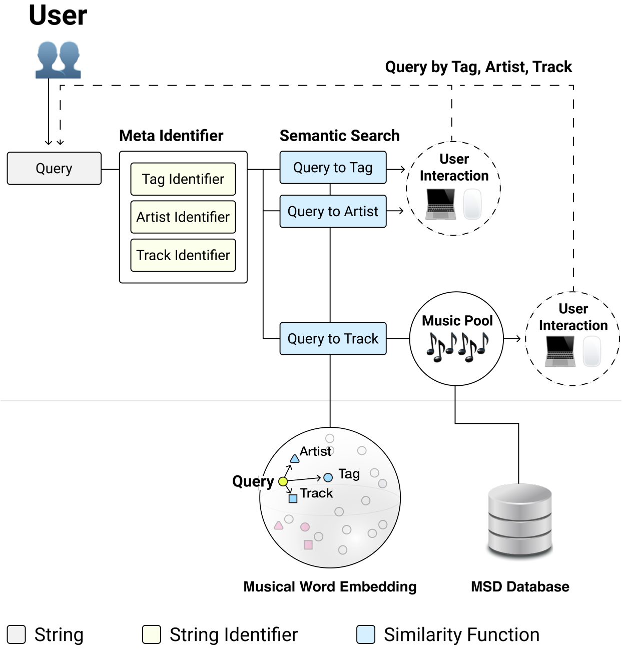


Figure 1. System architecture for Million Song Search.

Million Song Search is a web-based semantic music search system using the musical word embedding in the back-end [1] [2]. To design this system, we also upgraded the musical word embedding by including artist names and title names as semantic words. This updated musical word embedding allows users to type a query within 9.8M vocabulary words. The system also supports a multi-query blending function using a semantic averaging across different types of word terms (e.g., artist, track, tag, and general words) to provide more refined search.

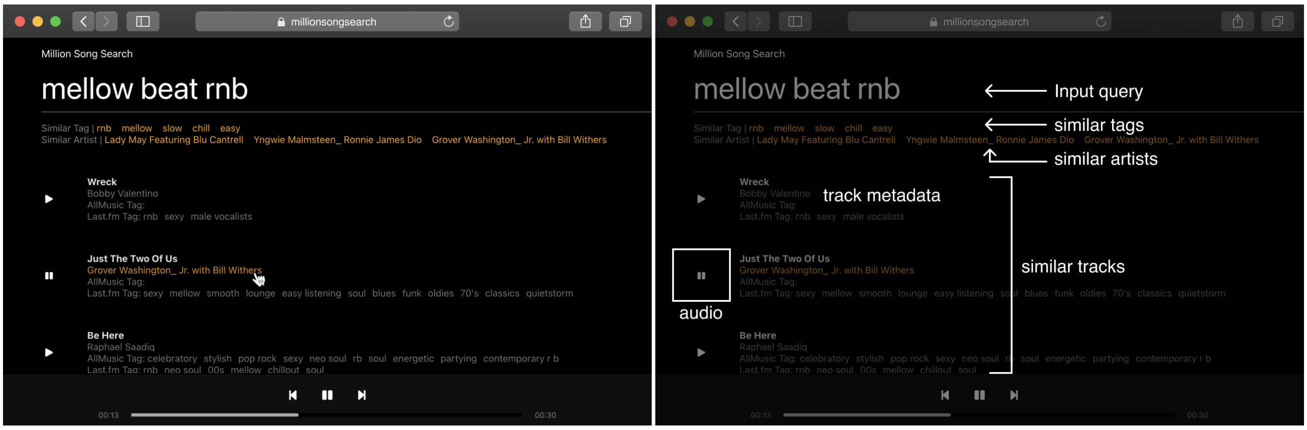
2. SYSTEM OVERVIEW

The power of the proposed web-based semantic search system is attributed to musical word embedding where every word (including artist or track IDs) is represented as a

Attribution: SeungHeon Doh, Jongpil Lee, and Juhan Nam, “Million Song Search: Web Interface for Semantic Music Search Using Musical Word Embedding”, in *Extended Abstracts for the Late-Breaking Demo Session of the 22nd Int. Society for Music Information Retrieval Conf.,* Online, 2021.

[1] We used the term “Million” because the artist and track information was obtained from Million Song Dataset (MSD) [6]. But, it also means that the system allows a large set of semantic query words.

[2] http://millionsongsearch.kaist.ac.kr/



millionsongsearch

Million Song Search

mellow beat rnb

Similar Tag | rnb mellow slow chill easy
Similar Artist | Lady May Featuring Blu Cantrell Yngwie Malmsteen_ Ronnie James Dio Grover Washington_ Jr. with Bill Withers

Wreck
Bobby Valentino
AllMusic Tag:
Last.fm Tag: rnb sexy male vocalists

Just The Two Of Us
Grover Washington_ Jr. with Bill Withers
AllMusic Tag:
Last.fm Tag: sexy mellow smooth lounge easy listening soul blues funk oldies 70's classics quietstorm

Be Here
Raphael Saadiq
AllMusic Tag: celebratory stylish pop rock sexy neo soul rb soul energetic partying contemporary r b
Last.fm Tag: rnb neo soul 00s mellow chillout soul

00:13 00:30

Figure 2. Million Song Search interface for *mellow beat rnb* query. Users can search over various musical words (e.g. tag, artist, track, etc.) from the semantic space.

dense vector and similarity scores between all word combinations can be calculated. Therefore, the system can take natural language-like queries. If the query has multiple words, the embedding vectors of the words are averaged and a similarity score between the query and other words is then calculated. The search results will be shown on the web interface panel. We show the results in three ranked lists for tags, artists, and tracks (see Figure 1). Therefore, the system can retrieve similar tag, artist, and track list for any type of query words. We used the following three submodules for building the web interface system.

Meta Identifier. The first module, meta identifier determines whether words in a query belong to an tag, artist, and track. If there exists the same name across tags, artists, and tracks, then it is difficult for the system to understand the user's intention. To solve this problem, we provide an auto-complete feature that allows the user to select the intended one. If the word does not belongs to tag, artist, or track, then it is regarded as a general word.

Semantic Search. The second module, semantic search is based on Musical Word Embedding. This model is trained with *Wikipedia*, *Amazon* album review, *AllMusic* tags, *Last.fm* tags, and artist/track IDs from the MSD dataset [6]. The entire collection provides 9.8M unique general words, 2,201 tags, 37,002 artists, and 0.7M tracks for the embedding space. We can then retrieve all items in this space by measuring the similarity score between the two items. We visualize a UMAP plot of the embedding space in Figure 3. We can see that they are mixed well, which makes various search scenarios possible.

Query by Tag, Artist, and Track. The last module provides a click interaction from the search results. Once the search results are displayed, users may want to explore further through the searched items. In this case, when the users click on one of the searched items, a new page of search results is presented based on the new query. This feature helps users keep finding similar music.

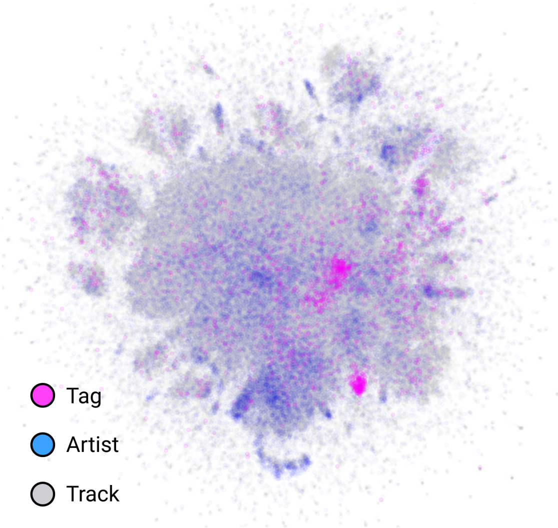


Figure 3. UMAP visualization of tag, artist, track embedding vectors in musical word embedding. We can see that they blend well. And, this makes the cross-search possible.

3. CONCLUSIONS & FUTURE WORK

We propose a web interface *Million Song Search* that supports semantic music search using musical word embedding. The system has a feature that quantifies the search experience of users. As future work, we will collect queries, query sequences, and listening time of users, and evaluate the user satisfactions of query-by-text task. Also, we plan to use the user query for active learning.

4. REFERENCES

[1] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, "Semantic annotation and retrieval of music and sound effects," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 467–476, 2008.

[2] J. Choi, J. Lee, J. Park, and J. Nam, "Zero-shot learning for audio-based music classification and tagging,"

in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2019, pp. 67–74.

[3] K. Watanabe and M. Goto, "Query-by-blending: a music exploration system blending latent vector representations of lyric word, song audio, and artist," in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2019, pp. 144–151.

[4] M. Won, S. Oramas, O. Nieto, F. Gouyon, and X. Serra, "Multimodal metric learning for tag-based music retrieval," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[5] S. Doh, J. Lee, T. H. Park, and J. Nam, "Musical word embedding: Bridging the gap between listening contexts and music," *Machine Learning for Media Discovery Workshop, International Conference on Machine Learning (ICML)*, 2020.

[6] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere, "The million song dataset," in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2011.