Muse-it: A Tool for Analyzing Music Discourse on Reddit

17 June 2025

Summary

Muse-it is a user-friendly software tool that helps researchers collect and analyze online discussions about music from Reddit. The software allows users to search for specific topics and gather posts and comments from across the platform into a csv file. Muse-it can then highlight patterns in the discussions—such as emotions, themes, and musical preferences—and link them to related tracks on Spotify. It also creates interactive visualizations to help users explore trends and group similar topics together. By focusing on naturally occurring online conversations, Muse-it makes it easier to study how people talk about music in real-life settings, without the influence of laboratory environments or questionnaires. While many of Muse-it's features are tailored to music research, the tool is flexible and can be used by researchers in other fields interested in analyzing Reddit discourse on any topic.

Introduction

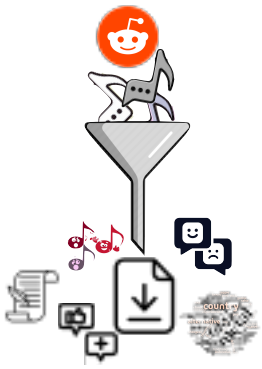

Figure 1: We introduce Muse-it, a tool to retrieve, process, visualize, and analyze music-related Reddit discourse using search queries. Furthermore, it provides

metadata of the tracks, albums, and playlists linked in the extracted data for further analysis.

Reddit is a social media platform built around user-created communities called subreddits, where users can interact anonymously. While other platforms impose character limits or induce social desirability bias, Reddit's anonymity encourages users to share experiences freely and openly. Users often post links to personal music choices, including playlists, providing an organic source of data on musical experiences and relationships with music.

A range of studies has demonstrated the versatility of Reddit data within MIR and music perception and cognition. For example, Cao and Bogers (2024) scraped ~5000 threads from r/ifyoulikeblank to compare human and algorithmic music recommendations, highlighting extra-musical factors in human suggestions. ASMR and music-induced frisson were analyzed via Reddit comments in Kovacevich and Huron (2019), supporting Huron's frisson theory Huron (2006).

Reddit's anonymity enables vulnerable populations to discuss sensitive topics openly. Singh (2023) analyzed r/depression and r/depressionMusic to identify music-listening strategies and their acoustic and lyrical properties. Varghese, Carlson, and Alluri (2024) examined r/autism discussions, revealing pop and electronica preferences among individuals with ASD, contrasting earlier findings. These studies showcase the value of first-person accounts in music psychology research.

The presence of subreddits also allows exploration of niche communities. Studies of genre-specific forums (e.g., metal, jazz) have examined identity construction and emotional associations Mishra et al. (2021). Veselovsky, Waller, and Anderson (2021) created a 1.3M-instance dataset of music sharing on Reddit, exploring macro-scale cultural dimensions of online music sharing.

Despite these insights, MIR research has yet to fully exploit Reddit data due to technical barriers in data retrieval, processing, and organization. Muse-it addresses this gap by streamlining the retrieval, analysis, and visualization of Reddit data for music research, lowering the technical barrier for interdisciplinary studies.

Statement of Need

Traditional laboratory-based research into music behavior and cognition often suffers from biases such as social desirability, demand characteristics, and limited ecological validity. Naturalistic settings like Reddit offer an opportunity to study music engagement as it organically occurs, reducing these biases. In recent years, Reddit has grown as a valuable resource for social science research due to its large user base, topic-specific communities, and candid discourse. However, its use in musicology remains limited. This is not due to a lack of potential, but rather the technical barrier to accessing and analyzing Reddit data. Easier access

to Reddit data could therefore unlock new possibilities for music researchers, enabling wider participation in large-scale, ecologically valid research.

Assembling and processing Reddit data, integrating it with music metadata, and generating meaningful visualizations pose significant technical challenges. Existing tools such as PRAW (Botzer, Gu, and Weninger 2022) facilitate basic Reddit data extraction, and SpotDL (Singh 2023) retrieves track metadata, but no unified framework bridges these tasks for music-focused research.

Moreover, a need for dialogue between Music Information Retrieval (MIR) and music psychology was raised over a decade ago (Aucouturier and Bigand 2013, 2012), but remains relevant even today. MIR often emphasizes computational accuracy and engineering performance, while overlooking psychological and contextual dimensions of music experience. Conversely, music cognition research tends to underutilize the large-scale analytical capabilities that MIR offers. Muse-it is designed to bridge this gap by providing a scalable, naturalistic framework for studying musical discourse that can be used by both MIR practitioners and psychological researchers.

Existing tools such as PRAW (Boe 2021) facilitate basic Reddit data extraction, and SpotDL ("spotDL: Spotify Downloader" 2016) retrieves track metadata, but no unified framework bridges these tasks for music-focused research. Muse-it addresses these gaps by offering:

- Automated retrieval of Reddit posts and comments matching custom queries across multiple subreddits.
- Optional NLP-based annotation of themes, emotions, and sentiment using TweetNLP (Camacho-Collados et al. 2022; Antypas and Camacho-Collados 2023).
- Topic modeling and hierarchical clustering via BERTopic (Camacho-Collados et al. 2022) and Sentence-BERT (Reimers and Gurevych 2020).
- Extraction of Spotify metadata (artist, genre, release date, popularity, lyrics) through SpotDL integration.
- A cross-platform GUI (SolidJS frontend, Flask backend) for configuring queries, filtering results, and generating interactive visualizations.

Muse-it's ability to organically combine Reddit discourse with rich track metadata from Spotify allows researchers to bridge the gap between qualitative insights and quantitative attributes. Muse-it can therefore serve as an invaluable tool for mixed-methods research in the field of musicology that empowers researchers to validate existing inferences in more ecologically valid settings and generate novel hypotheses. By lowering the technical barriers to big data analytics, Muse-it enables a broader community of researchers to explore the nuanced relationships between online discourse and music. It is also worth noting that the tool can be used by researchers outside the field of musicology for data retrieval and analysis.

Muse-it architecture and functionality

Muse-it comprises four modules (Figure 1):

1. **Data retrieval from Reddit**
2. **NLP-based metadata extraction (emotion, sentiment, topics)**
3. **Visualization and hierarchical topic clustering**
4. **Spotify track metadata extraction via SpotDL**

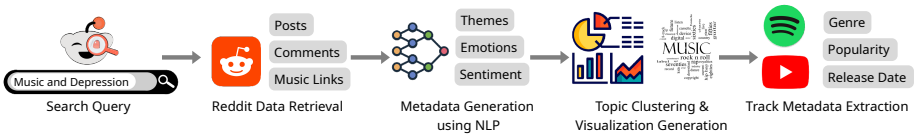

Figure 2: Muse-it pipeline.

Reddit Data Retrieval

Data retrieval uses the Reddit API (Developers 2008) and PRAW (Boe 2021). Muse-it searches for subreddits matching a query, allows user selection, and retrieves up to 1000 posts per subreddit. Filters for date range, comments, and Spotify URLs are provided. Metadata retrieved includes subreddit name, title, body, URL, comment count, and creation time. Spotify URLs in posts and (optionally) comments are identified for later metadata extraction. Retrieval of 10,000 posts typically takes 5–10 minutes.

Metadata Generation using NLP

Post titles are processed to generate themes, emotions, and sentiment using TweetNLP (Camacho-Collados et al. 2022). Titles are favored for efficiency. NLP processing of 10,000 posts takes ~4 minutes.

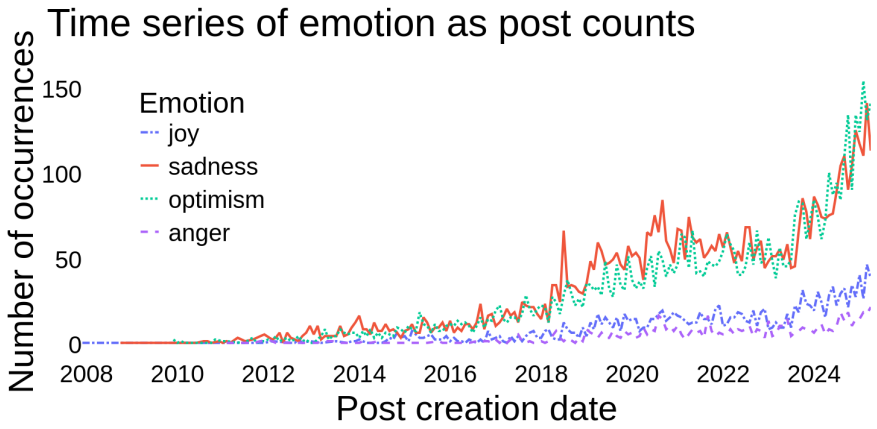

Figure 3: Monthly time-series of emotions for "sad music" posts.

Topic Clustering and Visualization

Muse-it visualizes metadata distributions and temporal trends. It uses BERTopic (Grootendorst 2022) with Sentence-BERT embeddings (Reimers and Gurevych 2020) for hierarchical clustering and 2D topic maps.

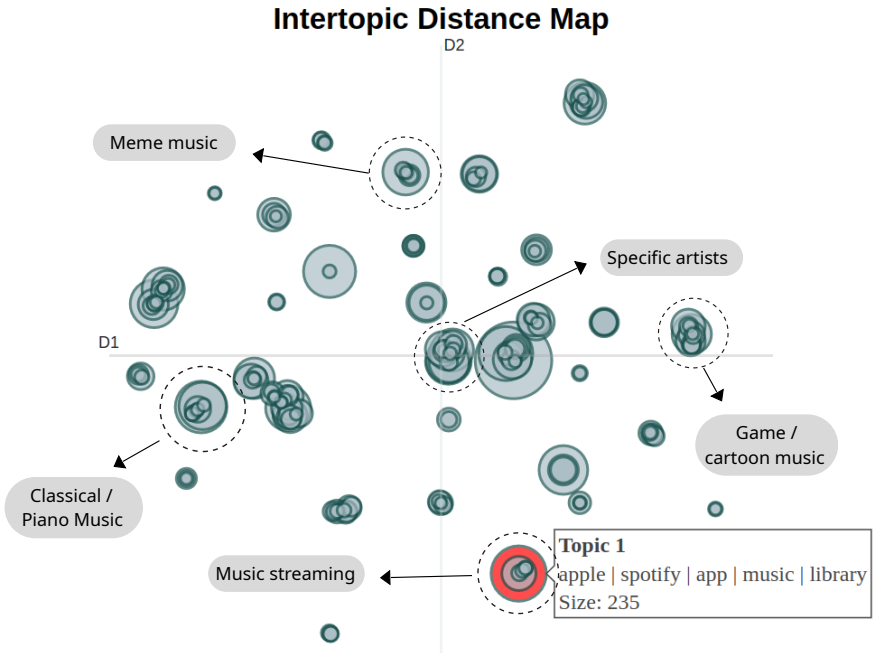

Figure 5: Hierarchical clustering of topics.

It also created wordcloud visualizations of the retrieved text.

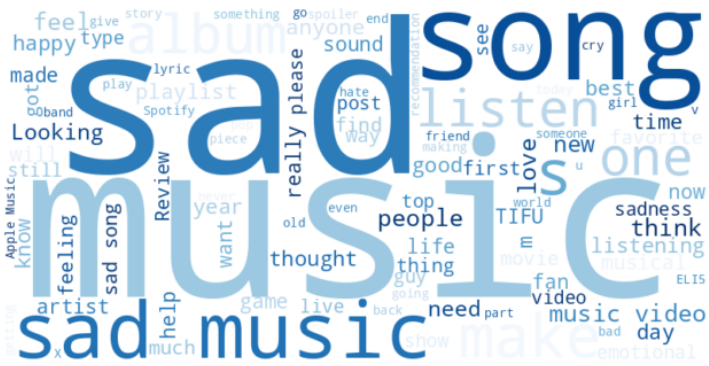

Figure 4: Word cloud for "sad music" posts.

Track Metadata Extraction

Spotify URLs are processed with SpotDL ("spotDL: Spotify Downloader" 2016) to extract track, album, and playlist metadata (e.g., name, artist, album, genres, release date, lyrics). Metadata extraction of 100 tracks takes ~2 minutes; timeouts prevent excessively long processing.

Data Download

Muse-it outputs a primary CSV (`data.csv`) with Reddit content and metadata, and separate CSVs for Spotify metadata organized by URI (Figure 6). CSV sizes vary but are typically tens of megabytes.

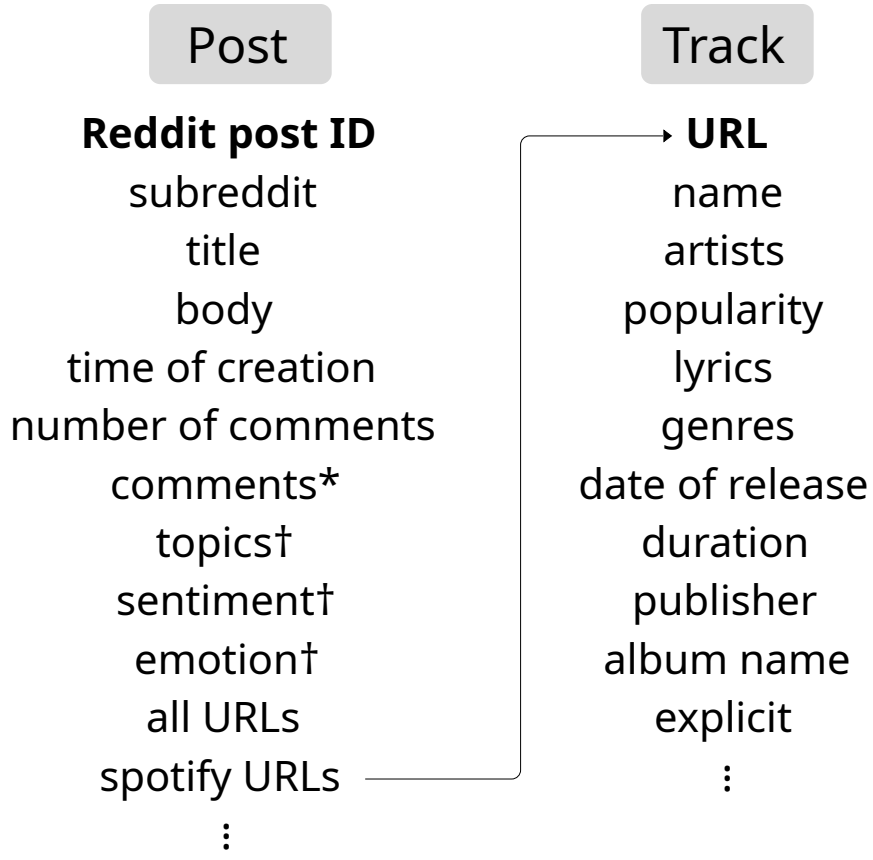

Figure 6: Key CSV headers for Reddit and Spotify metadata.

Comparison to Existing Software

While PRAW provides basic Reddit API access and SpotDL handles Spotify metadata, neither offers a unified pipeline for music-focused discourse analysis

for researchers. By automating the retrieval and organization of online discourse, Muse-it reduces the complexity that is traditionally associated with big data. It enables researchers to focus on extracting meaningful insights rather than grappling with data curation challenges.

Installation

1. Download the file appropriate for your platform from https://github.com/george-paul/MuseIt/releases.
2. Extract the zip file.
3. Obtain a Reddit API Secret and ID
4. . Populate this in the `config.yaml` file.
5. Open the launch script. This file will be named `launch-<platform>`.

This will open a console window in which the backend is running, as well as a browser windows through which one can interact with the frontend. The console window will show the progress of each request.

Usage

1. Enter search query (e.g., "sad music"), select subreddits, date range, and filters.
2. Click "Run" to retrieve and process data.
3. Explore visualizations for themes, sentiments, emotions, topics, and trends.
4. Download CSV files for further analysis.

We have created a demo of the tool along with the data used to create the visualizations, which can be found at https://drive.google.com/drive/folders/1DtQZ7oRwrp4LPSu8-EajPTWX_l93WOs4?usp=sharing

Instructions on how to use, develop, build, and test Muse-it can be found in the README of the repository.

Acknowledgements

We thank the developers of PRAW, SpotDL, TweetNLP, BERTopic, and Sentence-BERT for their open-source contributions.

References

Antypas, Dimosthenis, and Jose Camacho-Collados. 2023. "Robust Hate Speech Detection in Social Media: A Cross-Dataset Empirical Evaluation." https://arxiv.org/abs/2307.01680.

Aucouturier, Jean-Julien, and Emmanuel Bigand. 2012. "Mel Cepstrum & Ann Ova: The Difficult Dialog Between MIR and Music Cognition." *Proceedings*

of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012, January.

———. 2013. "Seven Problems That Keep MIR from Attracting the Interest of Cognition and Neuroscience." *J. Intell. Inf. Syst.* 41 (3): 483–97. https://doi.org/10.1007/s10844-013-0251-x.

Boe, Bryce. 2021. "PRAW: The Python Reddit API Wrapper." *GitHub Repository*. GitHub. https://github.com/praw-dev/praw.

Botzer, Nicholas, Shawn Gu, and Tim Weninger. 2022. "Analysis of Moral Judgment on Reddit." *IEEE Transactions on Computational Social Systems* 10 (3): 947–57.

Camacho-Collados, Jose, Kiamehr Rezaee, Talayeh Riahi, Asahi Ushio, Daniel Loureiro, Dimosthenis Antypas, Joanne Boisson, et al. 2022. "TweetNLP: Cutting-Edge Natural Language Processing for Social Media." In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Abu Dhabi, U.A.E.: Association for Computational Linguistics.

Cao, Thi Binh Minh, and Toine Bogers. 2024. "'If i Like BLANK, What Else Will i Like?': Analyzing a Human Recommendation Community on Reddit." In *Wisdom, Well-Being, Win-Win*, edited by Isaac Sserwanga, Hideo Joho, Jie Ma, Preben Hansen, Dan Wu, Masanori Koizumi, and Anne J. Gilliland, 70–83. Cham: Springer Nature Switzerland.

Developers, Reddit. 2008. "Reddit API." *Reddit Developers*. Reddit. https://developers.reddit.com/docs/0.9/api.

Grootendorst, Maarten. 2022. "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." *arXiv Preprint arXiv:2203.05794*.

Huron, David. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. The MIT Press. https://doi.org/10.7551/mitpress/6575.001.0001.

Kovacevich, Alexsandra, and David Huron. 2019. "Two Studies of Autonomous Sensory Meridian Response (ASMR): The Relationship Between ASMR and Music-Induced Frisson." *Empirical Musicology Review* 13 (January): 39. https://doi.org/10.18061/emr.v13i1-2.6012.

Mishra, Vipul, Kongmeng Liew, Elena V. Epure, Romain Hennequin, and Eiji Aramaki. 2021. "Are Metal Fans Angrier Than Jazz Fans? A Genre-Wise Exploration of the Emotional Language of Music Listeners on Reddit." In *Proceedings of the 2nd Workshop on NLP for Music and Spoken Audio (NLP4MusA)*, edited by Sergio Oramas, Elena Epure, Luis Espinosa-Anke, Rosie Jones, Massimo Quadrana, Mohamed Sordo, and Kento Watanabe, 32–36. Online: Association for Computational Linguistics. https://aclanthology.org/2021.nlp4musa-1.7/.

Reimers, Nils, and Iryna Gurevych. 2020. "Making Monolingual Sentence Embeddings Multilingual Using Knowledge Distillation." In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://arxiv.org/abs/2004.09813.

Singh, Kunal AND Paruchuri, Bhavyajeet AND Vaswani. 2023. "'Help! I Need Some Music!': Analysing Music Discourse & Depression on Reddit." *PLOS*

ONE 18 (7): 1–16. https://doi.org/10.1371/journal.pone.0287975.

"spotDL: Spotify Downloader." 2016. *GitHub Repository*. GitHub. https://github.com/spotDL/spotify-downloader.

Varghese, Sharon, Emily Carlson, and Vinoo Alluri. 2024. "LISTENING TO THE SPECTRUM: EXPLORING MUSIC PREFERENCES AND ASSOCIATIONS IN AUTISM SPECTRUM DISORDER COMMUNITIES ON REDDIT." https://doi.org/10.13140/RG.2.2.26401.62564.

Veselovsky, Veniamin, Isaac Waller, and Ashton Anderson. 2021. "Imagine All the People: Characterizing Social Music Sharing on Reddit." *Proceedings of the International AAAI Conference on Web and Social Media* 15 (1): 739–50. https://doi.org/10.1609/icwsm.v15i1.18099.