Enabling Music Search and Analysis: A Database for Symbolic Music Files

Yaolong Ju, McGill University
Gustavo Polins Pedro, McGill University
Cory McKay, Marianopolis College
Emily Hopkins, McGill University
Julie Cumming, McGill University
Ichiro Fujinaga, McGill University

Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)

Music Encoding Conference 2019
2019.5.30
Two questions

(1) How to store symbolic music files?
   Example: Beethoven’s Third Symphony — a musical work
   • Different movements
   • Different formats
   • Different encoding settings
   • Different sources

(2) As the database continues to grow, how to search music?
Introduction

Symbolic music files are invaluable resources for music research:

- Harmonic analysis (Condit-Schultz et al. 2018)
- Composer attribution (McKay et al. 2017b)

Using automated feature extraction software (e.g., jSymbolic), statistical analysis, and machine learning, symbolic music data allows us to study large quantities of music.

We need numerous, high-quality symbolic music files!
Existing symbolic music databases

Classical Archives
Musescore
ChoralWiki
The SEILS dataset
Kern Scores
Musedata
The Josquin Research Project
Introducing the SIMSSA* DB

*: Single Interface for Music Score Searching and Analysis
Introducing the SIMSSA* DB

Modelling Bibliographic Metadata for Music
Tracking Provenance
Searching Musical Content
Archiving Research

*: Single Interface for Music Score Searching and Analysis
Introducing the SIMSSA DB

Modelling Bibliographic Metadata for Music

• Allows modelling of complex relationships
Data model for complex relationships (McKay et al. 2017a)
Data model for complex relationships (McKay et al. 2017a)

We drew inspirations from FRBR* and IFLA-LRM**

*: Functional Requirement for Bibliographic Records
**: International Federation of Library Associations and Institutions-Library Reference Model
Data model for complex relationships (McKay et al. 2017a)
Uses authority control and controlled vocabulary

To ensure the quality of metadata:

• We use VIAF (Virtual International Authority File) entries for composers
  • Each entry is provided with linked data URIs (Uniform Resource Identifiers), which unambiguously identifies a resource
  • We also use controlled vocabulary for genres and instrumentation

This helps us:

• Guard against typographical errors
• Manage variant spellings of fields
• Increase interoperability
• Allow for the semantic query of data on a larger scale
Auto-suggest

Genre(s)

What type of piece is this? (e.g., song, symphony, motet)

- Madrigal
- Motet
- Create "M"
Bibliographic metadata for music & Upload

Contributions

Who created the work? Use the drop-down menu to choose between different kinds of contributions. Add more contributors with the green button.

Name:

Role:

Composer
Arranger
Author of Text
Transcriber
Improviser
Performer
Unknown

Location:

Date:
Tracking provenance

Specifies where the symbolic files are coming from

Where did these files come from?

Please indicate the provenance of the file you are uploading. If the music is part of a larger collection with multiple works (e.g., a complete edition of The Well-Tempered Clavier I, a Fake Book, or an online collection), please include the title of this collection below. If the music you are submitting is from a stand-alone work (e.g., a score of Bach's "Prelude and Fugue No. 2 in C minor"), then "Title" will be the same as the musical work, but you should still fill in the other fields.

Title of Collection *

Collection URL (if applicable):

Archive/Library where this source can be found (optional):

Portions:

*: International Music Score Library Project
Search

Example: retrieve all pieces with:
• “Missa ave” in the title (search for title)
• Midi format (search for additional metadata)
• With vertical tritone (search for musical content)
Search for title
Search for additional metadata

These musical contents (features) are automatically extracted with jSymbolic (McKay et al. 2018)
Search for musical content
Research archives

The work introduced above encourages high-quality data input

Once we use the database for research, how to archive it for:

- Our own future use
- For others to reproduce our results or conduct their own studies
Research archives

We use Zenodo to include a static dump of the music files as studied

- Zenodo is an open-access platform for “release quality” datasets
- Generate a DOI (Digital Object Identifier) for a stable dataset for citation

SIMSSA DB is great for:

- Finished corpora
- Store and search for metadata and musical contents
Research archive example (Zenodo)
Existing symbolic music files in SIMSSA DB

Currently, we have three datasets:

- Josquin La Rue Secure Duos Dataset (Cumming et al. 2018)
- Seven Renaissance Composers Dataset (McKay et al. 2017b)
- Florence 164 Dataset (Cumming and McKay, 2018)

700 musical works and ~2000 symbolic music files and still growing!

Once publicly available, we need your contribution!
Conclusion

SIMSSA DB: A high-quality database for symbolic music files search and analysis

- Permits the storage and distribution of a wide range of music in various symbolic formats
- Provides meaningfully structured metadata
- Uses authority control to ensure the quality of metadata
- Offers content-based search
- Emphasizes the provenance of resources
- Facilitates the archiving of research experiments
- Enables users to upload symbolic music
Future Work

Batch download and upload

Add more high-quality symbolic files, for example:

- Bach, Praetorius (Condit-Schultz et al. 2018) and Schutz chorales (721)
  - 721 chorales, in Kern and musicXML formats
- Late Medieval Liturgical Offices by Andrew Hughes (~5900)
  - About 5900 chant melodies
  - Translated and converted into MEI format (Ju and Helsen 2018)

Incorporate linked data
Future Work

Official release, made publicly accessible

Project URL: http://db.simssa.ca

Contact us: @simssaproject on Twitter!

We need your contribution!
Enabling Music Search and Analysis: A Database for Symbolic Music Files
References


- McKay, Cory, Tristano Tenaglia, Julie Cumming, and Ichiro Fujinaga. 2017b. “Using Statistical Feature Extraction to Distinguish the Styles of Different Composers.” Presented at the Medieval and Renaissance Music Conference, Prague, Czech Republic.