The OmniOMR Project

Jan Hajič, jr.
Institute of Formal and Applied Linguistics (ÚFAL)
Faculty of Mathematics and Physics, Charles University
Prague, Czech Republic
hajicj@ufal.mff.cuni.cz

Petr Žabička
Moravian Library
Brno, Czech Republic
petr.zabicka@mzk.cz

Jan Rychtář
Trinera, s.r.o.
Brno, Czech Republic
jan.rychtar@trinera.cz

Jiří Mayer
ÚFAL
Charles University
Prague, Czech Republic
mayer@ufal.mff.cuni.cz

Martina Dvořáková
Moravian Library
Brno, Czech Republic
martina.dvorakova@mzk.cz

Filip Jebavý
Moravian Library
Brno, Czech Republic
filip.jebavy@mzk.cz

Markéta Vlková
Moravian Library
Brno, Czech Republic
marketa.vlkova@mzk.cz

Pavel Pecina
ÚFAL
Charles University
Prague, Czech Republic
pecina@ufal.mff.cuni.cz

***Abstract*—In March 2023, the OmniOMR project started with the aim of applying OMR technologies in a digital library environment. The project aims to provide full-text search functionality in musical documents in diverse library collections, covering primarily Common Western Music Notation, but (white) mensural and diastematic chant notations as well, both printed and handwritten. To provide this capability, we aim to build an "OmniOMR Service" – not just for recognition, but also the preceding step of music notation detection and classification, as not all documents that contain music notation are marked in library collections with the appropriate metadata. A significant condition for the project's success is building a dataset that accurately reflects the challenges present in a broad library collection. This report lays out in more detail the goals of the project, its current status, and most importantly, we outline the opportunities for collaboration across the OMR community.**

***Index Terms*—Optical music recognition, digital library, music information retrieval**

I. Project Overview

The Moravian Library, like many others, maintains an extensive digitized collection that contains thousands of known music documents, ranging from medieval chant books over baroque manuscript scores and parts to prints of songs with piano accompaniments for early 20th century households and jazz lead sheets. On top of that, it contains an unknown number of music documents that have so far not been marked as such in the library's metadata scheme, especially in periodicals from the 19th or 20th centuries. In order to provide musical full-text search and music discovery functionality to its users, the library has sought out the OMR team at the Faculty of Mathematics and Physics of Charles University in Prague, resulting in the OmniOMR project.[1]

The OmniOMR applied research project aims to integrate OMR technology into the digital library environment to support these goals of providing access to and stewardship of this segment of cultural heritage in the digital domain, with an outlook of generalizing the developed models to other institutions as well. The project runs from March 2023 until the end of 2027, at a scale of roughly 5 full-time equivalents spread across two institutions: the Faculty of Mathematics and Physics (FMP) of the Charles University, in Prague, and the Moravian Library (MorL), based in Brno. The FMP is responsible for developing a functioning OMR service and overall project leadership, the MorL is responsible for developing search functionality and user interfaces, and for developing the legal framework under which images from as many institutions as possible can be processed and made searchable by the OMR service.

We aim to tackle a wide variety of input conditions (see Fig. 1), but only target the first two levels of comprehension: metadata extraction and search [1].

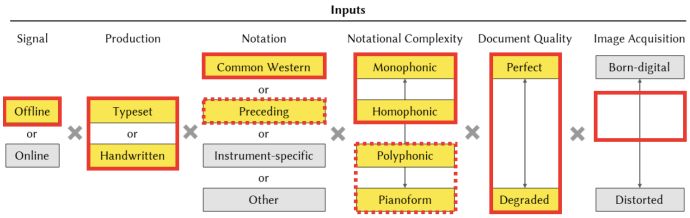

Fig. 1. Targeted music notation input condition.

The project is in its aims comparable to HISPAMUS [2],[2] although possibly with more weight on polyphonic manuscripts and search, and not as much on MEI/MusicXML output. We aim to utilize the available results of the SIMSSA [3] and OMMR4All projects [4],[3] for processing chant notation.

II. OMR Component

The OMR service will perform detection of music notation and its recognition as two separate steps, because each of these steps is subject to a different set of constraints. The detection step is significantly easier to train, but must be

[1]https://ufal.mff.cuni.cz/grants/omniomr

[2]https://grfia.dlsi.ua.es/hispamus/index.html

[3]https://simssa.ca/, https://ommr4all.informatik.uni-wuerzburg.de/en/

very fast. The MorL collection in which we aim to discover unmarked music notation has approximately 10 million pages. Processing such a collection for one entire calendar year would require a speed of 3 images per second. (Parallelization helps, but computational resources are scarce and the process likely will not run smoothly for an entire year.) Furthermore, due to expected class imbalance (less than 1 in 1000 pages is expected to contain music notation), especially precision on the "contains notation" class must be extremely high – each percentage point of false positives translates to 100 000 pages with music notation falsely detected, likely more than there are actual pages with music notation that do not have the appropriate metadata. Fortunately, music notation has a clear visual identity, so library-quality scans could be compressed without sacrificing accuracy.

The recognition step is subject to less severe computational efficiency and accuracy challenges. However, the task itself is more complex than binary classification and detecting relatively large rectangles. Computational resource constraints again apply, even though only tens of thousands of pages are expected to be processed in the recognition step, as it requires more complex models and may not admit reducing image quality as readily.

III. Search

The second technical aspect of the project is its musical search interface. Designing and selecting distance metrics in the presence of OMR mistakes and polyphonic music is in fact the most open research sub-topic in the project, and at the same time a key element of the user experience. To avoid assumptions of the project team [5], we are conducting a multi-stage user study with a focus group of digital music library users from various stakeholder groups (primarily musicians amateur and professional, teachers, and musicologists) to establish the UX priorities. So far, in the first stage, most significant is the interest in incipit- and keyboard-based search (already explored in the Musiconn project [6] or RISM [7]),[4] each soliciting enthusiasm from (different) 5 out of the 11 users.[5] The user study has implications beyond just the search interface: the focus on incipits, for instance, warrants OMR focusing on correctly identify where compositions start on a page (and which pages contain no start).

IV. Data

No public datasets sufficiently representative of the problem space are available for evaluating, much less training OMR systems. Therefore, while the MorL provides a representative set of digitized images of sheet music, FMP is responsible for annotating at least 1000 of these images with OMR ground truth. We are using the Labelbox web application.[6] The technical debt of MUSCIMarker [8], is too large, and outsourcing annotation software development and maintenance lightens the workload on the project development team. Labelbox was selected because it supports relations between objects in images, which is essential for the Music Notation Graph (MuNG) ground truth format [9] (at least for evaluation purposes, as notehead detection accuracy is one of the technical KPIs). Furthermore, Labelbox offers a free license for academic use.

V. Opportunities for Collaborations

We see multiple topics on which it seems prudent to consult and collaborate with the OMR community, and some of our outputs can be broadly useful for planning follow-up projects.

Output encodings. It would be a missed opportunity for the OMR community to make the dataset useful only for our limited levels of comprehension. A discussion is urgent on the ground truth formats (particularly improving MuNG, also in light of the new DoReMi dataset [10]), especially other projects that are in the process of data acquisition.

Data. We anticipate publishing the manually annotated data in yearly updates. Furthermore, as we aim for a broad applicability of our models, we are open to directing some of our annotation resources towards images provided by other OMR stakeholders.

Data synthesis. One prerequisite for the success of OmniOMR is building a synthetic data generator that can simulate full pages under the various input conditions and enabling the user to specify a sample document that should be imitated. We are extending the already published Mashcima system [11].

OMR service and indexing for search. We aim to make the recognition service publicly available (up to some manageable scale), with the option to index a user-provided collection alongside all other indexed collections of music documents. This is to enable tasks such as identifying anonymous compositions by comparing them across libraries or mapping transmission of repertoire, such as demonstrated by the F-TEMPO project for menusral part books.[7] Coordination on persistent music notation image identifiers is thus welcome, especially in light of existing projects' work on a broader systematization of musical culture in the digital domain (i.a. Trompa [12], LinkedMusic [13],[8] as well as on legal frameworks for delivering music notation images to users.

User study. Through participation in the user study, we can take into account the manifold priorities of the OMR and the broader music library communities.

It is our hope that the OmniOMR project significantly advances the state of OMR in the following years. In order to best utilize whatever resources we have available to achieve that, we believe that coordination with the broader OMR community is needed – and very much welcome.

Acknowledgment

This paper was supported by the applied research grant of the Czech Ministry of Culture programme NAKI III no. DH23P03OVV008, OmniOMR - optical music recognition using machine learning for digital libraries.

[4] https://rism.online/?mode=incipits

[5] Our target focus group size is, however, 20.

[6] https://labelbox.com/

[7] https://f-tempo.org/

[8] https://trompamusic.eu/, https://linkedmusic.ca/

REFERENCES

[1] J. Calvo-Zaragoza, J. Hajič Jr., and A. Pacha, "Understanding optical music recognition," *ACM Comput. Surv.*, vol. 53, no. 4, 2020.

[2] J. M. Iñesta, P. J. P. de León, D. Rizo, J. Oncina, L. Micó, J. R. Rico-Juan, C. Pérez-Sancho, and A. Pertusa, "Hispamus: Handwritten spanish music heritage preservation by automatic transcription," in *1st International workshop on reading music systems*, pp. 17–18, 2018.

[3] I. Fujinaga, A. Hankinson, and J. E. Cumming, "Introduction to simssa (single interface for music score searching and analysis)," in *Proceedings of the 1st international workshop on digital libraries for musicology*, pp. 1–3, 2014.

[4] C. Wick and F. Puppe, "Ommr4all-a semiautomatic online editor for medieval music notations," in *2nd International workshop on reading music systems*, pp. 31–34, 2019.

[5] A. Green and F. Lawrence, "The shock of the new: Testing the pan-archival linked data catalogue with users.," in *Proceedings of the Workshops and Doctoral Consortium of the 26th International Conference on Theory and Practice of Digital Libraries*, pp. 108–115, 2022.

[6] J. Diet, "Innovative mir applications at the bayerische staatsbibliothek," in *Proceedings of the 5th International Conference on Digital Libraries for Musicology*, 2018.

[7] K. Keil and L. Pugin, "Das internationale quellenlexikon der musik, rism," *Bibliothek Forschung und Praxis*, vol. 42, no. 2, pp. 309–318, 2018.

[8] J. Hajič jr. and P. Pecina, "Groundtruthing (not only) music notation with MUSICMarker: A practical overview," in *14th International Conference on Document Analysis and Recognition*, (Kyoto, Japan), pp. 47–48, 2017.

[9] J. Hajič jr. and P. Pecina, "The MUSCIMA++ dataset for handwritten optical music recognition," in *14th International Conference on Document Analysis and Recognition*, (Kyoto, Japan), pp. 39–46, 2017.

[10] E. Shatri and G. Fazekas, "Doremi: First glance at a universal omr dataset," in *Proceedings of the 3rd International Workshop on Reading Music Systems* (J. Calvo-Zaragoza and A. Pacha, eds.), (Alicante, Spain), pp. 43–49, 2021.

[11] J. Mayer and P. Pecina, "Synthesizing training data for handwritten music recognition," in *Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part III 16*, pp. 626–641, Springer, 2021.

[12] D. M. Weigl, T. Crawford, A. Gkiokas, W. Goebl, G. Emilia, N. Guti, C. C. Liem, and P. Santos, "Fair interconnection and enrichment of public-domain music resources on the web," *Empirical Musicology Review*, vol. 16, no. 1, pp. 16–33, 2021.

[13] L. Pugin and C. Bacciagaluppi, "An analysis of musical work datasets and their current level of linkage," in *7th International Conference on Digital Libraries for Musicology*, pp. 32–39, 2020.