READ

RECOGNITION & ENRICHMENT OF ARCHIVAL DOCUMENTS

D8.3

Open Innovation Forum P3 (DocScan and ScanTent)

Florian Kleber, Markus Diem, Stefan Fiel, and Günter Mühlberger
CVL

Distribution: http://read.transkribus.eu/

READ
H2020 Project 674943

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 674943

Project ref no.	H2020 674943
Project acronym	READ
Project full title	Recognition and Enrichment of Archival Documents
Instrument	H2020-EINFRA-2015-1
Thematic priority	EINFRA-9-2015 - e-Infrastructures for virtual research environments (VRE)
Start date/duration	01 January 2016 / 42 Months

Distribution	Public
Contract. date of delivery	31.12.2018
Actual date of delivery	28.12.2018
Date of last update	18.12.2018
Deliverable number	D8.3
Deliverable title	Open Innovation Forum P3 (DocScan and ScanTent)
Type	Report, Demonstrator
Status & version	in progress
Contributing WP(s)	WP8
Responsible beneficiary	CVL
Other contributors	CVL, UIBK
Internal reviewers	UCL, StAZh
Author(s)	Florian Kleber, Markus Diem, Stefan Fiel, and Günter Mühlberger
EC project officer	Christopher DOIN
Keywords	Crowd-Scanning, Android App, Transkribus

Contents

1 Executive Summary

The Open Innovation Forum focuses on the development of the ScanTent and the DocScan app due to the demand by scholars and genealogists as well as the positive feedback of project partners. An open source Android app for the scanning of documents has been developed which can now be connected to Transkribus and Dropbox. The main innovative feature is the auto-shoot mode which detects if a page is turned over and automatically takes a picture of the new page. Although new features will/can be integrated into the app, a final version is available in the Google PlayStore. In addition to the app, a mobile scanning device - the ScanTent - has been developed. Based on the 45 created prototypes and the feedback of the scholars and project partners, the ScanTent v3.0 is now ready to go into production.

The development of the Crowd-Scanning app was originally foreseen as a sub-contract for the Russian company ABBYY. This plan was changed due to the fact that CVL was able to reuse recently developed technology. The app is provided as open source on GitHub. The task of improving the functionality of DocScan and the ScanTent has been continued under Task 8.3.

Section 2 describes the final prototype of the ScanTent developed between 2016 and 2018, which is now ready for production. DocScan is presented in Section 3. Future work is presented in Section 5.

2 ScanTent

Based on the experiments in 2016 and 2017 the, for the time being, final prototype was developed in 2018.

The ScanTent from the end of 2017 is shown in Figure 1. Due to the expensive production using tarpaulin, the canvas was replaced by a fabric which is transparent. This allows the use of traditional sewing machines for production and enhances the lighting condition due to the transparent fabric which acts as a diffuser (see Figure 2).

Felt is used as a base for the tent to provide a homogeneous background. LED strips sewed into the tent are used as an additional lighting system. Figure 2 shows an image of the final prototype. The technical description and design goals of the ScanTent are presented in D8.2 (resolution, camera distance, Depth-of-Field (DoF), etc.). A professional sewing pattern was manufactured and also the final measurements and production of the mount. Figure 3 shows the design drawings of the sewing pattern and the mount.

We would like to thank all participating organizations and individuals for the valuable feedback, especially

- University Archive Greifswald
- Niederösterreichisches Landesarchiv
- Staatsarchiv des Kantons Zürich
- Goethe-Uni Frankfurt

Figure 1: The ScanTent in the library.

- BBAW Berlin
- UIBK
- scholars (historians)
- UCL
- NAF

In November 2018 the Vienna Scanathon was organized and hosted by the Austrian Centre for Digital Humanities (ACDH). This was the successor of the first international Scanathon held in London, Helsinki and Zurich. Scanathons were designed to promote the tools and gather feedback from potential users. In the first International Scanathon three parallel Scanathons were held at The National Archives of the UK, the National Archives of Finland and the State Archives of Zurich. The events were led by Louise Seaward (University College London), Maria Kallio (National Archives of Finland) and Tobias Hodel (State Archives of Zurich). Presentations and digitisation went on in each location and the three archives also shared their progress on a Skype call. The ScanTent was also presented at the *Veneto Research Night* at the Piazza Dante in Verona. There was also an article in the Verone newspaper (30.09.2018), see Figure 4.

Based on the success of the Scanathons and the requests from institutions/individual researchers about 300 ScanTents will be finalized by the beginning of 2019. Thus, ScanTents will be available for sale and renting.

3 DocScan

DocScan is a document scanner app, which has a live view and detects in real time the document page. Furthermore, it evaluates if a picture is in focus to give the user feedback

Figure 2: Final prototype of the ScanTent.

and to assure a certain picture quality which will be used for further processing with document image analysis methods. Additionally, it detects page turns and automatically take new pictures. Thus, a book can be scanned very quickly by flipping through the pages. The DocScan app is directly accommodated to the ScanTent.

New features of the DocScan app developed in 2018:

- Transkribus batch upload
- Dropbox upload
- Crop Images (also as batch)
- QR Code Reader improvements
- performance improvements

Figure 5 shows screenshots of the app for the batch upload and the image gallery/crop function.

The QR Code generator/reader allows libraries/archives the following workflow:

- A library/archive creates automatically a QR code with the main metadata of a record (signature, record identifier, Transkribus identifier), which is displayed on the institution's website.
- A user will scan the relevant QR code with the DocScan app and then begin digitizing the document.

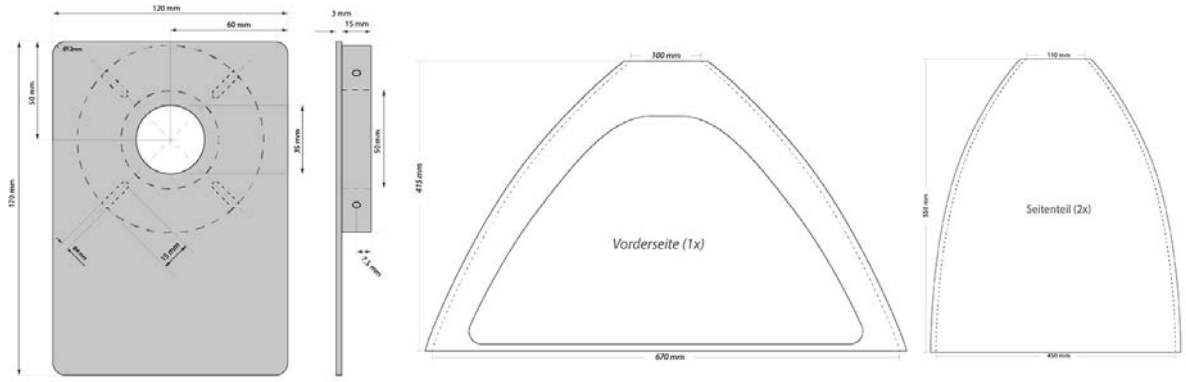


Figure 3: Mount design and sewing pattern of the ScanTent.


La curiosità

Il set fotografico portatile che consente l'analisi del testo dei libri

La ScanTent, un sistema di digitalizzazione innovativo

Figure 4: Article of ScanTent in the Verone newspaper.

- A copy of each digitised image and its appropriate metadata will then become directly available to the library or archive, from where it can be connected with their digital repository.

The University Archive Greifswald has already experimented with this workflow which can be seen in a short video[1].

4 Resources

The Transkribus DocScan App is OpenSource and available at the Google Playstore, as well as in the Transkribus Github repository: `https://github.com/TUWien/DocScan`.

All information regarding the ScanTent are presented at `https://scantent.cvl.tuwien.ac.at/en/`.

[1] `https://youtu.be/BVnKnsWUHOM`

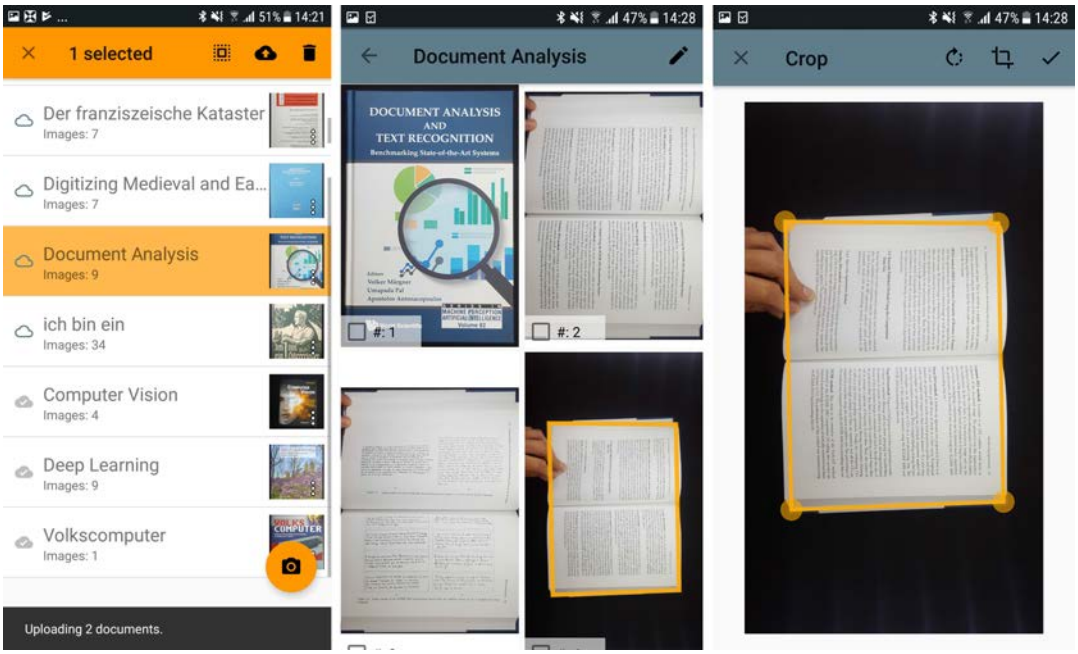


Figure 5: Screenshot of the DocScan app showing batch upload, image gallery/crop function.

5 Future Work

In 2018 the material selection for the ScanTent and the workflow for the production has been defined. In 2019 the marketing of the ScanTent is planned. ScanTents will also be available for rent for libraries and archives interested in organizing "Scanathons". The events will work as follows:

- A library or archive contacts potential volunteers among their users and rents (or buys) ScanTents
- A specific (historical) collection is selected for scanning, e.g. documents relating to a topic
- Volunteers come with their own smartphones and receive instructions from the library or archive staff about using DocScan and the ScanTents
- A scanning session takes place
- The Scanathon could be repeated the next day or even become a weekly activity for volunteers

Within 2 hours, each volunteer will be able to scan up to 1000 images (or 2000 book pages). This means that 20 volunteers will produce up to 20.000 images or 40.000 book pages within one session. The speed may be slower with more delicate archival material but volunteers will still be able to generate tens-of-thousands of images of good/sufficient quality.

Additionally, DocScan will also be ported to iOS in the future.

READ

RECOGNITION & ENRICHMENT OF ARCHIVAL DOCUMENTS

D8.3 (Annexe A)

Open Innovation Forum

Handwritten Music Retrieval and Indexing

J. Calvo-Zaragoza, A.H. Toselli and E. Vidal
UPVLC

Distribution: http://read.transkribus.eu/

READ
H2020 Project 674943

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 674943

Project ref no.	H2020 674943
Project acronym	READ
Project full title	Recognition and Enrichment of Archival Documents
Instrument	H2020-EINFRA-2015-1
Thematic priority	EINFRA-9-2015 - e-Infrastructures for virtual research environments (VRE)
Start date/duration	01 January 2016 / 42 Months

Distribution	Public
Contract. date of delivery	21.11.2018
Actual date of delivery	31.12.2018
Date of last update	31.12.2016
Deliverable number	D8.3 (Annexe A)
Deliverable title	Open Innovation Forum
Type	report
Status & version	in process
Contributing WP(s)	WP6
Responsible beneficiary	ABP
Other contributors	UPVLC
Internal reviewers	Günter Mühlberger
Author(s)	J. Calvo-Zaragoza, A.H. Toselli and E. Vidal
EC project officer	unknown
Keywords	probabislitic indexing, information extraction, handwritten music scores

1 Introducction

Huge amounts of manuscripts containing handwritten music notation pile up in thousands of libraries and archives, public and private alike. Following cultural heritage conservation and dissemination efforts, large quantities of these manuscripts have been or are being digitized in many countries. However, for these sources to be really useful, not only the digital images but also their very *musical notation contents*, need to be made accessible.

Current automatic transcription technologies –such as Optical Music Recognition (OMR) or Handwritten Music Recognition (HMR)– are far from offering results that are sufficiently accurate [7, 2]. Therefore, if a certain degree of accurateness is required, significant user effort is needed to supervise the transcription process. Although tools for automatic transcription may obviously help relieving such work-load, this scenario is unfeasible when approaching large-scale manuscript collections.

Quite often, however, the interest is not so much in having an exact transcript of the musical content, but in being able to perform a content-based search with a certain reliability.

Traditional attempts towards this target have more or less tried to adopt classical concepts and techniques of Optical Character Recognition (OCR). Most of the proposed methodologies rely on knowledge-based approaches to: a) segment the music sheet images into individual music-symbol regions, and b) try to recognize each of these image regions, without taking into account the (musical) context in which it appears in the image. The noisy symbols and symbol sequences recognized in this way, are then indexed using classical techniques available for digital symbol sequences such as plain text, DNA sequences, etc.

However, this idea is flawed in two main respects. First, as it has painfully conceded in recent years in the field of handwritting text recognition, here the automatic segmentation of music images into individual symbols also proves extremely difficult and/or unreliable. And second, music manuscripts are very heterogeneous and knowledge-based, heuristic methods tend to fail to generalize well over different styles. Therefore, new systems need to be painstakingly built manually, almost from scratch, for every new manuscript collection.

In contrast with these segmentation-based and knowledge-based heuristics, we have recently proposed a holistic, segmentation-free and principled, machine learning framework, which is providing promissing results [4, 5].

Here we adopt these ideas for the task of *content-based indexing and search* for *untranscribed* music sheet images. Following successful UPVLC developemts in the tasks of Key-Word Spotting (KWS) and Text Indexing on handwritten text documents [8, 1, 6][1], pioneering efforts have been made in the READ project towards the development of innovative technologies for *probabilistic* Music-Symbol Spotting (MSS) and Music Symbol Sequence Indexing (MSSI). Folloing [3], MSS is here understood as the task of finding the likley locations of a given musical symbol in a set of music sheet images. In addition, MSSI is assumed to consist in indexing MSS results so as to enable fast search for *melodic patterns*, represented as *music symbole sequences*, on (large) collections of music manuscripts.

[1] See also the section "Query by String (QbS) KWS work at UPVLC" of the READ deliverables D7.15 and the annex on Large-scale Probabilistic Word Indexing of D8.12

2 A Case Study: The VORAU-253 Manuscript

In order to explore the ideas outlined in Section 1 we consider the task of retrieval of music symbol sequences, representing melodic patterns, from early music manuscripts. As a test-bed collection we used the music manuscript referred to as Cod. 253 of the Vorau Abbey library, which was provided by the Austrian Academy of Sciences. It is written in German gothic notation and dated around year 1450. Fig. 1 shows example images of the VORAU-253 manuscript. It consists of about 450 page images which contain close to one thousand 4-line *staves*, and only about one hundred 5-line staves. The notation is based on 19 different music symbols, as discussed in the coming sections. More details of the dataset derived from this manuscript are reported in READ Deliverable D7.15.

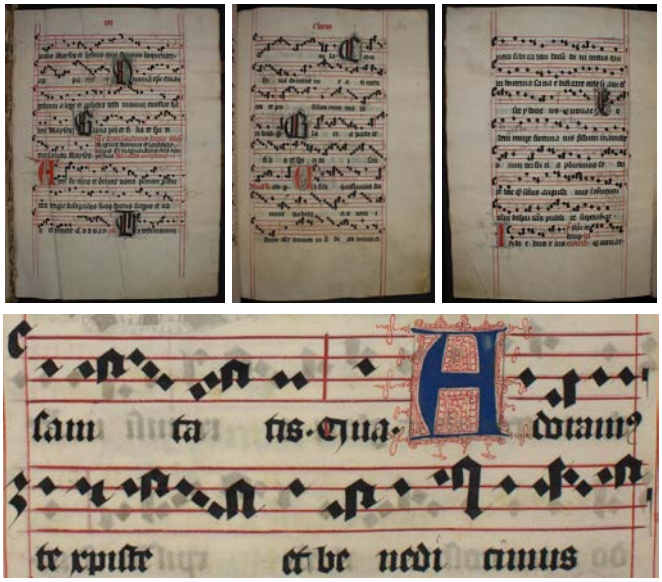


Figure 1: Examples of page images of VORAU-253 music manuscript.

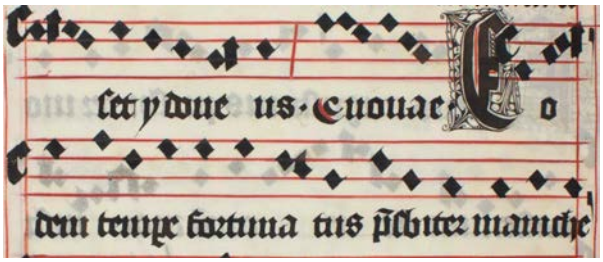
3 German Gothic Notation and Volpiano Ground Truth

In *modern music notation*, two main naming schemes are generally adopted to textually specify notes (pitch levels) of the *diatonic scale*.

- Romance and major Slavic languages: Do, Re, Mi, Fa. Sol, La, Si
- English: C, D, E, F, G, A, B

In German Gothic notation, notes are represented as squares drawn at different vertical positions, generally in a four-line *"stave"*. The actual meaning (pitch) of this notation depends, not only of its vertical position, but also of the *clef* of the stave. Two main clefs are used in VORAU-253: *C=(Do)* and *F(=Fa)*. See examples in Fig. 1 and 2.

The original Ground Truth used in the VORAU-253 manuscript is annotated with "semantic" meaning (i.e., with the intended pitch of each note, rather than its graphical position in the stave image). The so called *Volpiano* encoding was adopted, where each note is represented as a single letter, without time or rhythm information. However, hyphens are used to synchronize note sequences with syllables of the chant lyrics, as illustrated in Fig. 3.



C4 Si Do Si La Si La Sol Fa . . .
(C4 B C B A B A G F . . .)

C3 Re Mi Do Re Re Re Do . . .
(C3 D E C D D D C . . .)

Figure 2: German Gothic notation

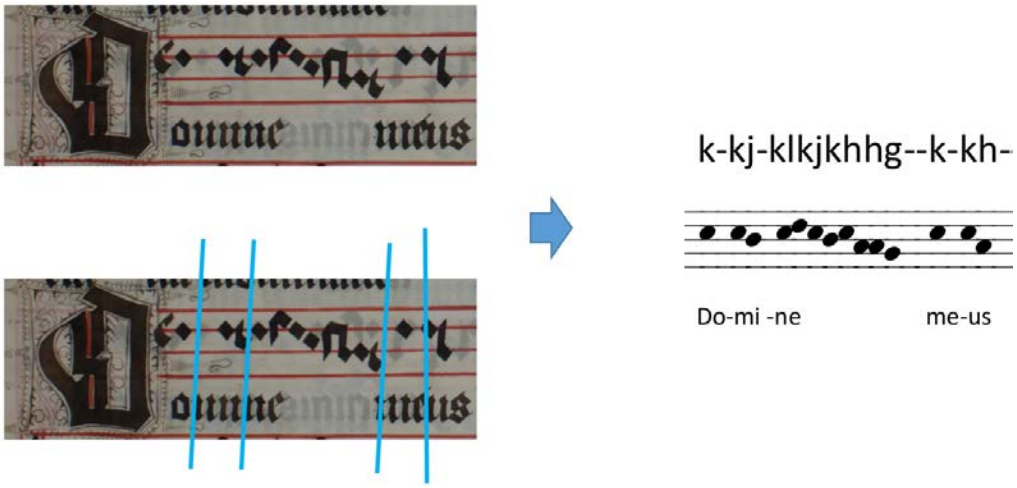


Figure 3: Volpiano encoding.

4 Adapting GT Annotation for Optical Modeling

To allow for adequate optical model training, the Volpiano encoding was converted into a graphical annotation, where vertical positions of notes are made explicit. It should be noted that vertical position can be inferred from pitch and clef. Conversely, pitch can be deduced from vertical position and clef. Therefore, Volpiano and vertical-position encodings are reversible if the clef is known. Fig. 4 illustrates the proposed encoding transformations. Finally, in this initial work, lyrics syllables and the associated hyphens are ignored.

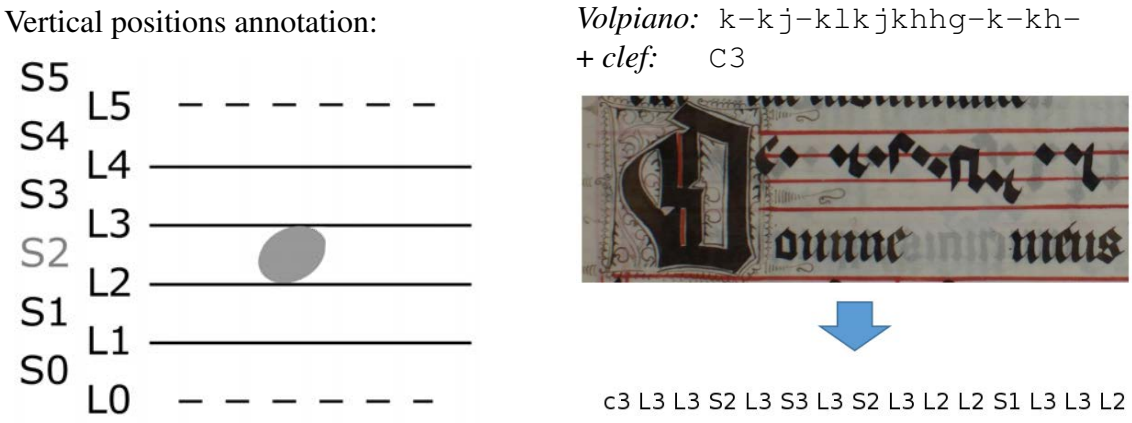


Figure 4: Adapting Volpiano encoding to the needs of optical modeling.

5 Optical and Language Modeling

Two kinds of models were trainied from annotated images. First a *Convolutional-Recurrent Neural Network* (CRNN) *optical music symbol model*, which was trained using the TensorFlow toolkit. This kind of optical model is reminiscent of the successful models used nowadays in handwritting text recognition (HTR). However, significant research effort had to be devoted to find an adequate architecture for the convolutional part of the CRNN, which allows the network to learn the specific features of music symbols. More specifically, while character vertical position invariance is important for optical models used in HTR, the vertical position of musical symbols constitute perhaps the most important feature to distinguish among the different musical symbols. Similarly, certain amount of scale invariance is beneficial for HTR, but not so much for music notation optical modeling. Therefore, conventional CRNN architectures which provide excellent optical models for HTR, tend to fail dramatically in the case of music notation. The details of the CNRR architectures which finally were succesful for music notation optical modeling in our experiments exceed the scope of the present report and will be published in a forthcoming technical paper.

On the other hand, much in the same way as language models are used in HTR to model the concatenation regularities of character and/or words, music notation also exhibit similar regularities and similar language models can therefore be used. In our experiments and demonstrator we used a *2-gram symbol language model*, estimated from training sequences of tokens representing vertical symbol positions within the staves.

6 Laboratory Experiments

Empirical assessment was carried out under laboratory conditions. First, the dataset was divided into 422 annotated page images for training the optical and language models and 44 annotated pages for evaluation. On these images, automatic stave segmentation was carried out using Transkribus tools (which did not always provide accurate resuts in this dataset). Then objective query sets were established as follows: all the 15 symbols seen in the test set were used as *single symbol queries*, and all the 615 sequences with lengths ranging from 3 to 15 which appear in the test set more than once were selected for *symbol sequence queries*. Finally Precision-Recall performance was evaluated at stave level for sequence queries and at relative symbol position level for single-symbol queries.

The search performance achieved under these conditions, measured in terms of *Average Precision* (AP), was: 89% for single-symbol queries and 86% for symbol sequences. These are very good results which predict an excellent degree of search and retrieval usability in practice, as can be experienced first hand using the on-line demonstrator described in Section 7.

More details of these experiments and results are reported in READ Deliverable D7.15.

7 Public WEB Indexing and Search Demonstrator

Using the models trained as described in Sec. 6, the whole VORAU-253 manuscript was probabilisticaly indexed. Information of the resulting probabilistic index is summarized in Table 1.

Table 1: Relevant data of the probabilistic index obtained for the VORAU-253 manuscript.

Computed values		Estimated from index probabilities	
Number of indexed pages	490	Running symbols	88 751
Number of spots	262 660	Avg. num. of runn. symbols / Page	181
Average num. of spots / page	536	Avg. num. of spots / runn. symbol	3.0

Fig. 5 shows the front page of the UPVLC (PRHLT) music symbol sequence search and retrieval interface for the probabilistic index of the VORAU-253 manuscript, which is available at: `http://prhlt-carabela.prhlt.upv.es/music`. The interface provides a link for help about essential indexing and search concepts and details about search options, along with symbol sequence query examples. Some query examples are also listed below Fig. 5.

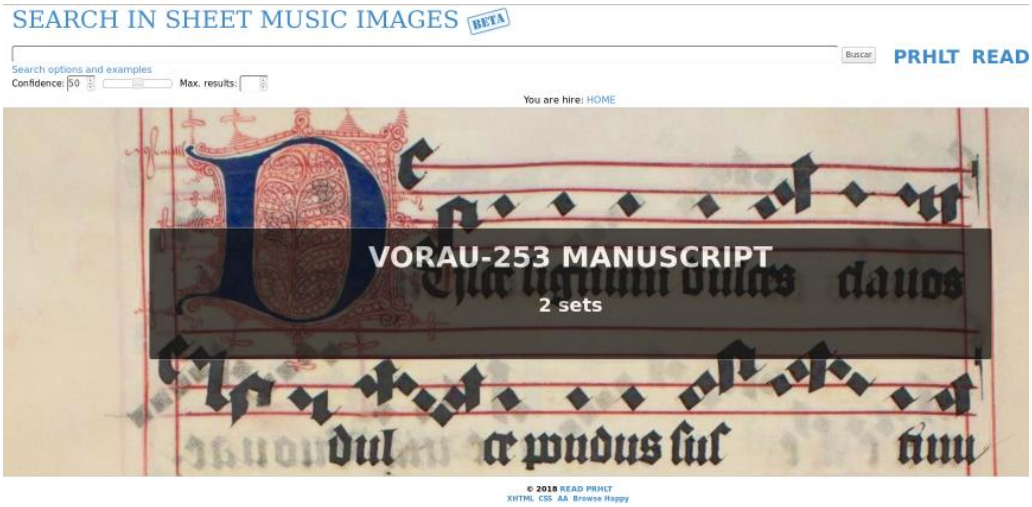


Figure 5: PRHLT (UPVLC) Music Indexing and Search Live Demonstrator

EXAMPLES OF MUSIC SYMBOL AND MUSIC SYMBOL SEQUENCE QUERIES:

Single symbols (obviously not very useful);
S3 (Si / B in C4 clef, or Re / D in C3 clef)

F4 (Fa / F clef in the fourth line)

Symbol sequences:
[S1 S3 S2 S3 L4 S3]
(Sol Re Si Re Mi Re / G D B D E D – in C3 clef)

[L3 L3 L3 L3 L3]
(Do Do Do Do Do / C C C C C – in C3 clef)

[L2 L2 S2 L3 S2 L3 S2 S1 L2 S1]
(Fa Fa Sol La Sol La Sol Mi Fa Mi / F F G A G A G E F E – in C4 clef, or
Do Do Re Mi Re Mi Re Si Do Si / C C D E D E D B C B – in C2 clef)

Sequences with alteration:
[L2 FLAT S2 L2]
(La Si*b* La / A B*b* A – in C3 clef, or Re Mi*b* Re / D E*b* D – in F3 clef)

[FLAT S3 L3 S2 L3]
(Si*b* La Sol / B*b* A G – in C4 clef, or Re*b* Do Si / D*b* C B – in C3 clef)

8 Discussion and Conclusions

With a fair amount of ingenuity, handwritten text images indexing and search technologies can be adapted to deal with handwritten musical documents. Initial developments and results look promissing, but many specificities still need substantial fundamental and engineering research. In particular, how to properly annotate *ground truth* music transcripts requires in-deep thinking, discussion and experimentation. Similarly, adequate, user-friendly ways to express music-meaningful queries need to be envisaged and devised. Moreover, in this type of music manuscripts, both music (staves) and chant (lyrics) are needed together to adequately predict timing and rhythm.

As an interesting byproduct of this work, we observe that Probabilistic Indices produced for music search and retrieval might be advantageously used to *automatically discover music patterns* which appear frequently in untranscribed images. Research on this issue is also left for future works.

References

[1] T. Bluche, S. Hamel, C. Kermorvant, J. Puigcerver, D. Stutzmann, A. H. Toselli, and E. Vidal. Preparatory kws experiments for large-scale indexing of a vast medieval manuscript collection in the himanis project. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 311–316, Nov 2017.

[2] Donald Byrd and Jakob Grue Simonsen. Towards a standard testbed for optical music recognition: Definitions, metrics, and page images. *Journal of New Music Research*, 44(3):169–195, 2015.

[3] Jorge Calvo-Zaragoza, Alejandro H Toselli, and Enrique Vidal. Probabilistic music-symbol spotting in handwritten scores. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 558–563. IEEE, 2018.

[4] Jorge Calvo-Zaragoza, Alejandro Héctor Toselli, and Enrique Vidal. Handwritten music recognition for mensural notation: Formulation, data and baseline results. In *14th IAPR International Conference on Document Analysis and Recognition, Kyoto, Japan*, pages 1081–1086, 2017.

[5] Jorge Calvo-Zaragoza, Alejandro Héctor Toselli, and Enrique Vidal. Hybrid Hidden Markov Models and Artificial Neural Networks for handwritten music recognition in mensural notation. *To be published*, 2019.

[6] Ernesto Noya-García, Alejandro H. Toselli, and Enrique Vidal. Simple and effective multi-word query spotting in handwritten text images. In Luís A. Alexandre, José Salvador Sánchez, and João M. F. Rodrigues, editors, *Pattern Recognition and Image Analysis: 8th Iberian Conference, IbPRIA 2017, Faro, Portugal, June 20-23, 2017, Proceedings*, pages 76–84. Springer International Publishing, Cham, 2017.

[7] Ana Rebelo, Ichiro Fujinaga, Filipe Paszkiewicz, André R. S. Marçal, Carlos Guedes, and Jaime S. Cardoso. Optical music recognition: state-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, 1(3):173–190, 2012.

[8] Alejandro H. Toselli, Enrique Vidal, Verónica Romero, and Volkmar Frinken. Hmm word graph based keyword spotting in handwritten document images. *Information Sciences*, 370(C):497–518, November 2016.