Optical Music Recognition

Andrew Hankinson

Bodleian Libraries, University of Oxford

SIMSSA | Single Interface for Music Score Searching and Analysis

Download this file

https://goo.gl/NHxtwd

Overview

Part I

(Pre-) History of OMR

The Great Transition (or not…?)

What's Next?

Part II

Techniques

Challenges

What's Next?

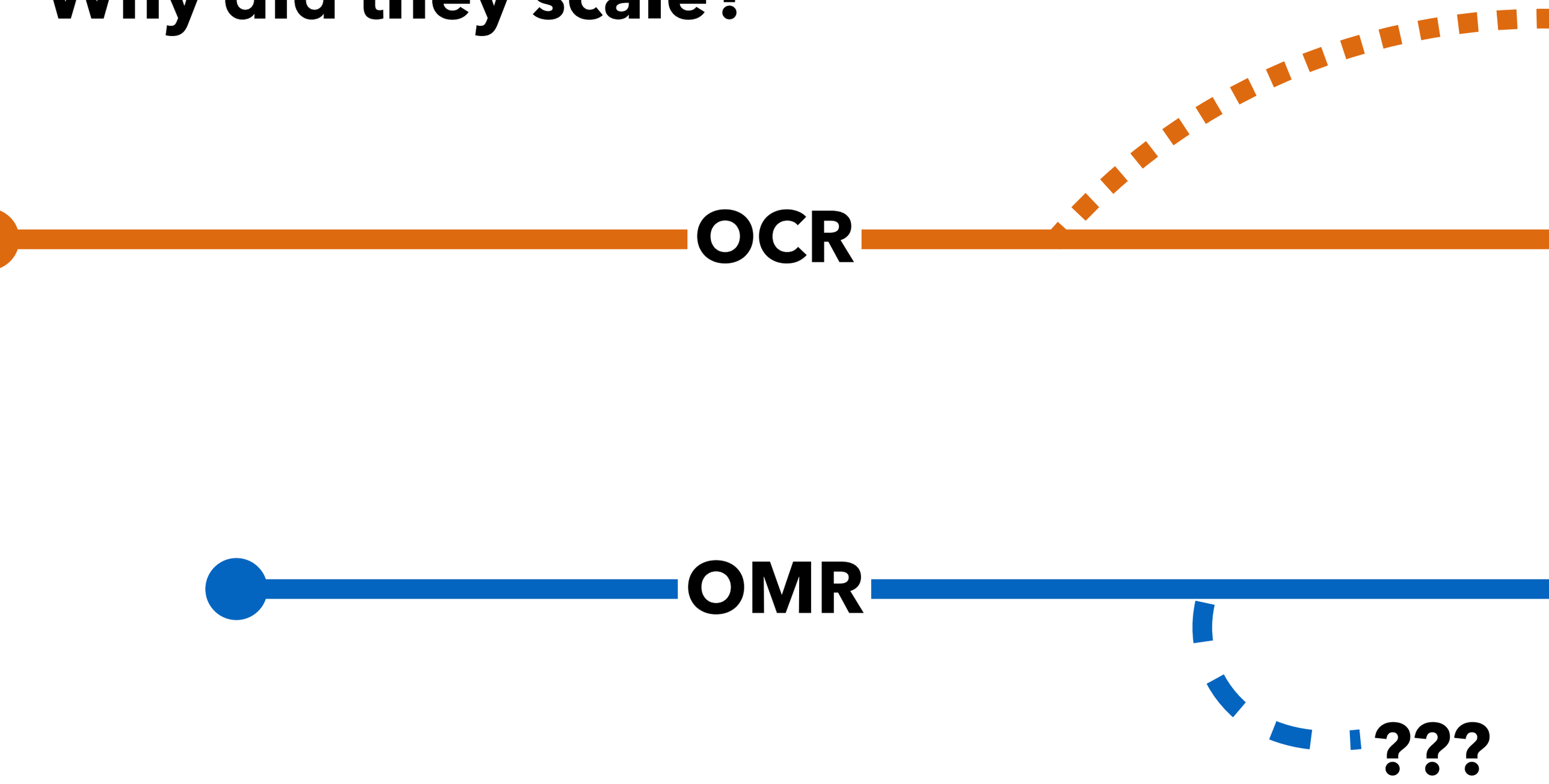

Notation-to-Sound
Playback


Notation Editors
(Sibelius, MuseScore, etc.)

SALE

When did they scale?

Why did they scale?



Hypothesis

Image + Text Alignment is the secret to scaling recognition systems.

Optical Character Recognition

Transcribes text from

physical objects

into

machine-readable data.

Optical Music Recognition

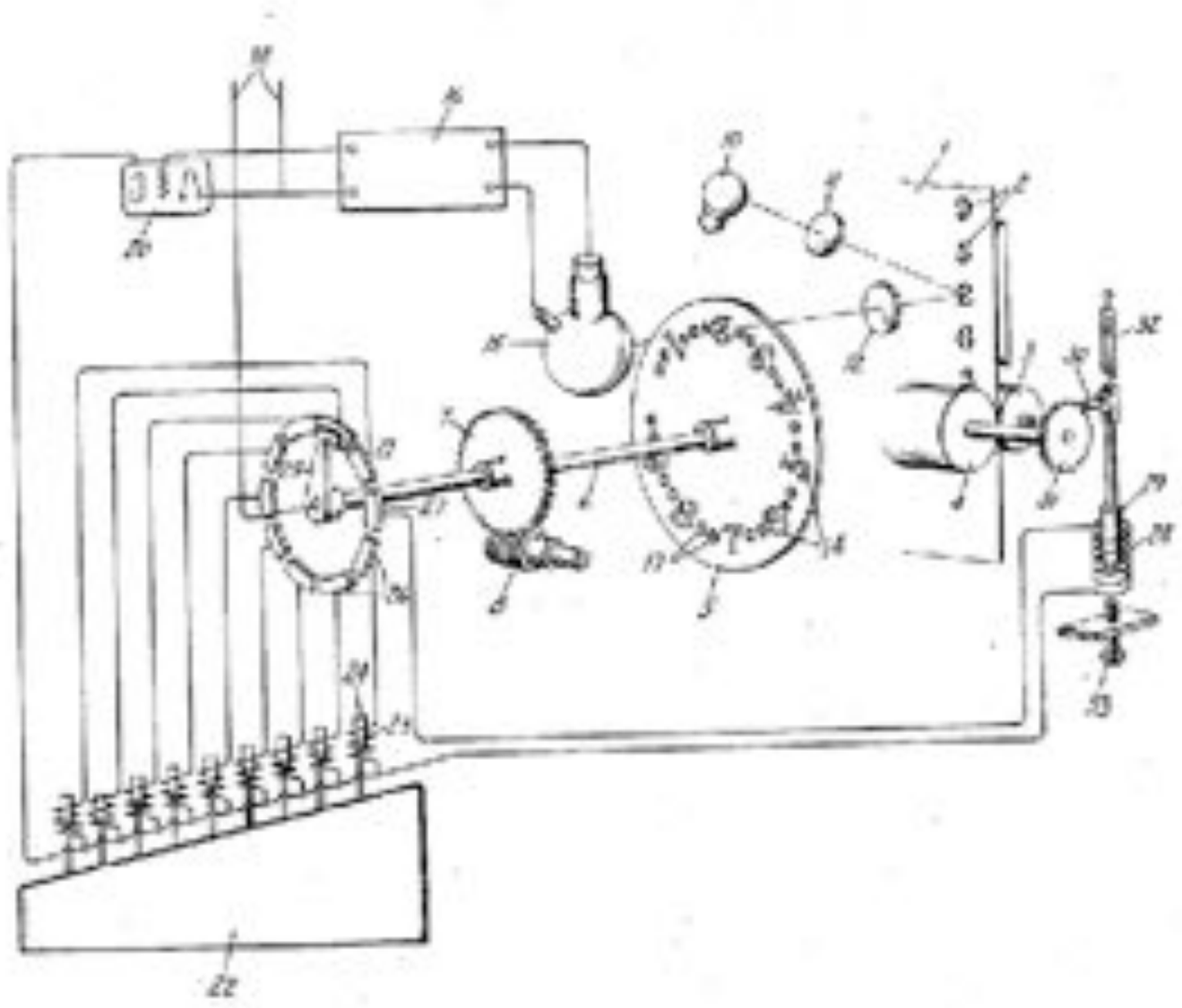
Transcribes symbolic music from

physical objects

into

machine-readable data.

Pre-History: OCR



Early electro-mechanical OCR system (Handel 1931)

"At the beginning stage it was thought that it would be easy to develop an OCR, and the introduction of a very rigorous reading machine was expected in the 1950's. Roughly speaking, the 1950's and 1960's... were periods when researchers imagined an ideal OCR, even though they were aware of the great difficulty of the problem. Actually this is an instance of a common phenomenon which occurred in the research field of artificial intelligence in general."

Mori et al. 1992, 1033.

OCR: The Highlights

1950s: Reader’s Digest, AT&T (billing systems)

First commercially-available computerized OCR by the Intelligent Machines Research Corporation

https://www.flickr.com/photos/smithsonian/2550431145

OCR: The Highlights

1970s: Defense Advanced Research Projects Agency (DARPA)

Military procurement process and document management.

"The future use of Optical Character Recognition (OCR) during the next decade will produce great cost saving for certain areas. The use of multi-font machines is just coming into being, and will be perfected within the next few years."

Varley 1969, 43.

OCR: The Highlights

1970s: Kurzweil Data Entry Machine (KDEM)

Lexis (Law Indexing) / Nexis (News Indexing)

Purchased by the University of Oxford

“Up to 20 pages of A4 per hour”

Hockey, S. 1986

http://www.achievement.org/achiever/ray-kurzwell/

OMR: Origins

1966: Dennis Pruslin

Automatic recognition of Sheet Music

1970: Daniel Prerau

Computer pattern recognition of standard engraved music notation

1972: Michael Kassler

Optical Character-Recognition of Printed Music: A Review of Two Dissertations Automatic Recognition of Sheet Music…

Kassler 1972

"Both authors have essayed to solve less than the entire problem, so the work of each should be judged by its extensibility to, rather than by its non-realization of, an actual working machine."

Kassler 1972

Pruslin: “Clefs, time-signatures, grace-notes, dynamic marks, phrase marks, and other special signs of CCMN are disallowed.”

Kassler 1972

"Prerau limits attention to a subsystem of CCMN in which each of only two parallel staffs bears monolynear music composed of notes, rests, treble and bass clefs, certain time-signatures, [sharp/flat/natural], and dots of prolongation—but not tempo indications, dynamic or phrase marks, or certain other 'special signs'."

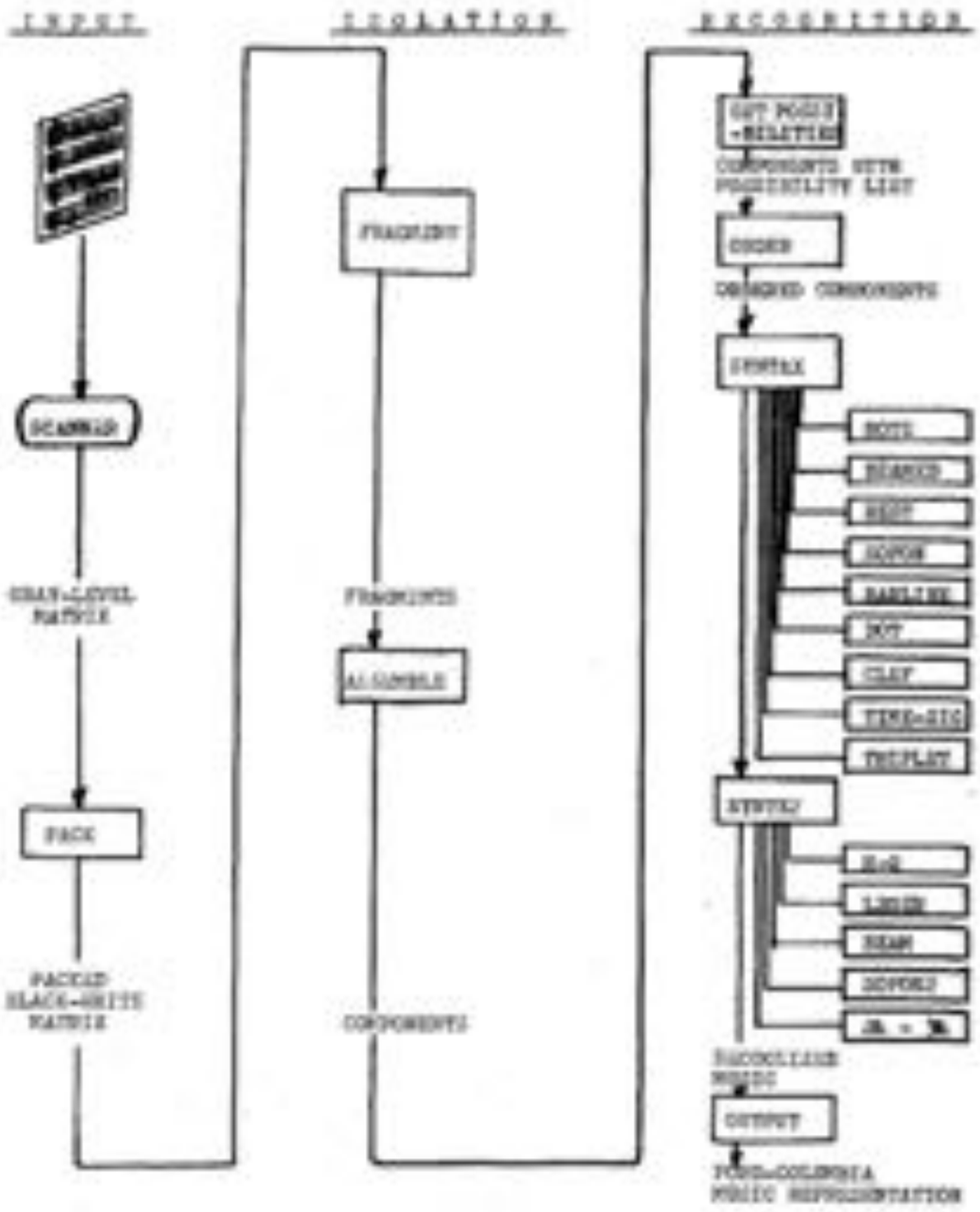


Figure 3—DO-RE-MI flowchart

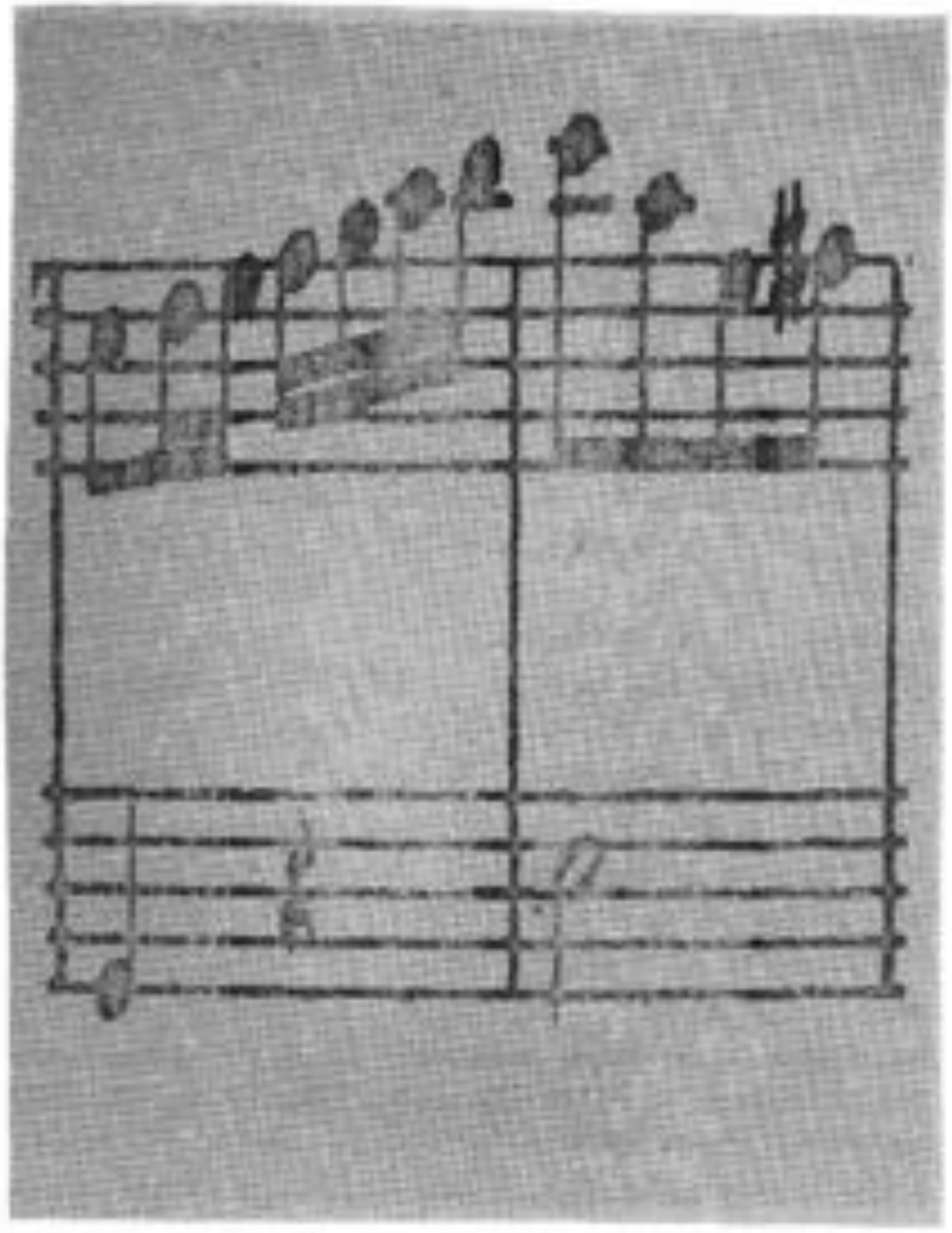


Figure 4—Printout of a picture

The OMR Process in the DO-RE-MI system (Prerau 1971)

Kassler 1972

"Perhaps the greatest accomplishment of the authors is that, as a result of their work, the logic of a machine that 'reads' multiple parallel staffs bearing polylynear printed music in at least one 'fount' and size can be seen to be no further than another couple of M. I. T. dissertations away. Quite possibly such dissertations may get completed before much thought is directed toward deciding what wisely to do with the masses of musical data that an operational OCR system could make available for computer processing."

1980s

OCR/OMR is for perfect transcription.

Computers could not (cheaply) store images ($$$$).

Computers could not (easily) display images ($$$)

Data transfer was 

Text (ASCII) was easy, cheap, and fast.

After 40 years of intense work, OCR still did not work perfectly.

"I see no occasion for that. You and the girls may go, or you may send them by themselves, which perhaps will be still better; for as you are as handsome as any of them, Mr. Bingley might like you the best of the party."

"My dear, you flatter me. I certainly have had my share of beauty, but I do not pretend to be anything extraordinary now. When a woman has five grown-up daughters, she ought to give over thinking of her own beauty."

"In such cases a woman has not often much beauty to think of."

"But, my dear, you must indeed go and see Mr. Bingley when he comes into the neighborhood."

"It is more than I engage for, I assure you."

"But consider your daughters. Only think what an establishment it would be for one of them! Sir William and Lady Lucas are determined to go, merely on that account; for in general, you know, they visit no new-comers. Indeed you must go; for it will be impossible for us to visit him, if you do not."

"You are over-scrupulous, surely. I dare say Mr. Bingley will be very glad to see you; and I will send a few lines by you to assure him of my hearty consent to his marrying whichever he chooses of the girls; though I must throw in a good word for my little Lizzy."

"I desire you will do no such thing. Lizzy is

https://archive.org/stream/prideandprejudi06austgoog#page/n14/mode/2up

'^ I see no occasion for that. You and the girls
may go, or you may send them by themselves,
which perhaps will be still better; for as you are
as handsome as any of them, Mr. Bingley might
like you the best of the party."

<'My dear, y ou flatter me. I certainly have
had my share of b ^ftiity , ^'^ ^ ^^ "^^^ prete n d -^ -fee
anything extraordinary now. ^Yhfiii a woman baa

fi ve grown-up daughters, s he j)ug ht tcL^gii.a avfix
thinking of her own beautj^"

^^In such cases a woman has not often much
beauty to think of."

* ^ But, my dear, you must inde ed ^{} and s ^^ ^is..
Bingley when he comers into the neighborhood* '*

^^ it is more than I engage for, I assure you."

https://archive.org/stream/prideandprejudi06austgoog/prideandprejudi06austgoog_djvu.txt

Then...


PHILIPS

"Yet the scholar may prefer to work with the page images because, in general, it is the only digital representation that maintains the full information content of the original, including illustrations, layout, and uncommon markings (as in musical or mathematical notation). It is vitally important for historically significant documents such as hand-written manuscripts, illuminated books, and fine press materials, where the typography materially contributes to the value of the work. Even were the OCR to be perfect, translations into other digital representations inevitably lose information. And it is difficult to predict what needs preserving; at the extreme, it is possible that in the course of time that even the worm holes become significant, as a variant interpolation of the text eaten out may yield a different translation and, perhaps, a new interpretation."

Phelps and Wilensky 1996, 101

There is, and never will be*, a perfect automatic recognition system.

*Caveat

STOR

OCR: The Highlights

1. Create [illegible] ving page layout: Very expensive!

2. Use OCR for text extraction and lose page layout: Cheaper, searchable, but unreadable results.

Image-Text Alignment

JSTOR (ca.1993)

But JSTOR's commitment to be responsive to user needs pushed it to add text files that would be searchable while remaining invisible behind the images. The layer of text could substantially enhance JSTOR's usefulness to scholars and students, who would be able to search the text of the journal for phrases.... With images in place for display, the fulltext's accuracy was of less concern—it could be, at least to some degree, "dirty."

Schonfeld 2003, 28–9

JSTOR (ca.1993)

The pilot librarians worried that ‘even a single user printing a full article composed of bitmapped images may seriously degrade performance’ on the campus Internet gateway—slowing network traffic to a halt for all campus users.

Schonfeld 2003, 30

RightPages (ca.1993)

While we do use [OCR] to obtain the text for searches, the OCR results are never visible to the user, but are spatially associated with the location of the text on each page image... The main reasons for displaying the image and not the ASCII is that most readers are already familiar with general graphical layout conventions, especially those used in journals they have read before, so they can rely on this familiarity when they scan the page images for content. A second practical reason is that OCR and page layout analysis results are not guaranteed to be flawless. Rather than display OCR errors to the users, the problem is sidestepped by showing only the image, and "hiding" the associated OCR text and layout planes.

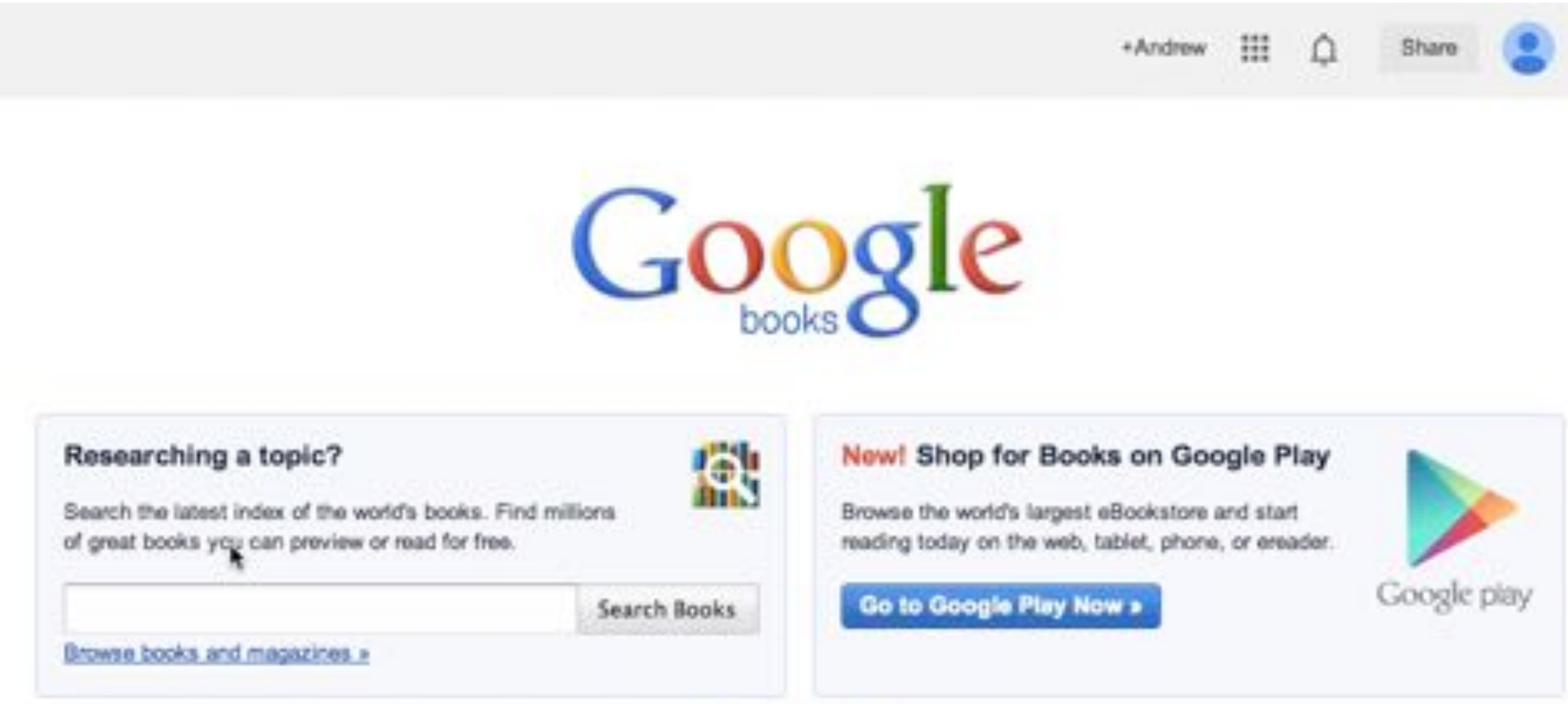
Hoffman et al. 1993, 447.

British Library Newspapers (ca.2001)

"Readability," defined as the user's capacity to view and comprehend historic text, and "searchability," defined as the user's capacity to reach relevant content through provision of search criteria, can be said to be the two components of "accessibility," or the user's capacity to retrieve and read relevant content....

In the past, it was thought that text generated by OCR (Optical Character Recognition) could provide both readability and searchability. Due to the difficulty of extracting high-quality text from historic scans, this approach is now known to be impractical.

Deegan et al. 2001, 6


+Andrew
Share
Google
books
Researching a topic?
Search the latest index of the world's books. Find millions of great books you can preview or read for free.
Search Books
Browse books and magazines »
New! Shop for Books on Google Play
Browse the world's largest eBookstore and start reading today on the web, tablet, phone, or ereader.
Go to Google Play Now »
Google play
My library
About Google Books - Privacy Policy - Terms of Service - Blog - Information for Publishers - Report an issue - Help - Sitemap - Google Home
©2012 Google

Why is this important (at scale)?

- OCR doesn't have to be perfect; humans can read the images
- Preserves non-textual page components *in situ*: Diagrams, tables, etc.
- No dependency on human intervention.

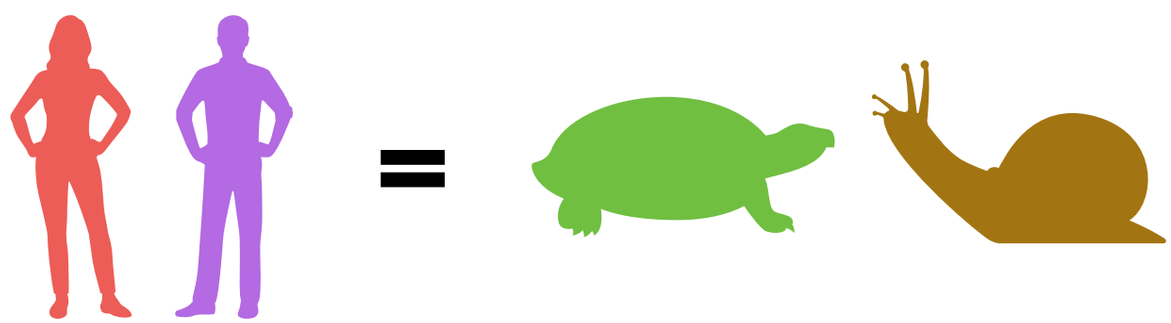
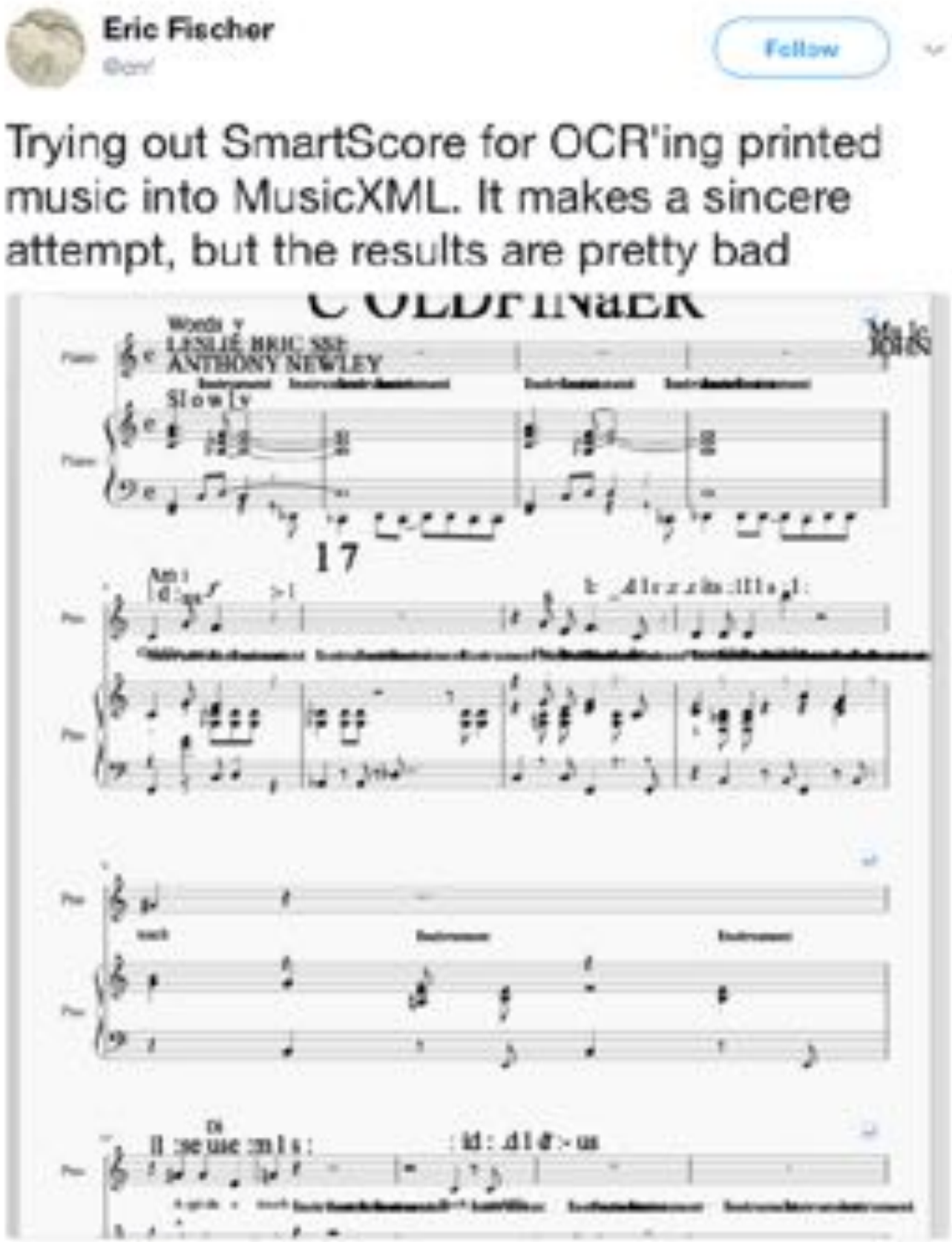


Image + Text Alignment is the secret to scaling recognition systems.

Has OMR made the same transition?

Can we use OMR only and throw away the images?

Not really...



https://twitter.com/enf/status/852399638548054017

Is it good enough if we keep alignment with the images?

Probably!

Demo Time.

International Image Interoperability Framework

http://iiif.io

IIIF: Two Parts

1. Image API

Specified URL format for requesting images

Supports “zooming” image viewers

Images are usable in contexts outside of institutional websites

IIIF: Two Parts

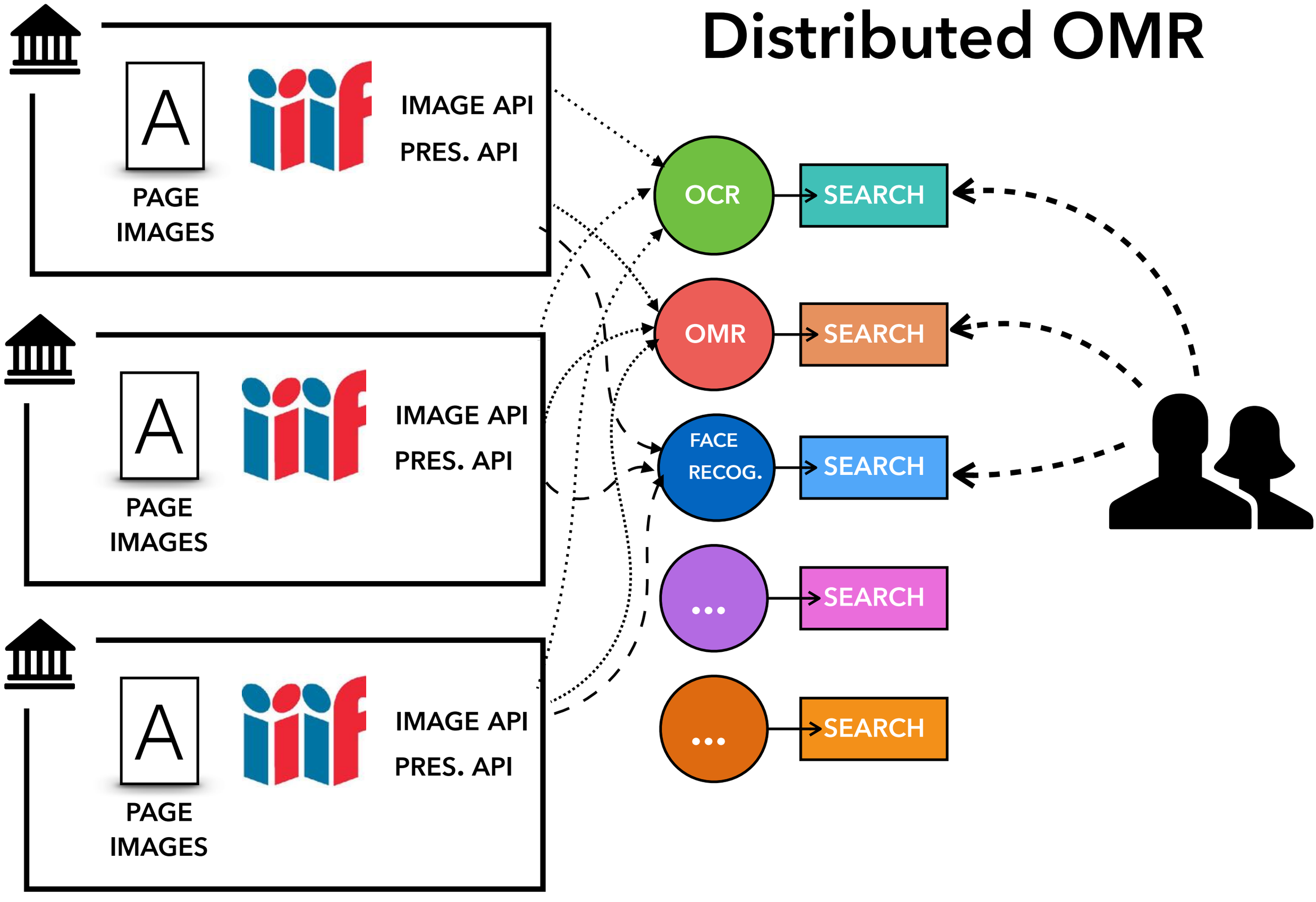
2. Presentation API

Representation of document structure and metadata as a “manifest”

Functions like PDF, but points to images on external servers (Image API)

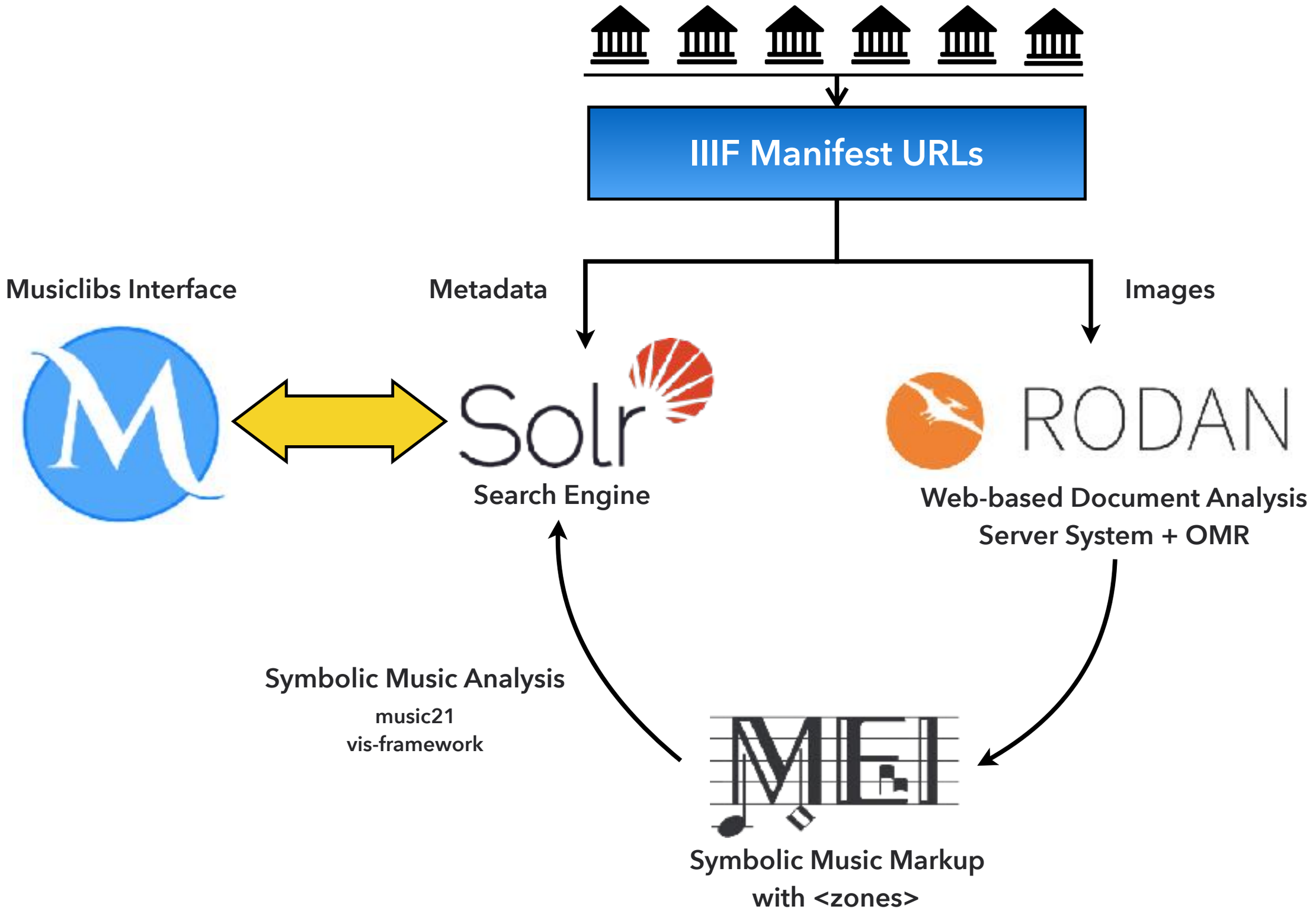
Usable by IIIF-compatible image viewers (Mirador, Universal Viewer, Diva.js)

Distributed OMR



Demo Time.

Musiclibs Recognition & Indexing Process



Part I: Summary

- The perfect is the enemy of the good.
- Image and symbol alignment may be the secret to scale.
- New OMR tools are needed to provide scalability.
- IIIF is awesome.
- What do we do when we have all the music data in the world? “What is a musical query?”

Part II

OMR is a process.

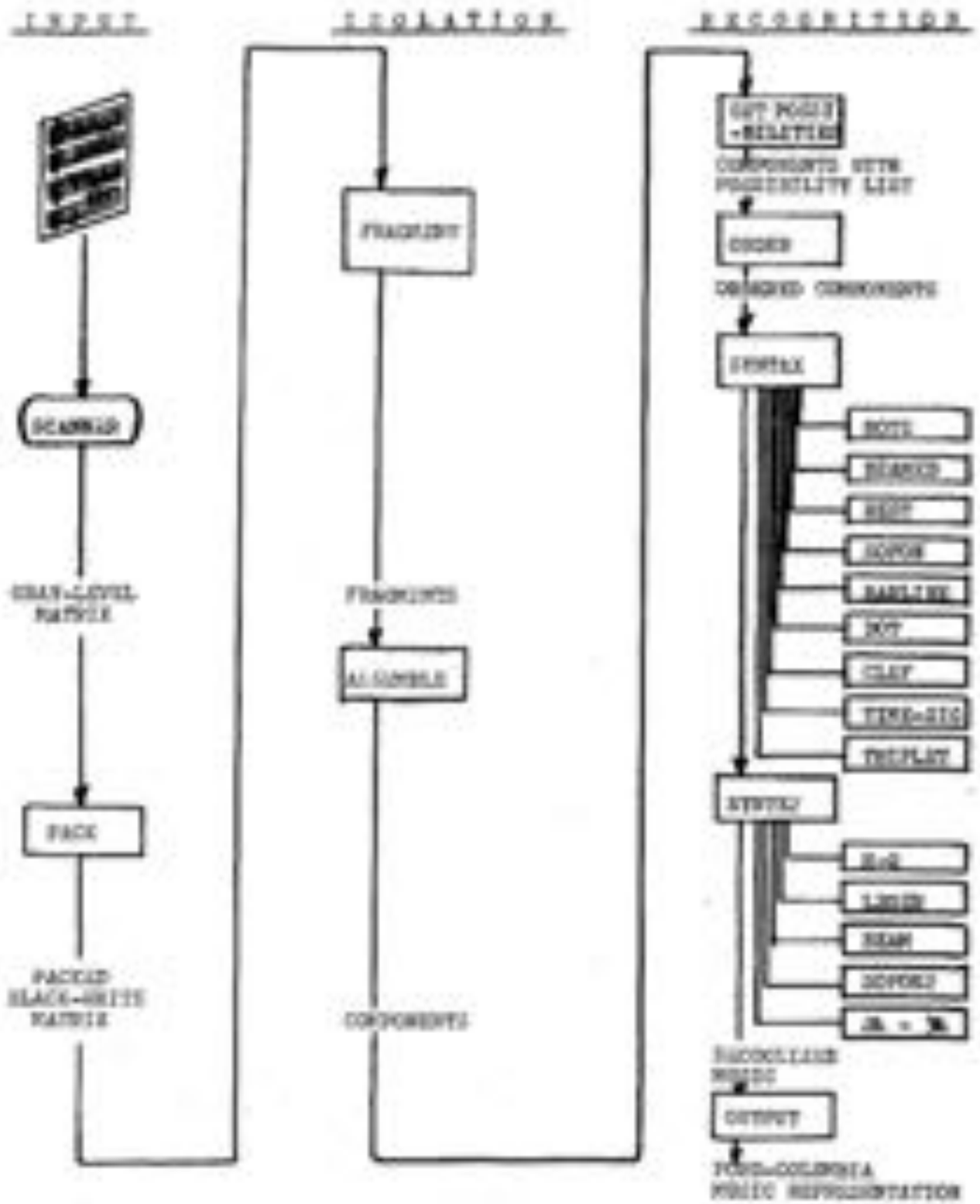
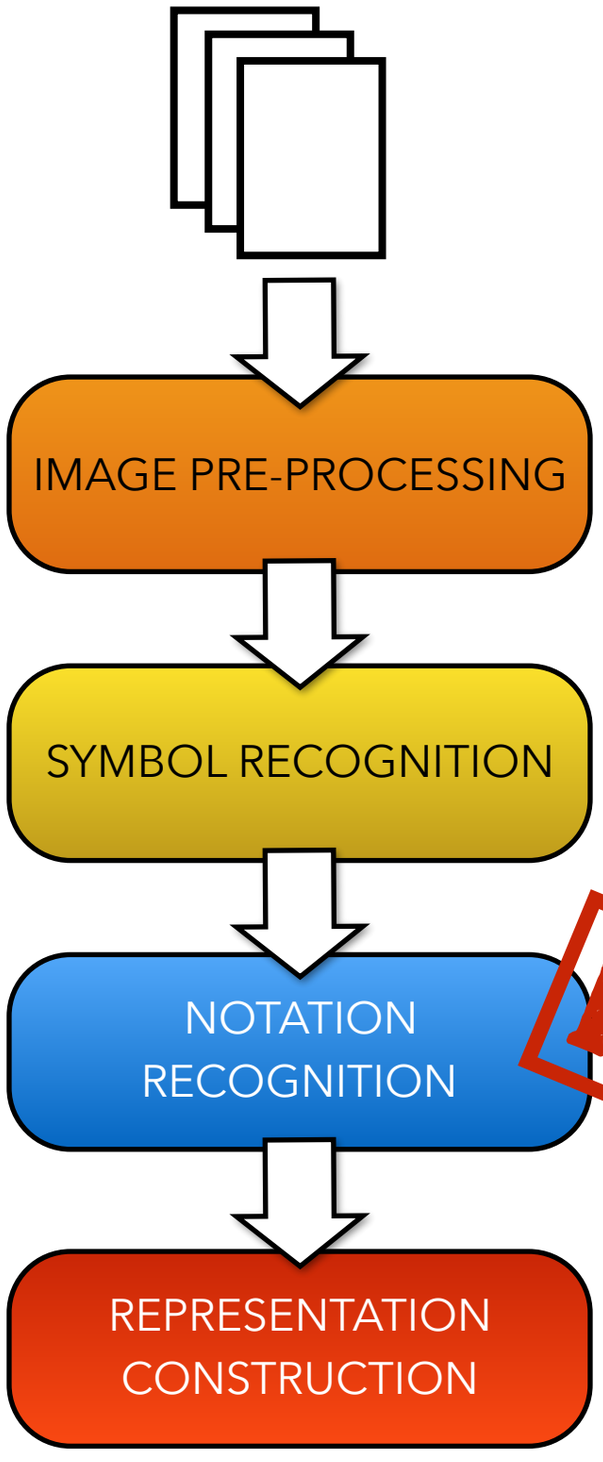


Figure 3—DO-RE-MI flowchart

The OMR Process



BINARIZATION, DE-SKEWING, DE-WARPING PAGE ROTATION, FOREGROUND/BACKGROUND ESTIMATION, STAFFLINE REMOVAL

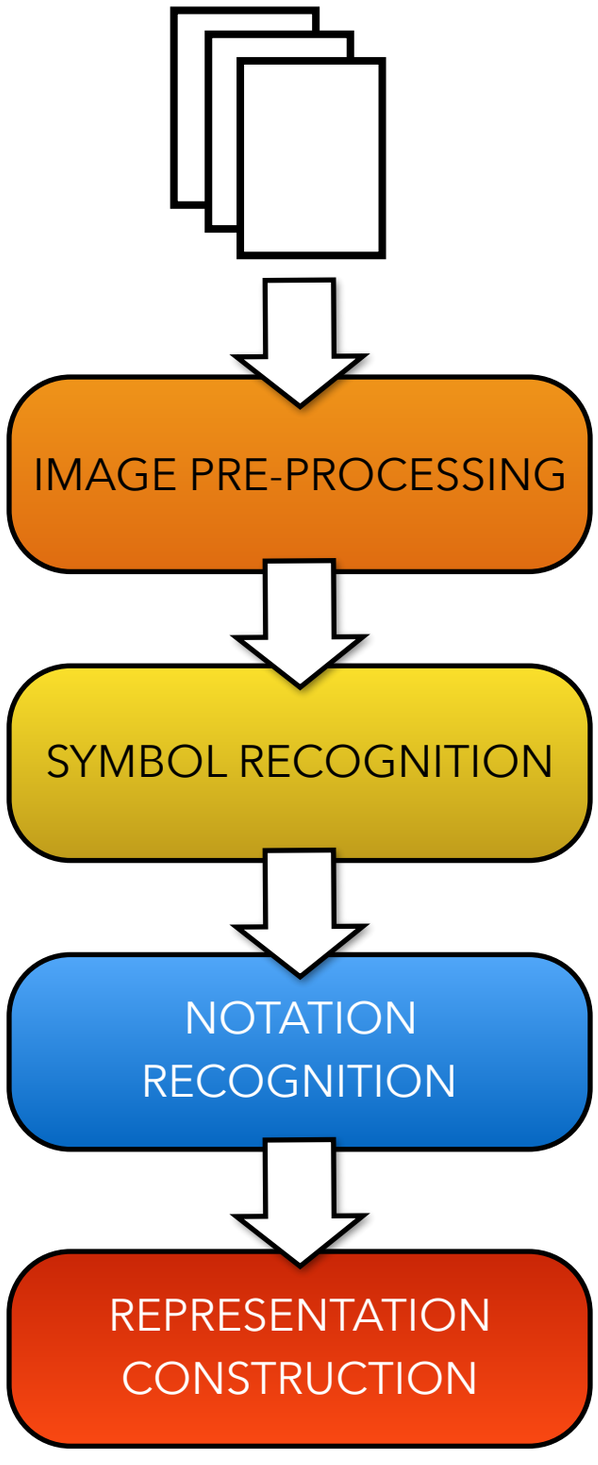
CONNECTED COMPONENT ANALYSIS, FEATURE EXTRACTION, CLASSIFICATION, TEMPLATE MATCHING

INTER-SYMBOL RELATIONSHIPS, NOTE PITCH, DURATION, CHORD VOICING

REALLY HARD

TRANSLATE TO ENCODING FORMAT (MIDI, MUSICXML, MEI)

The OMR Process



BINARIZATION, DE-SKEWING, DE-WARPING PAGE ROTATION, FOREGROUND/BACKGROUND ESTIMATION, **STAFFLINE REMOVAL**

CONNECTED COMPONENT ANALYSIS, FEATURE EXTRACTION, **CLASSIFICATION**, TEMPLATE MATCHING

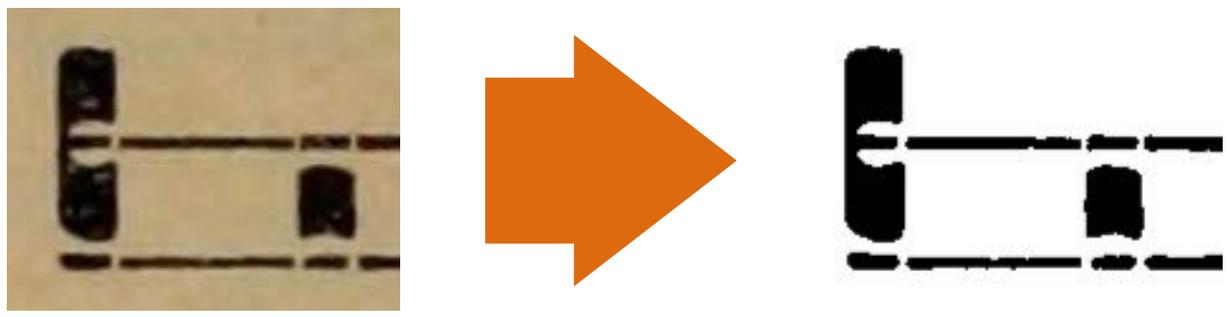
INTER-SYMBOL RELATIONSHIPS, NOTE PITCH, DURATION, CHORD VOICING

TRANSLATE TO ENCODING FORMAT (MIDI, MUSICXML, MEI)


Gamera!
映像の大怪獣ガメラ

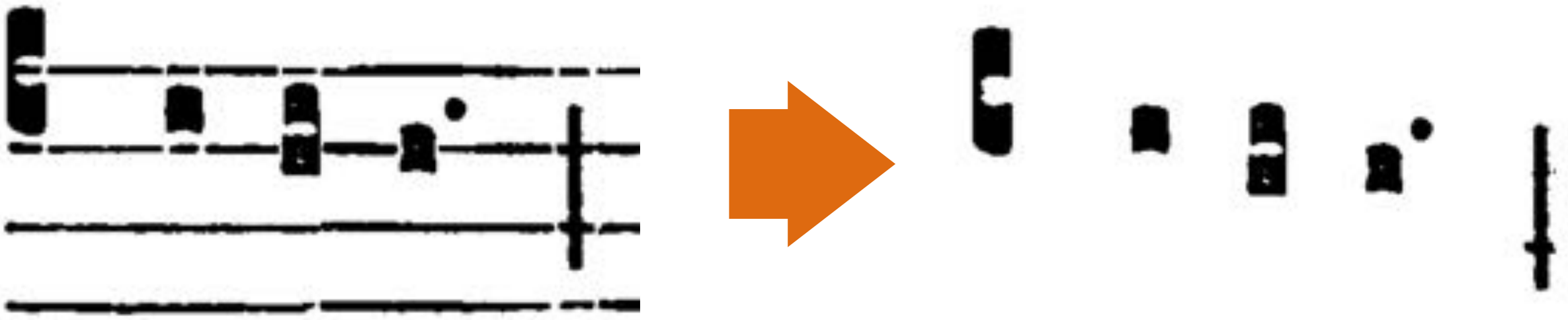
Binarization

(Foreground/Background Separation)



Staffline Removal

(Everything is Connected)*

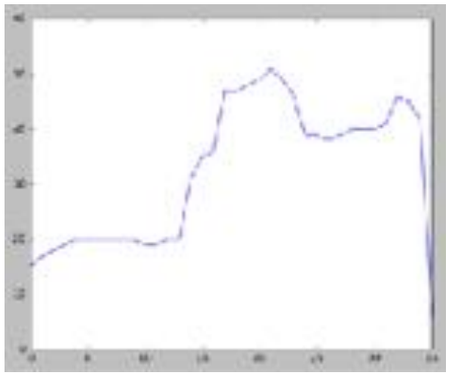


*The Zen of OMR

Connected Component Analysis [Not Everything is Connected :( ]

Features

(Computational Description of Blobs)



Features

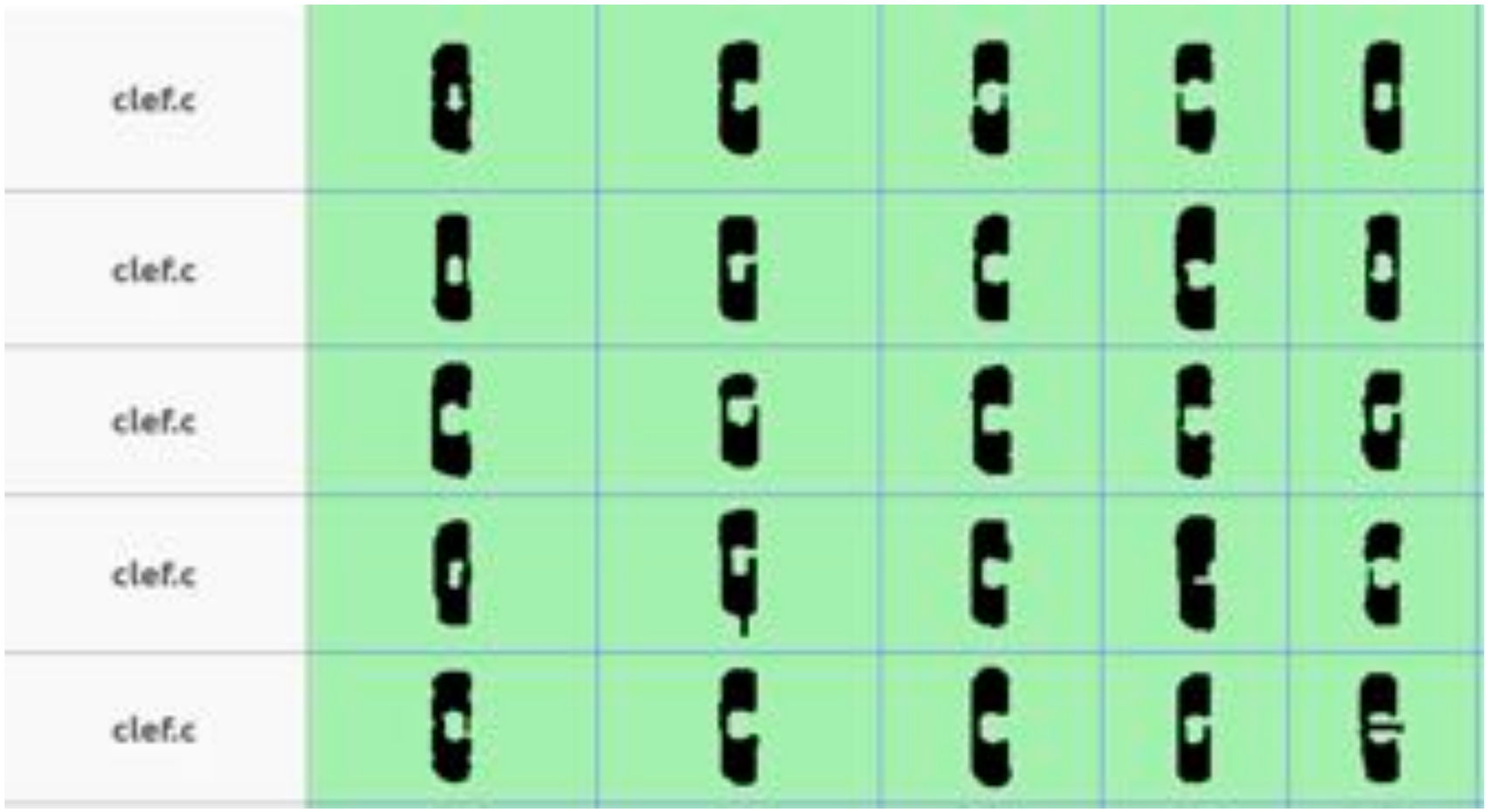
(Computational Description of Blobs)

```
area: 1944.0

nholes: [0.416, 0.277]

zernike_moments: [0.3258829217998120 4,
0.1226942921445565 4, 0.0772662387064154 7,
0.1401601631805497 2, 0.0274400100001215 32,
0.0860754582976985 8, 0.0578217146025021 8,
0.1047561270348521 8, 0.1986143145993405 6,
0.0898182635376208 2, 0.0891821975932056 7,
0.0571916065030057 5, 0.0948485965289453 1,
0.0392342006750479 1]
```

http://gamera.informatik.hsnr.de/docs/gamera-docs/features.html

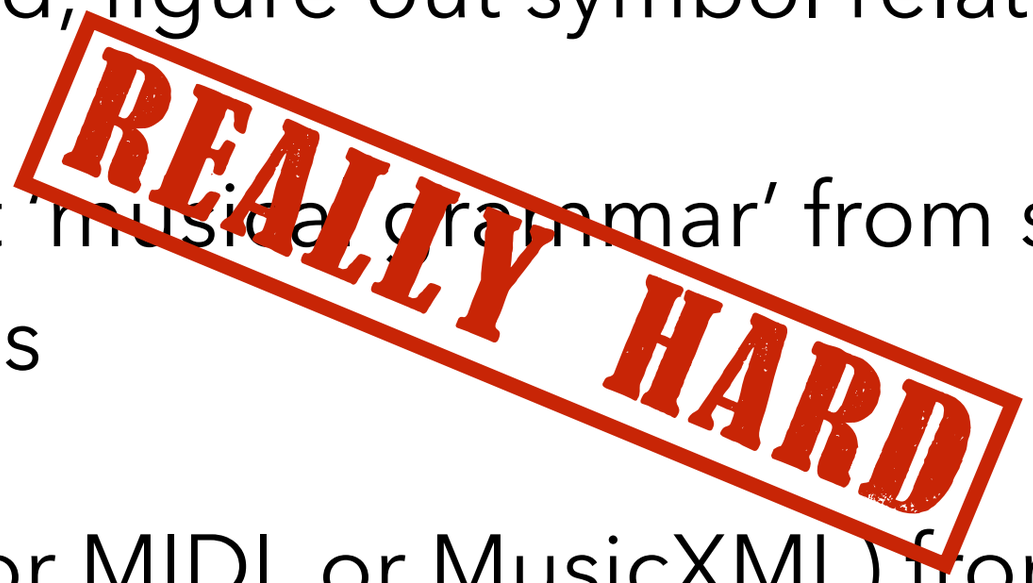
Classification

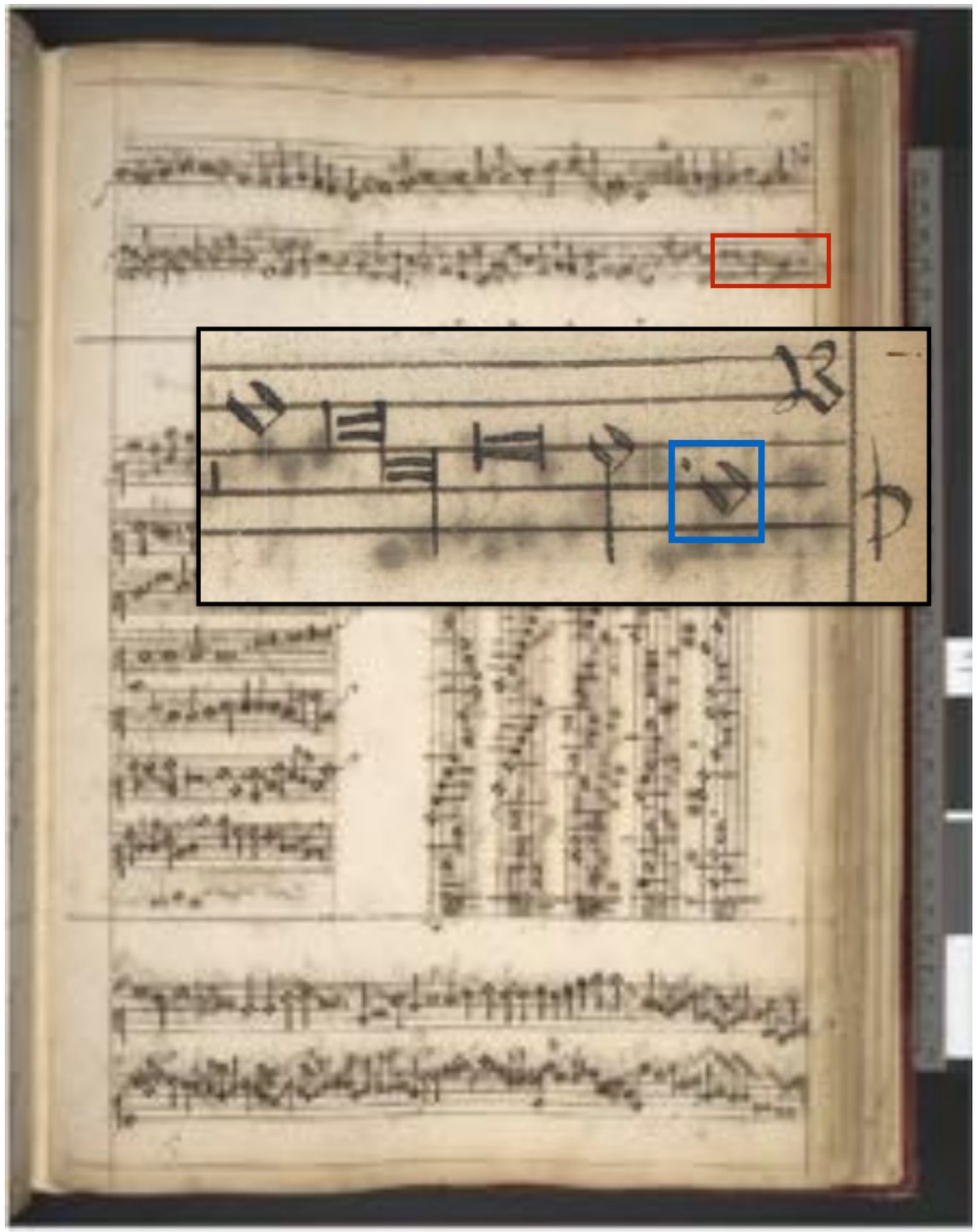
(Assigning Names to Blobs)



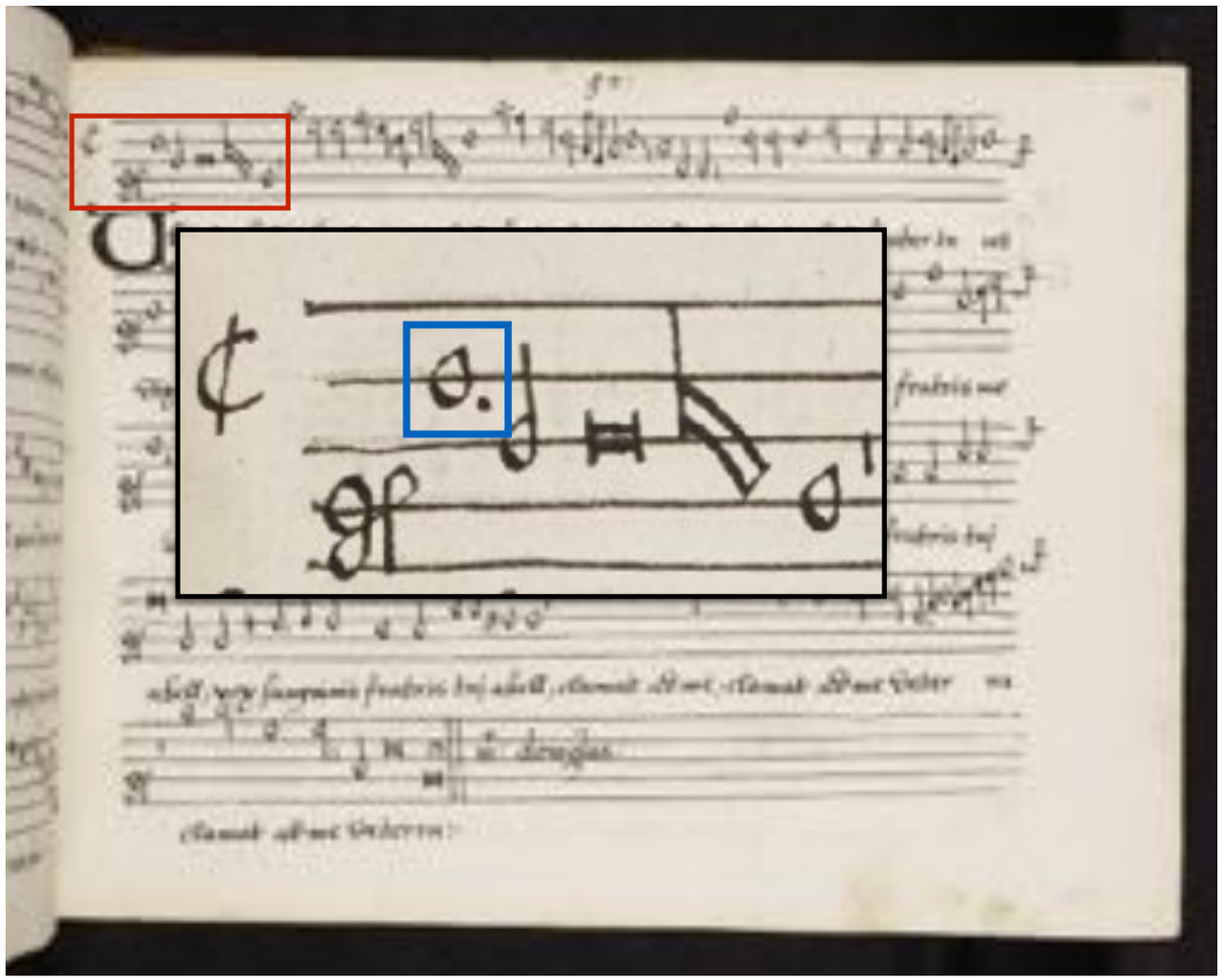
The Other Stuff

- Once named, figure out symbol relationships
- Reconstruct 'musical grammar' from symbol relationships
- Write MEI (or MIDI, or MusicXML) from musical grammar.





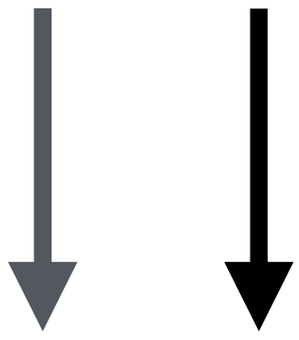
***Ubi est Abel frater tuus dixit Dominus*, Orlando di Lasso**
f. 80r, GB-Lbl Add. MS. 31390
https://www.diamm.ac.uk/sources/1888/



***Ubi est Abel frater tuus dixit Dominus*, Orlando di Lasso**
f. 103, GB-Och Mus. 979 (Baldwin Partbooks) [Superius]
https://www.diamm.ac.uk/sources/2348/

```
facsimile
 surface
   graphic@xml:id="g1" target="[Dow image]"
   zone coordinates="xywh" data="#n36"
```

```
facsimile
 surface
   graphic@xml:id="g2" target="[Add. 31390 image]"
   zone coordinates="xywh" data="#n36"
```



```
<note xml:id="n36" ... pname="d" />
```

Thank you.

Bodleian Libraries
UNIVERSITY OF OXFORD

McGill

Schulich School of Music
École de musique Schulich

SIMSSA | Single Interface for Music Score Searching and Analysis

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada

Canada

Fonds de recherche Société et culture Québec

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB

Bibliography

Deegan, M., E. King, and E. Steinvil. 2001. Project report: British Library microfilmed newspapers and Oxford grey literature online. The British Library, London.

Handel, P. 1931. Statistical machine. US Patent 1,915,993, filed 27 April 1931, and issued 27 June 1933.

Hockey, S. 1986. OCR: The Kurzweil data entry machine. *Literary and Linguistic Computing* 1 (2): 63–7.

Hoffman, M., L. O'Gorman, G. Story, J. Arnold, and N. Macdonald. 1993. The RightPages™ service: An image−based electronic library. *Journal of the American Society for Information Science* 44 (8): 446–52.

Kassler, M. 1972. Optical character recognition of printed music: A review of two dissertations. *Perspectives of New Music* 11: 250–4.

Mori, S., C. Suen, and K. Yamamoto. 1992. Historical review of OCR research and development. *Proceedings of the IEEE* 80 (7): 1029–58.

Phelps, T., and R. Wilensky. 1996. Toward active, extensible, networked documents: Multivalent architecture and applications. In *Proceedings of the First ACM International Conference on Digital Libraries*. Bethesda, MD, 20–23 March, 100–8.

Prerau, D. 1971. Computer pattern recognition of printed music. *AFIP Joint Computer Conferences* 39: 153–62.

Pruslin, D. 1966. Automatic recognition of sheet music. Sc. D. diss., Massachusetts Institute of Technology.

Schonfeld, R. 2003. *JSTOR: A history*. Princeton, NJ: Princeton University Press.

Varley, T. 1969. Data input error detection and correction procedures. George Washington University Logistics Research Project. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=AD0689365.