

Glyph Miner Slides

frame001

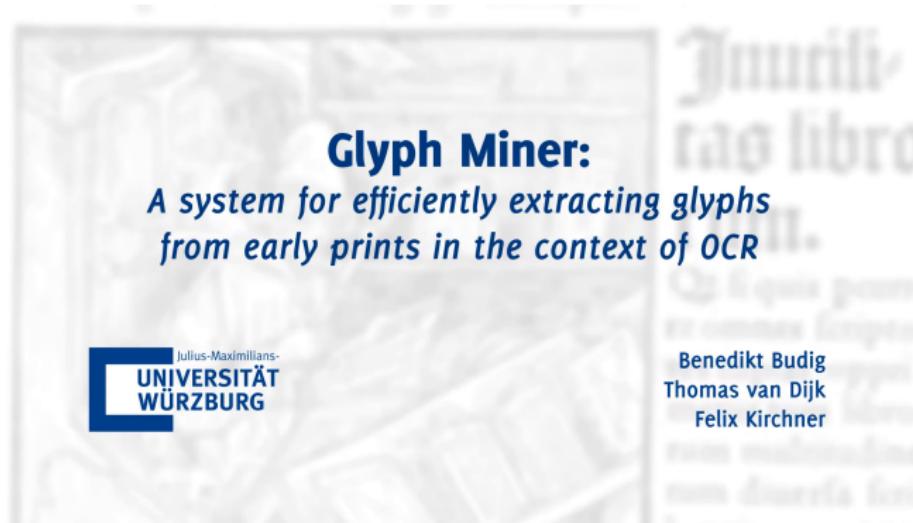
Glyph Miner Slides

2025-04-26

└ frame001

In this video, we showcase a system for extracting glyphs from early typeset prints.

frame002

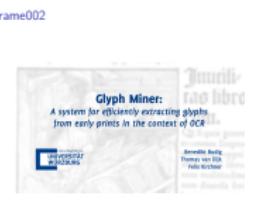


Glyph Miner Slides

2025-04-26

└ frame002

The system helps for training OCR software by rapidly locating sample occurrences of given glyphs.



frame002



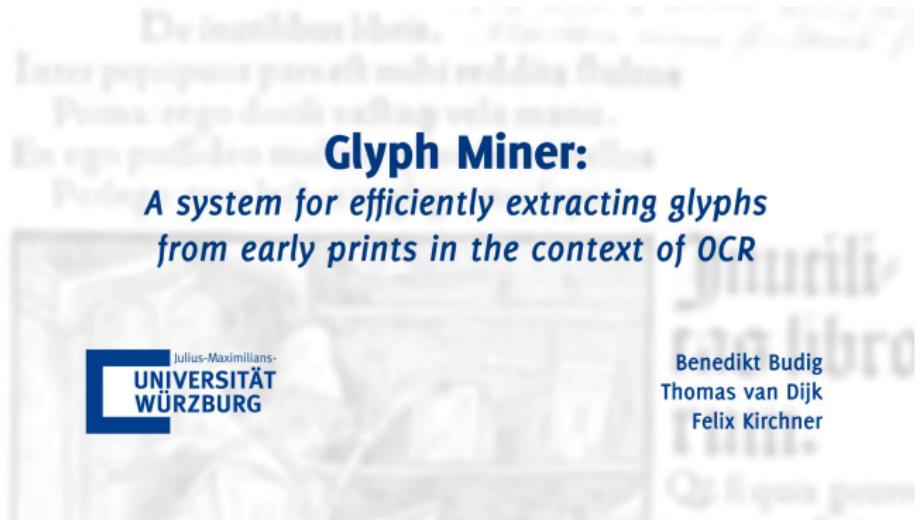
Glyph Miner Slides

2025-04-26

└ frame003

Let's have a look at how this works.

For this demonstration, we use scans of an incunable printed in Basel in 1497,



2025-04-26

└ frame005



Inutilitas librum.

containing a Latin translation of Sebastian Brandt's famous Narrenschiff.

XI Et blattas abigit pulvinis ulpius
De inutilibus librīs. Vix dicens anima huius col-
Inter prēcipuos pars est mihi redditā stultos
Prima:rego docili vastaq̃ vela manu.
En ego possideo multos/quos raro libellos
Perlego:tum lectos negligo:nec sapio:



Inutilitas librorum.

Glyph Miner Slides

2025-04-26

└ frame006

Here you see an overview of the documents that have been uploaded to our system.

The screenshot displays the 'Document Overview' section of the 'Historical Image Viewer & Glyph Library'. It features a grid of document thumbnails and two side-by-side form panels for adding new documents and collections.

- Thumbnail Grid:** Contains 10 thumbnails representing various historical documents, including maps and manuscripts. Each thumbnail includes a title, a brief description, and a 'Select' button.
- Add new document:** A form with fields for Title, Subtitle, Author, Year, Signature, Collection, and two file upload fields for Image (color) and Image (bw). A 'Upload document' button is at the bottom.
- Add new collection:** A form with fields for Title, Subtitle, Author, Year, and Signature. A 'Create collection' button is at the bottom.

frame006

Glyph Miner Slides

2025-04-26

└ frame007

We select the incunable and go to the next screen.

The screenshot shows the 'Document Overview' section of the Glyph Miner Slides application. It displays a grid of historical documents, each with a thumbnail, title, and brief description. The documents include:

- Das Narrenschiff (Test)**: Narrengrau, Text mit wenigen Seiten. Beaten Brot -- 1494. Select button.
- Das Narrenschiff**: Stultorum natalis : narrenpsis perfectioria rumpens satis lacuata macta / per ... Beaten Brot -- 1487. Select button.
- Circulus Franconicus**: In quo sunt episcopatus Würzburg, Bamberg, et Aachens, Status Equites Teutonici, Ducatus Coloniae, Marchionatus Brandenburgi, et ducatus Saxe, Henneberg, Wertheim, Holzsch, Reuss, Papehein, Eptach, Schwartzenberg, et Frederik De Wit -- 1706. Select button.
- Anonyme Homilie über M...**: Kölner Psalter, 81 ff., 260 x 195 mm, Würzburg, 1. Drittel 9. Jahrhundert. Select button.
- Das Bisthum Wurtzburg I...**: I H. S. dehewet I Scholast. ex Nürnberg Joh. Hoffm. 1516 - Kopienbuch, foliant. 32 x 38 cm - Maßstab: ca. 1280000. Select button.
- Das Frankenland**: Chorographia Francie (Inselbild) Sebastian von Rottenhan - Aufs new wideretur in Tract. München: Peter Küng, ca. 1620 - Holzschnitt und Kupferstich, 56,4 x 35,6 cm. Select button.
- Franciae Orientalis**: (vgl. Frankenland) descriptio Auctore Sebastian a Rotenhan. Antwerp: Abraham Ortelius, Antwerpen 1570 und 1612 - Kupferstich, holzs. 24,8 x 30,2 cm. - Maßstab ca. 1:900.000. Select button.
- Franciae Orientalis (crop...**: Sebastian von Rottenhan/Abraham Ortelius -- 1570. Select button.

To the right of the document grid is a sidebar for adding new documents and collections. The 'Add new document' section includes fields for Title, Subtitle, Author, Year, Signature, Collection, and two file upload fields for Image (color) and Image (bw). The 'Upload document' button is blue. Below it is the 'Add new collection' section with similar fields and a 'Create collection' button.



Graph Miner Slides

-frame008



Now we see the title page.



Glyph Miner Slides

2025-04-26

frame009

On the left, we can navigate to the other pages of this collection.

Document Overview Historical Image Viewer Glyp Library API Docs

Collection "Das Narrenschiff"

R0001r Size: 2166 x 3219 Select

R0001v Size: 2166 x 3219 Select

R0002r Size: 2166 x 3219 Select

R0002v Size: 2166 x 3219 Select

R0003r Size: 2166 x 3219 Select

R0003v Size: 2166 x 3219 Select

R0004r Size: 2166 x 3219 Select

R0004v Size: 2166 x 3219 Select

R0005r Size: 2166 x 3219 Select

R0005v Size: 2166 x 3219 Select

R0001r Size: 2166 x 3219 Select

Stultifera Navis.

Marragonice pfermonis numip
fina laudata Nois per Sethaftani Brate versando val-
gatis ferme & chytrimo p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &
glare prope sterne & lumen p' vltos mortalit facinantis
et falegans impulsi dianone speculo vltinodig &

447.
Natalia Olafsa.
Jo. de Opr.

Detect Threshold

Navigation icons: back, forward, search, etc.

Lyph Miner Slides

—frame010

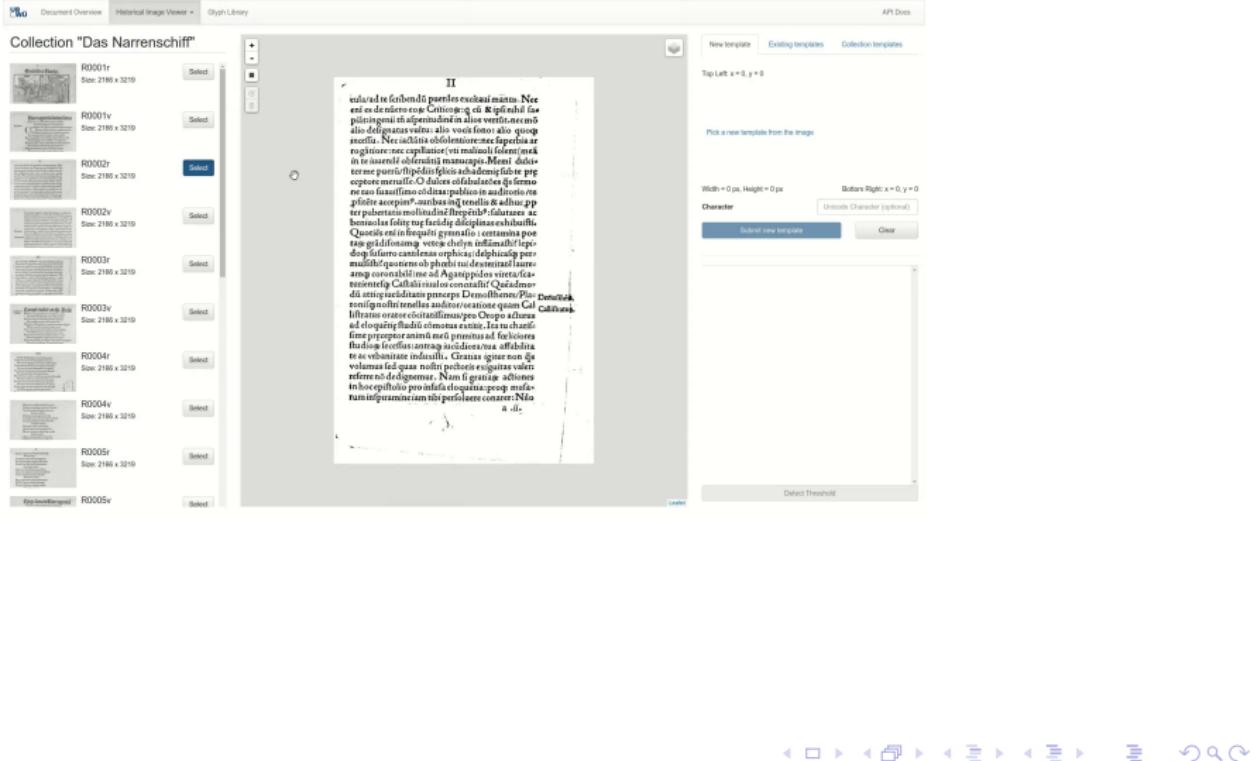


Let's pick a page that contains a lot of text.

Glyph Miner Slides

└ frame011

In this example, we want to find occurrences of the character E to train an OCR engine on.



frame012

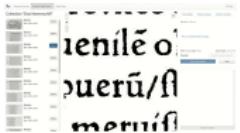


Glyph Miner Slides

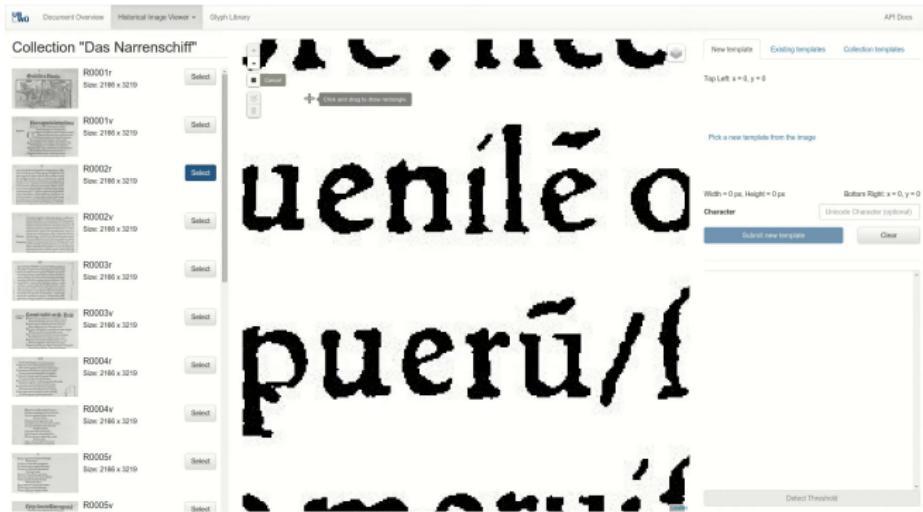
2025-04-26

frame012

First, we select a decent E somewhere in the text by selecting it with the rectangle tool.



frame013



Glyph Miner Slides

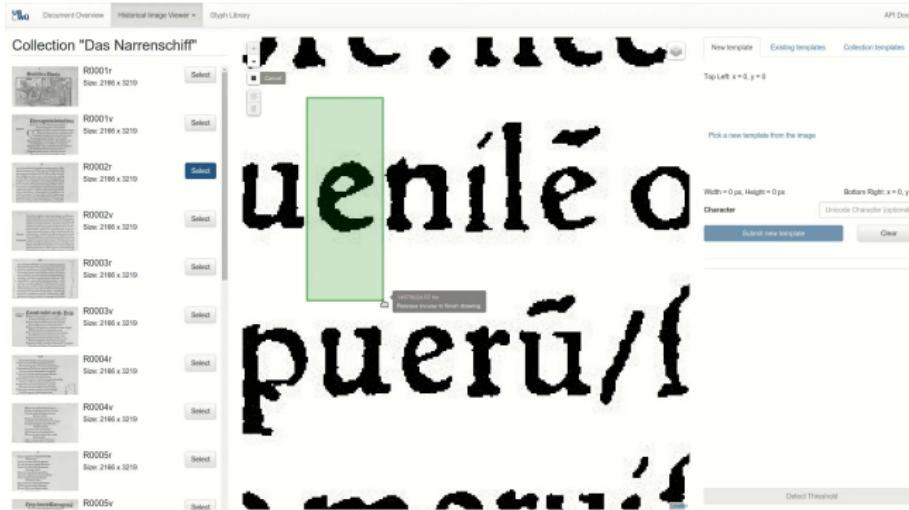
2025-04-26

frame013

Note that we crop it generously to include the white space above and below.



frame014

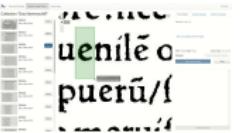


Glyph Miner Slides

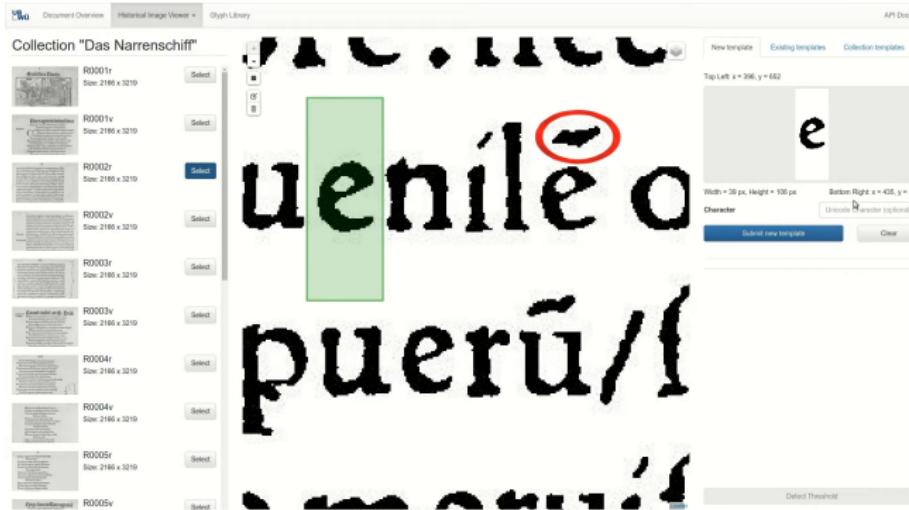
2025-04-26

frame014

This white space distinguishes it from similar glyphs but with accents, such as the one to the right.



frame015



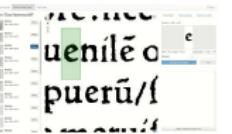
Glyph Miner Slides

2025-04-26

frame015

Next, we tell the system that what we cropped was an E so that we can find it more easily later,





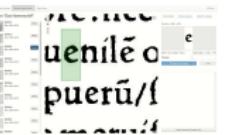
Glyph Miner Slides

2025-04-26

frame016

and then we submit this as a new template.

frame016



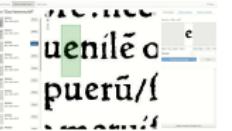
Glyph Miner Slides

2025-04-26

frame017

In the background, the system now starts a template matching algorithm

The screenshot shows the 'Glyp Library' tab of the Glyph Miner Slides application. On the left, a sidebar lists document thumbnails with names like 'R0001r', 'R0001v', etc., up to 'R0005v'. In the center, a preview area displays the text 'uenilé o puerú/f'. A green rectangular box highlights the letter 'e' in 'uenilé'. To the right of the preview, a small window shows a template for the character 'e' with dimensions: Width = 38 px, Height = 100 px, Top Left x = 296, y = 652, Bottom Right x = 435, y = 750. Below this window are buttons for 'Select new template' and 'Clear'. At the bottom of the preview area, there is a button labeled 'Detect Threshold (Template 310)'. The bottom of the screen features a navigation bar with various icons.



Glyph Miner Slides

2025-04-26

frame018

that searches for occurrences over E on all of the pages in the collection.
We have only uploaded 20 pages for this demonstration,

frame018



Glyph Miner Slides

2025-04-26

frame020

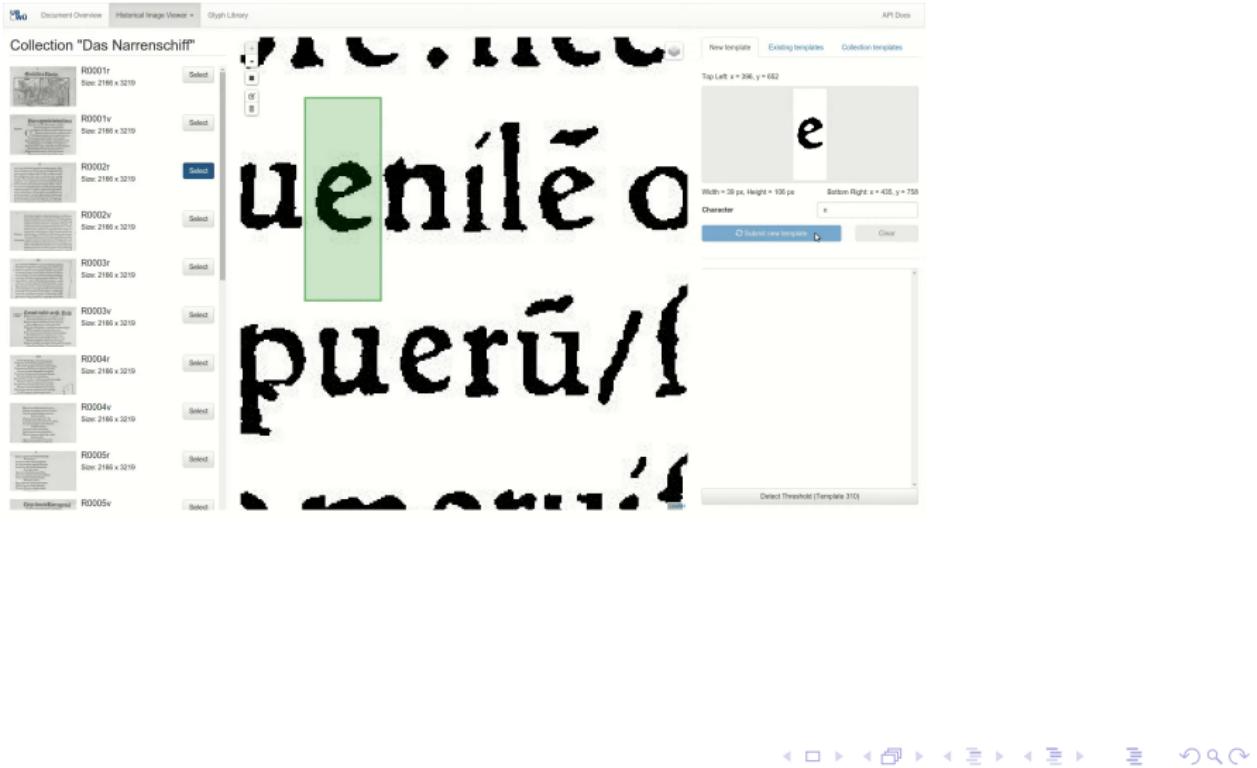
but our algorithm could as well search all 320 pages of this incunable.

frame020

2025-04-26

└ frame021

This would just take a little longer.



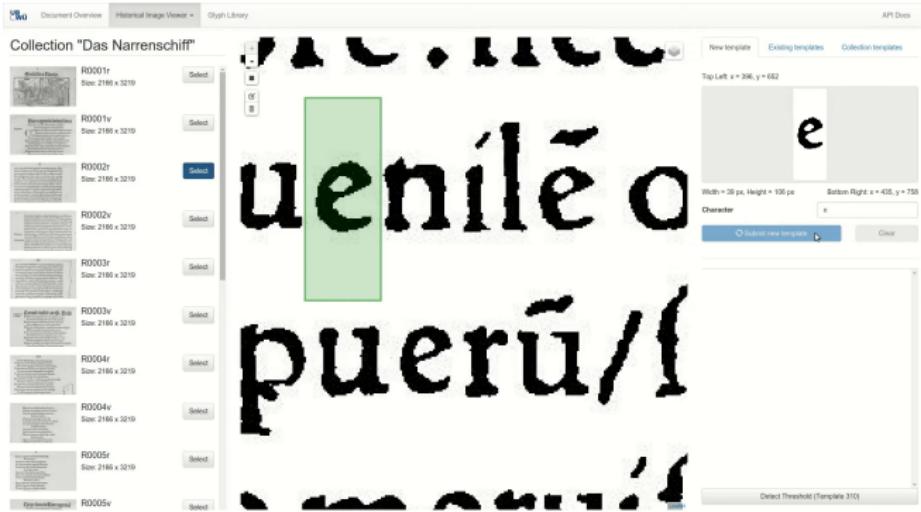


Glyph Miner Slides

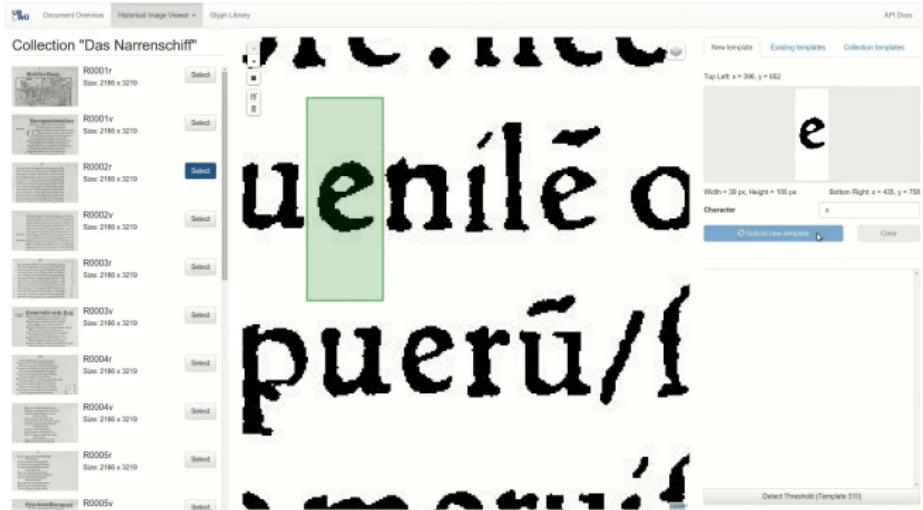
2025-04-26

frame022

Once the template matching is done, we are presented with a set of candidate matches.



frame023

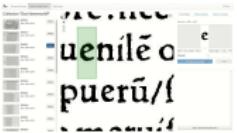


Glyph Miner Slides

2025-04-26

frame023

They are color-coded according to their similarity to the template we indicated.

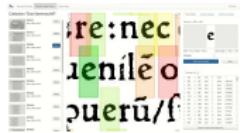
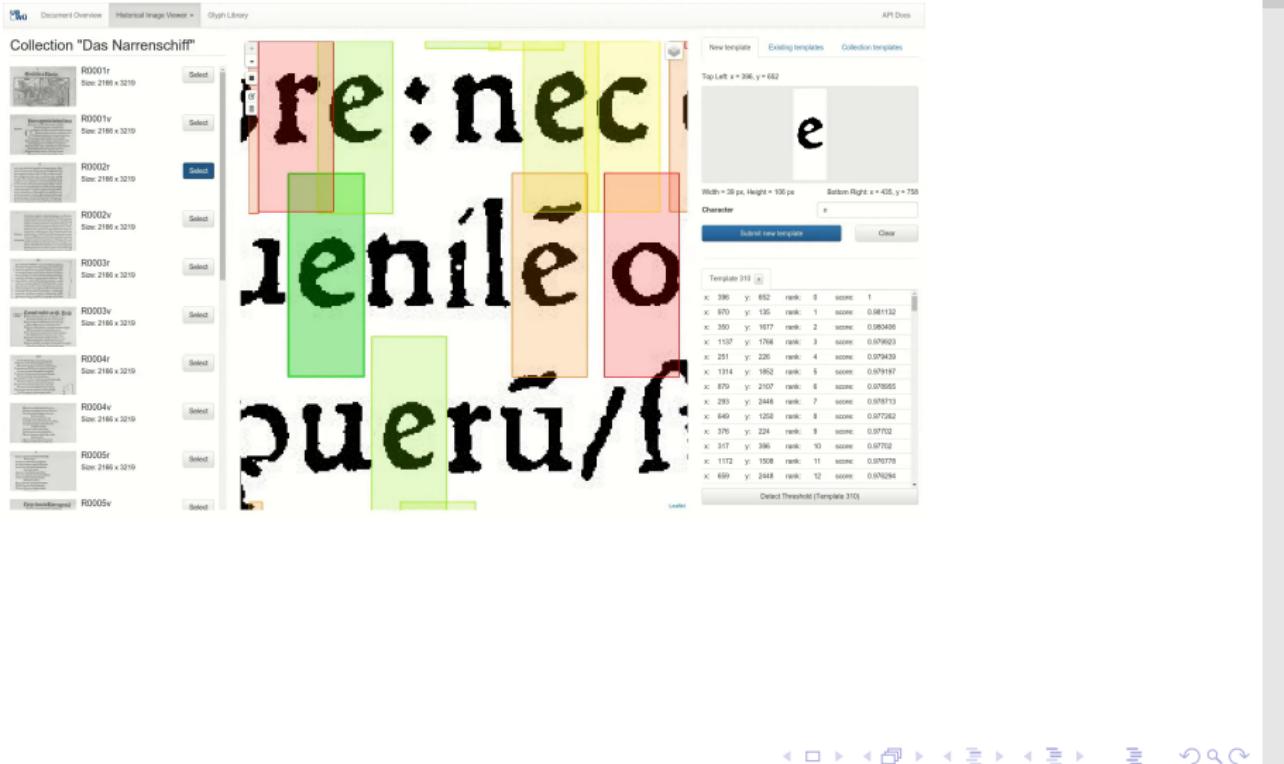


Glyph Miner Slides

2025-04-26

└ frame024

As you can see, our system correctly located some Es.





frame025

Character	x	y	rank	score
e	396	652	1	0.981132
e	870	135	2	0.980495
e	350	1777	3	0.979693
e	157	1764	4	0.979433
e	251	226	5	0.979197
e	1314	1852	6	0.979055
e	879	2157	7	0.979113
e	293	2444	8	0.97702
e	649	255	9	0.97702
e	376	224	10	0.97702
e	317	296	11	0.97702
e	1172	1508	12	0.976773
e	659	2448	13	0.976794

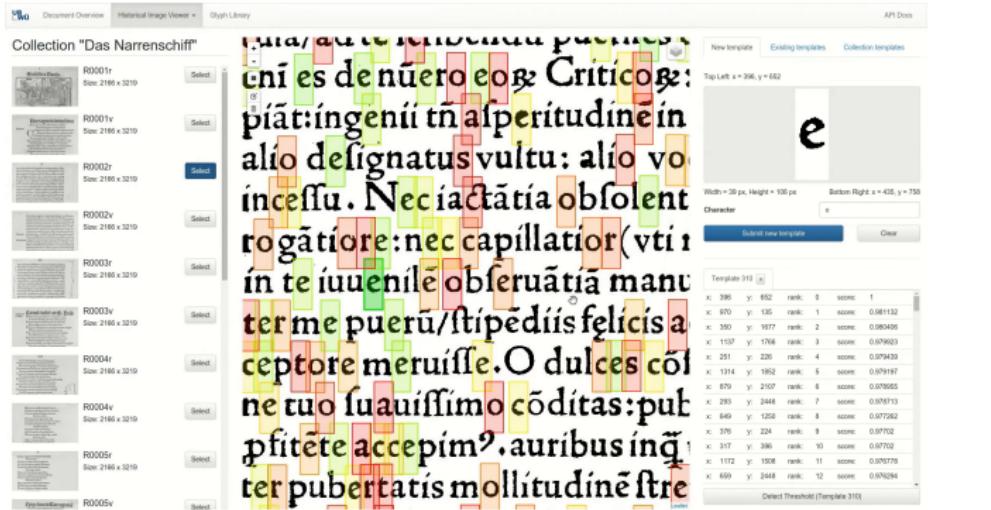
Glyph Miner Slides

2025-04-26

frame025

There are also a whole lot of false positives.

frame026



Glyph Miner Slides

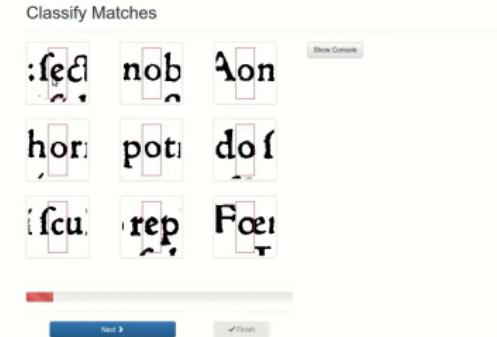
2025-04-26

└ frame026

At this point, we switch to the classification interface, which will help us to get rid of these false positives.



frame028



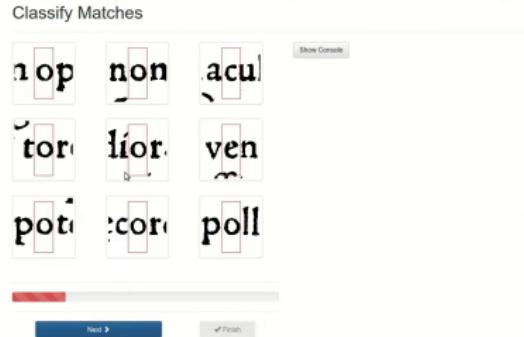
Glyph Miner Slides

2025-04-26

└ frame028

Here, we are presented with a batch of nine candidates

frame029



Glyph Miner Slides

2025-04-26

frame029

and have to tell the system which of them are correct and which are not. The system uses this information to iteratively train a logistic regression model

frame031



Glyph Miner Slides

2025-04-26

└ frame031

from which it eventually learns to distinguish between the correct and the incorrect matches.

frame032

Glyph Miner Slides

2025-04-26

└ frame032

For details on the algorithm involved, see the link in the description below.



frame033



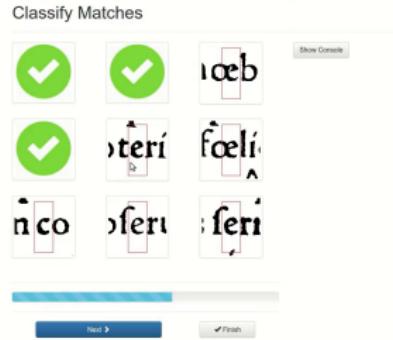
Glyph Miner Slides

2025-04-26

└ frame033

After about 10 iterations, the model is sufficiently trained,

frame034



Glyph Miner Slides

2025-04-26

frame034

and we go back to the document view.

frame035

Classify Matches

r/co	✓	✓
ocíe	✓	vicia
✓	✓	sco

Show Console

Next >  Finish

Glyph Miner Slides

2025-04-26

frame035

Now, only those matches that the system classifies as correct are shown.



Glyph Miner Slides

2025-04-26

frame036

Classify Matches

Show Console

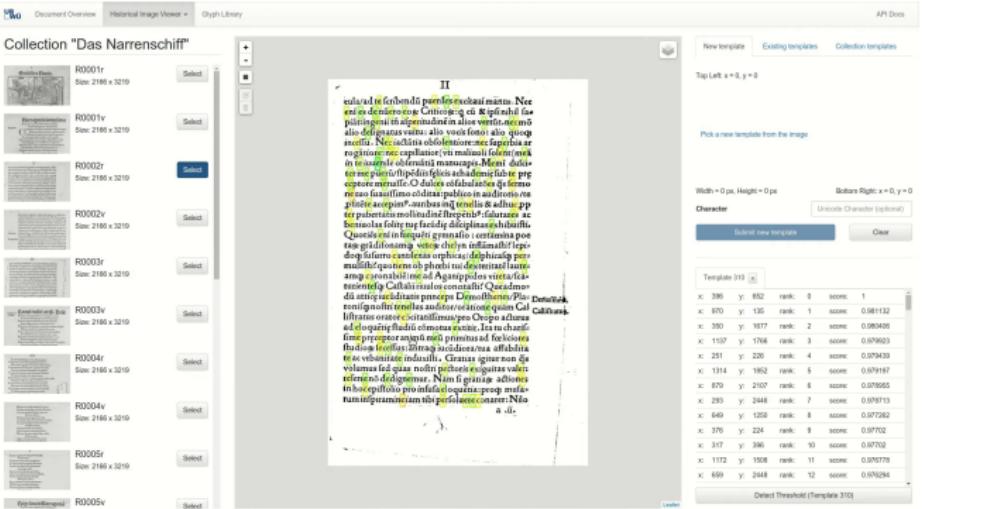
Next > Finish

We can see that almost all false positives have been removed,

Graph Miner Slides

frame037

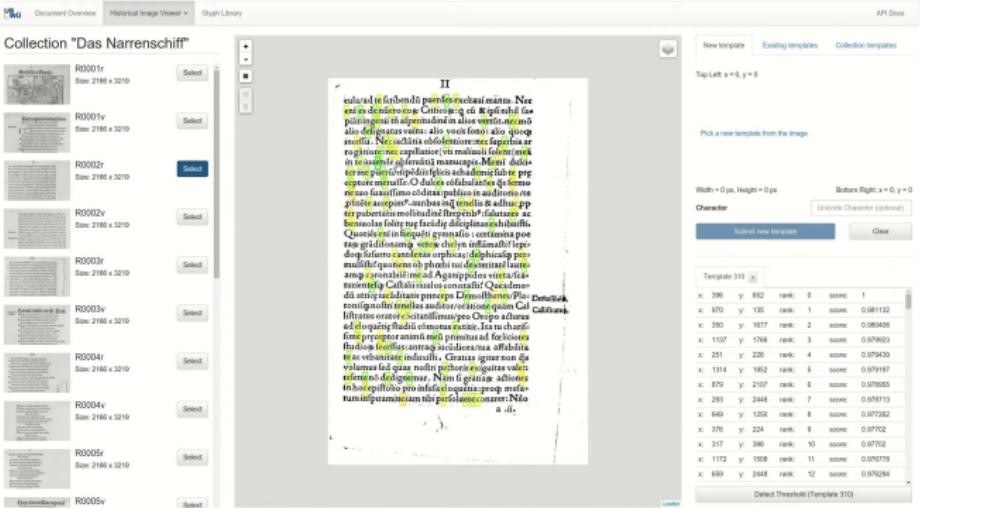
While most of the correct matches are still there.



Graph Miner Slides

frame038

This is also true for the other pages in the collection.



frame039

The screenshot shows the "Document Overview" tab selected in the top navigation bar. Below it, a list of pages from "Das Narrenschiff" is shown, each with a thumbnail and a "Select" button. The main area displays a portion of a page with Latin text. Overlaid on the text are several green rectangular boxes highlighting specific characters or groups of characters. A "Template" tool interface is overlaid on the right side of the text area. It includes fields for "Top Left: x = 0, y = 0", "Width: 0 px, Height: 0 px", and "Bottom Right: x = 0, y = 0". A "Character" dropdown is set to "Unselect Character (placeholder)". Buttons for "Create new template" and "Clear" are present. Below this is a table titled "Template 313" with columns for "x", "y", "rank", "score", and "score". The table lists 12 points found in the template, with the first point having the highest score of 0.981132.

x	y	rank	score	score
396	182	0	0.981132	
870	135	1	0.980132	
350	1777	2	0.980495	
1157	1764	3	0.979623	
251	226	4	0.979433	
1314	1852	5	0.979197	
879	2157	6	0.979055	
293	2444	7	0.979113	
649	255	8	0.977082	
376	224	9	0.97702	
317	390	10	0.97702	
1172	1508	11	0.976773	
659	2448	12	0.976794	

Detect Threshold: Template 313

Glyph Miner Slides

2025-04-26

frame039

We have found thousands of occurrences of this glyphs in mere minutes.

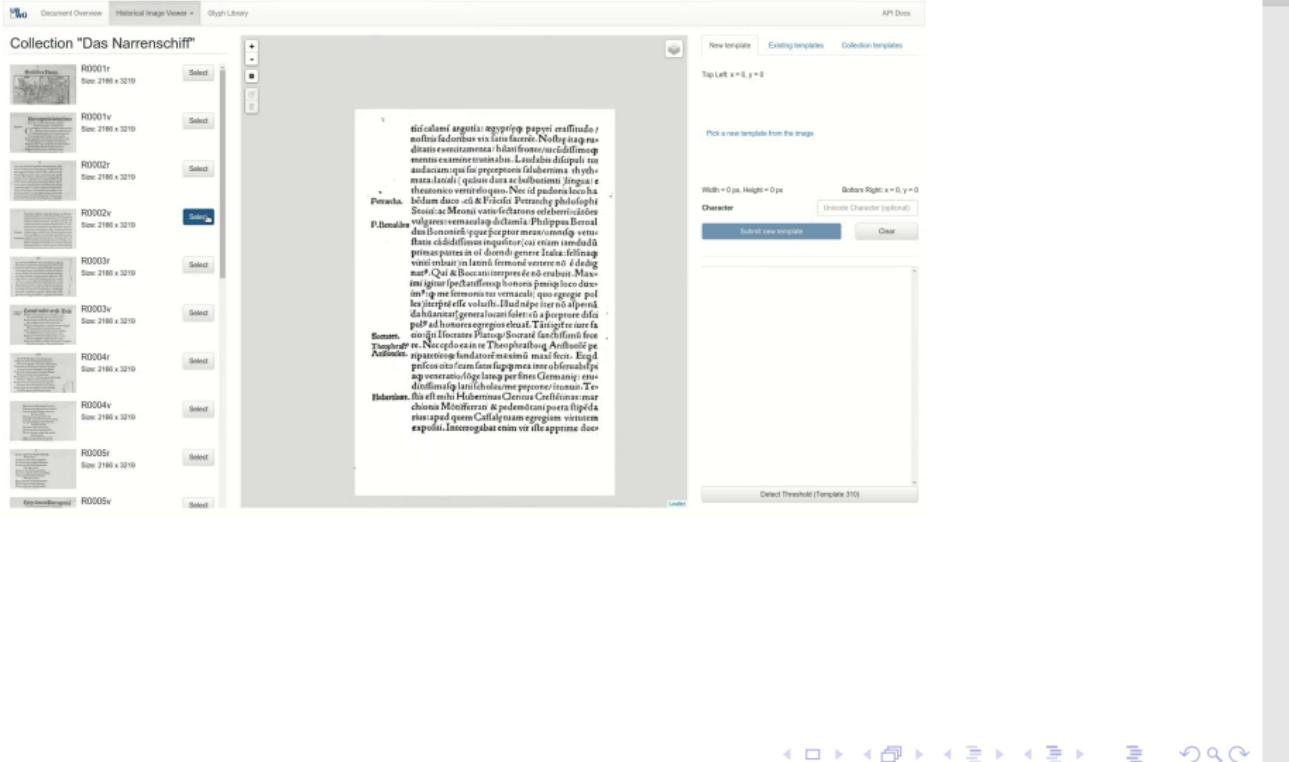


Glyph Miner Slides

—frame040

We now go to the glyph library,

卷二



Glyph Miner Slides

—frame041

where the results from the previous step are aggregated. All detected Es are displayed in the list on the right, sorted by quality.

Collection "Das Narrenschiff"

	R0001r Size: 2180 x 3219	<input type="button" value="Select"/>
	R0001v Size: 2180 x 3219	<input type="button" value="Select"/>
	R0002r Size: 2180 x 3219	<input type="button" value="Select"/>
	R0002v Size: 2180 x 3219	<input type="button" value="Select"/>
	R0003r Size: 2180 x 3219	<input type="button" value="Select"/>
	R0003v Size: 2180 x 3219	<input type="button" value="Select"/>
	R0004r Size: 2180 x 3219	<input type="button" value="Select"/>
	R0004v Size: 2180 x 3219	<input type="button" value="Select"/>
	R0005r Size: 2180 x 3219	<input type="button" value="Select"/>
	R0005v Size: 2180 x 3219	<input type="button" value="Select"/>

New template Existing templates Collection templates

Top Left: x = 0, y = 0

Pick a new template from the Image

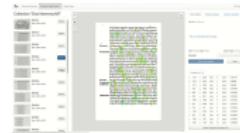
Width = 0 px, Height = 0 px Bottom Right: x = 0, y = 0

Character Unicode Character (optional)

Template 310

x	y	rank	size	score	
1318	2264	0	source	0.981574	
1315	2436	1	source	0.980808	
1417	2522	2	source	0.980258	
748	1496	3	source	0.974439	
883	8	2522	4	source	0.972229
680	286	5	source	0.967829	
1908	1838	6	source	0.972782	
881	2906	7	source	0.970778	
532	374	8	source	0.970336	
1395	2002	9	source	0.970284	
847	8	804	10	source	0.970062
1958	1921	11	source	0.970052	
1502	1158	12	source	0.970048	

Direct Threshold (Template 310)



Glyph Miner Slides

2025-04-26

└ frame043

About two-thirds of the way down the list,

The screenshot shows the 'Glyph Library' tab of a software application. On the left, there's a sidebar with 'Documents' containing thumbnail previews of various historical manuscripts. The main area is titled 'Templates' and lists characters 'g', 'a', 'e', 'p', 'o', 'v', 'u', 'n', 'i', 'c', 'e', and 'p' each associated with a specific template from a document. Each entry includes the template number, character name, position, and size. At the bottom, there are buttons for 'Select all', 'Deselect all', and 'Filter by character...', along with tabs for 'Template', 'User Positive', 'Predicted', 'Ranking', and 'Group'. A 'Download' button is also present.

Character	Template	Position	Size
g	Template 248 – "g"	(794, 1080)	42 x 69
a	Template 249 – "a"	(490, 1080)	39 x 46
e	Template 250 – "e"	(626, 1072)	32 x 39
p	Template 251 – "p"	(341, 2364)	44 x 71
o	Template 252 – "o"	(1112, 2356)	40 x 42
v	Template 253 – "v"	(164, 1080)	47 x 40
u	Template 254 – "u"	(871, 1080)	44 x 46
n	Template 263 – "n"	(1118, 2326)	30 x 45
i	Template 264 – "i"	(844, 1195)	45 x 67
c	Template 265 – "c"	(627, 1175)	33 x 40
e	Template 266 – "e"	(801, 1047)	42 x 45
p	Template 267 – "p"	(794, 2033)	34 x 39
p	Template 268	(1903, 1084)	52 x 67



frame044

The screenshot shows a software interface for managing historical documents and their associated glyphs. On the left, a sidebar lists "Documents" with thumbnail previews and details. On the right, a main panel titled "Glyphs" displays a list of characters (e.g., 't', 'g', 'a', 'c', 'i', 'g', 'e', 'e', 'e', 'e') each associated with a "Template" and its position and size. A vertical scroll bar is visible on the right side of the main panel.

Document	Template	Position	Size
Das Narrenschiff (Test)	Template 301 - "t119"	(458, 858)	30 x 38
Das Narrenschiff Salomon Brant - 1494	Template 302 - "012b"	(1914, 858)	34 x 57
Das Narrenschiff Salomon Brant - 1494	Template 303 - "a"	(1915, 2281)	25 x 55
Testcollection berndt - 2015	Template 304 - "c"	(1748, 858)	36 x 43
Circulus Francorum	Template 305 - "i"	(916, 778)	54 x 69
Arznei Hornisse über MI 11,3	Template 306 - "g"	(215, 588)	44 x 68
Das Bistum Würzburg in Franken	Template 307 - "e"	(320, 427)	34 x 43
Das Franckenland	Template 308 - "e"	(354, 1089)	32 x 81
Francise Orientalis	Template 309 - "e"	(398, 658)	36 x 92
Francise Orientalis (cropped)	Template 310 - "g"	(399, 653)	39 x 105
Kronen Albrecht			

Glyph Miner Slides

2025-04-26

└ frame044

we see that our glyphs have started to show significant variance,
but almost all of them are still correct.



frame046

The screenshot shows the 'Glyph Library' tab of a software application. On the left, there is a sidebar with a list of 'Documents' containing historical image viewer thumbnails. The main area displays a grid of 'e' and 'g' glyphs from various templates. Each template entry includes its name, position, and size. A vertical scroll bar is visible on the right side of the grid.

Template	Position	Size
e	(458, 858), Size: 30 x 38	
Template 301 – "0110"	(1914, 2116), Size: 34 x 67	
Template 302 – "0125"	(1855, 2281), Size: 25 x 55	
Template 303 – "a"	(1745, 1937), Size: 36 x 43	
Template 304	(596, 738), Size: 54 x 69	
Template 305	(215, 581), Size: 44 x 68	
Template 306	(320, 427), Size: 34 x 43	
Template 307 – "g"	(354, 1089), Size: 32 x 81	
Template 309 – "e"	(398, 608), Size: 36 x 92	
Template 310 – "e"	(398, 653), Size: 39 x 105	

Glyph Miner Slides

2025-04-26

frame046

We can see that the glyphs have started to show significant variance,



frame047

The screenshot shows the 'Glyph Library' tab of a software application. On the left, there's a sidebar with 'Documents' containing several historical image viewer entries. The main area is titled 'Templates' and displays a grid of character shapes. To the right of the grid, individual characters are listed with their template names and bounding box coordinates. The characters shown include 'e', 'f', 'g', 'h', 'i', 'j', 'k', 'l', 'm', 'n', 'o', 'p', 'r', 's', 't', 'u', 'v', 'w', and 'x'. Each character has a 'Select' button next to its name.

Character	Template Name	Position
e	Template 301 – "e"	(458, 858), Size: 30 x 38
f	Template 301 – "f"	(1914, 858), Size: 34 x 67
g	Template 302 – "f12b"	(1855, 2281), Size: 25 x 55
h	Template 303 – "g"	(1710, 1987), Size: 36 x 43
i	Template 304	(516, 738), Size: 54 x 89
j	Template 305	(215, 588), Size: 44 x 68
k	Template 306	(320, 427), Size: 34 x 43
l	Template 307 – "g"	(354, 1089), Size: 32 x 81
m	Template 309 – "g"	(398, 608), Size: 36 x 92
n	Template 310 – "g"	(398, 652), Size: 39 x 195

Glyph Miner Slides

2025-04-26

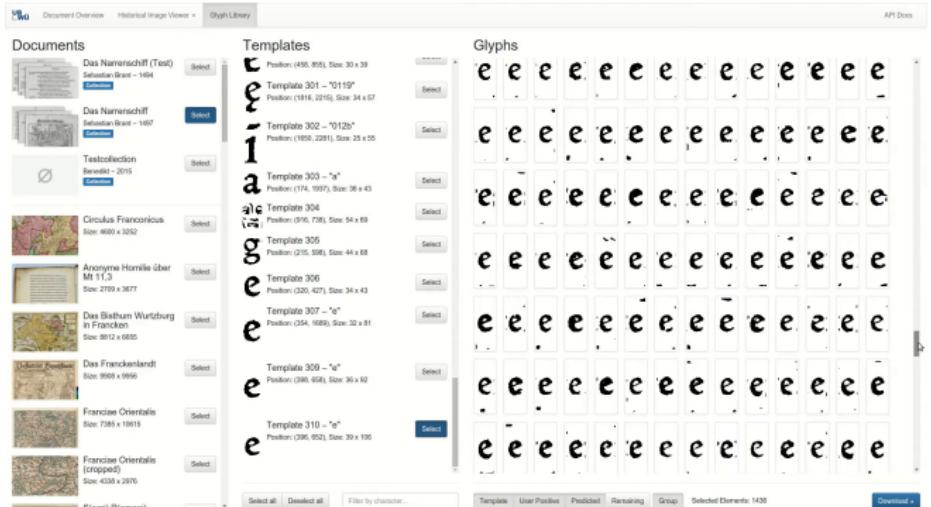
└ frame047

but almost all of them are still correct.

All the way to the bottom, there are also some false positives,



frame049



Glyph Miner Slides

2025-04-26

└ frame049

but relative to the number of glyphs found, this is not a significant fraction.
Here is another glyph we prepared earlier using this process.





Glyph Miner Slides

2025-04-26

└ frame051

Document	Template	Position	Size
Das Narrenschiff (Test)	Template 285 – "0075"	(328, 1588)	36 x 33
Das Narrenschiff Selbstdruck – 1464	Template 284 – "0072"	(1289, 1811)	31 x 41
Das Narrenschiff Selbstdruck – 1467			
Testcollection Berndt – 2015			
Circulus Francorum	Template 285 – "0075"	(328, 1588)	36 x 33
Das Narrenschiff Selbstdruck – 1464	Template 284 – "0072"	(1289, 1811)	31 x 41
Das Narrenschiff Selbstdruck – 1467			
Anonymous Hornisse über MI 11,3	Template 285 – "0075"	(328, 1588)	36 x 33
Anonymous Hornisse über MI 11,3	Template 284 – "0072"	(1289, 1811)	31 x 41
Das Bistum Würzburg In Franken	Template 285 – "0075"	(328, 1588)	36 x 33
Das Bistum Würzburg In Franken	Template 284 – "0072"	(1289, 1811)	31 x 41
Das Franckenland	Template 285 – "0075"	(328, 1588)	36 x 33
Das Franckenland	Template 284 – "0072"	(1289, 1811)	31 x 41
Francise Orientalis	Template 285 – "0075"	(328, 1588)	36 x 33
Francise Orientalis	Template 284 – "0072"	(1289, 1811)	31 x 41
Francise Orientalis (cropped)	Template 285 – "0075"	(328, 1588)	36 x 33
Francise Orientalis (cropped)	Template 284 – "0072"	(1289, 1811)	31 x 41
Kirchenkarte	Template 285 – "0075"	(328, 1588)	36 x 33
Kirchenkarte	Template 284 – "0072"	(1289, 1811)	31 x 41

We see much fewer occurrences, because the CT ligature is relatively rare.
These data can now be downloaded as image files,

frame053

The screenshot shows the 'Glyph Library' interface. On the left, there's a sidebar with 'Documents' containing nine thumbnail images of historical documents, each with a 'Select' button. The main area is titled 'Templates' and shows a grid of character glyphs. A specific row of 'a' characters is highlighted with a black border. Below the grid, there's a section for 'Character "e"' with a grid of 'e' characters. At the bottom, there are buttons for 'Selected Elements: 1516', 'Download', and navigation icons.

Glyph Miner Slides

2025-04-26

└ frame053

or exported in the page.xml file.



frame054

The screenshot shows the 'Glyph Library' interface. On the left, there's a sidebar with 'Documents' containing several thumbnail images of historical documents like 'Das Narrenschiff (Test)', 'Das Narrenschiff (Sebastien Brant - 1494)', 'Das Narrenschiff (Sebastien Brant - 1497)', 'Testcollection (Berndolt - 2015)', 'Circulus Francoricus (Size: 4600 x 352)', 'Anonyme Hornisse über MI 11,3 (Size: 2189 x 3877)', 'Das Bistum Würzburg in Franken (Size: 8812 x 6655)', 'Das Franckenland (Size: 3688 x 9556)', 'Francise Orientalis (Size: 7385 x 19815)', 'Francise Orientalis (cropped) (Size: 4338 x 3976)', and 'Kronen Albrecht'. Below this is a 'Templates' section with a list of items for 'Character "d"' and 'Character "e"'. The main area displays a grid of character templates for 'd' and 'e', each with a preview image, template number, character code, position, size, and a 'Select' button.

Character	Template ID	Character Code	Position	Size
d	Template 283	"0075"	(322, 1588)	Size: 36 x 30
d	Template 284	"0072"	(1289, 1811)	Size: 31 x 41
d	Template 285	"elff"	(342, 1980)	Size: 65 x 84
e	Template 286	"0081"	(827, 2037)	Size: 37 x 36
e	Template 287	"ot"	(1126, 2190)	Size: 58 x 61
e	Template 288	"0065"	(543, 2348)	Size: 33 x 70
d	Template 289	"0043"	(820, 2333)	Size: 36 x 37
m	Template 290	"0084"	(863, 2307)	Size: 61 x 42
f	Template 291	"fb05"	(1161, 1889)	Size: 52 x 68
e	Template 292	"at76"	(1116, 2476)	Size: 59 x 63
g	Template 293	"0067"	(279, 1988)	Size: 42 x 65
e	Template 294	"0066"		

Glyph Miner Slides

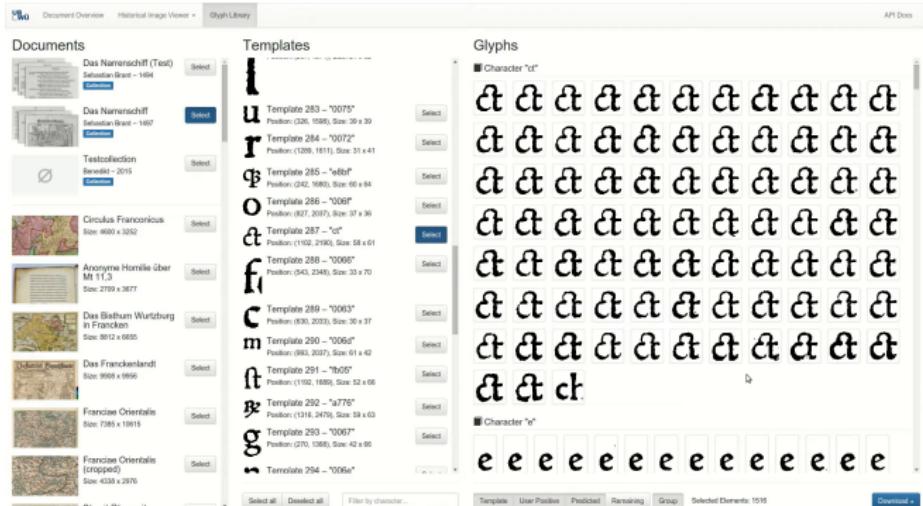
2025-04-26

└ frame054

As an example of what you can do with the output of our system,



frame055



Glyph Miner Slides

2025-04-26

└ frame055

we import the detected glyphs into FrankenPlus.
This is software for generating synthetic draining data for the Tesseract OCR engine.



frame057

The screenshot shows the 'Glyph Library' interface. On the left, there's a sidebar with 'Documents' containing thumbnail images of historical documents like 'Das Narrenschiff (Test)' by Sebastian Brant - 1494, 'Das Narrenschiff' by Sebastian Brant - 1497, 'Testcollection' by Berndolt - 2015, 'Circulus Francoricus' (Size: 4600 x 352), 'Anonyme Hornisse über MI 11,3' (Size: 2109 x 3877), 'Das Bistum Würzburg In Franken' (Size: 8812 x 6655), 'Das Franckenland' (Size: 3668 x 9556), 'Francie Orientalis' (Size: 7385 x 19815), and 'Francie Orientalis (cropped)' (Size: 4338 x 3876). The main area is titled 'Templates' and 'Glyphs'. It displays a grid of characters for the letter 'd' (Template 285 - "0075", Position: (328, 1598), Size: 38 x 33) and 'n' (Template 293 - "0067", Position: (270, 1598), Size: 42 x 85). Below the grid are buttons for 'Select', 'Character "d"', 'Character "n"', 'Template', 'User Posizer', 'Predicted', 'Resizing', 'Group', 'Selected Elements: 1516', and 'Download'. At the bottom, there are buttons for 'Select all', 'Delete all', 'Filter by character...', 'Template', 'User Posizer', 'Predicted', 'Resizing', 'Group', 'Selected Elements: 1516', and 'Download'.

Glyph Miner Slides

2025-04-26

└ frame057

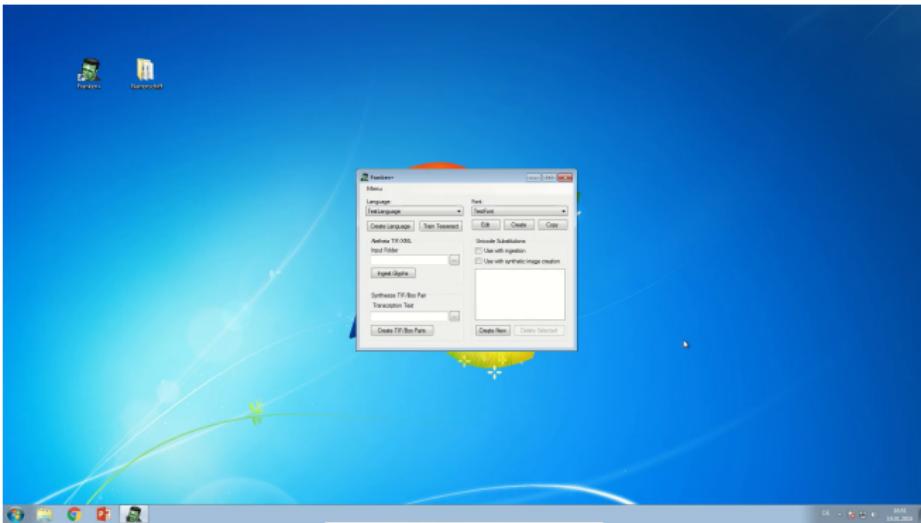
It was developed at the Texas A&M University,



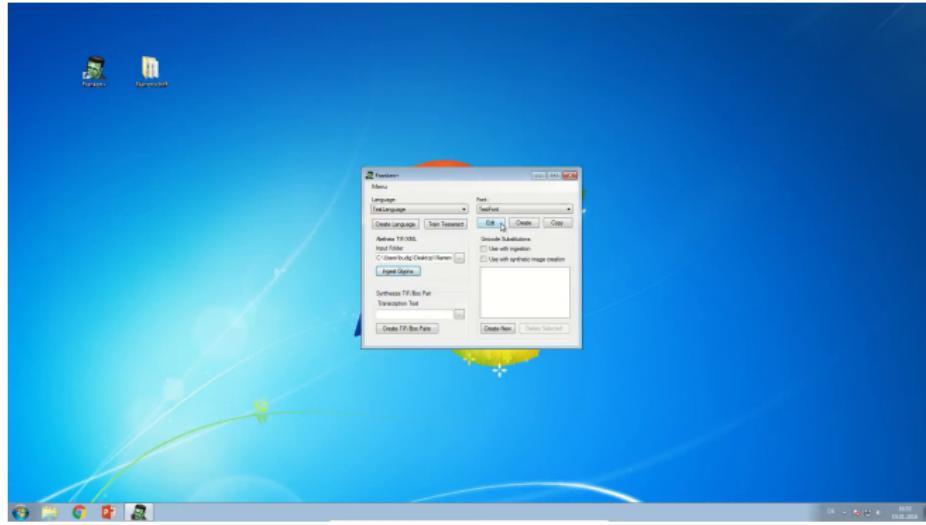
Glyph Miner Slides

└ frame058

and is designed to be used in a digitization pipeline for early prints.
As you can see here, our results are similar to those of the previous step.



frame060



Glyph Miner Slides

2025-04-26

└ frame060

This concludes our video, in which we gave a quick impression of the features of our system.

frame060



frame061



Glyph Miner Slides

2025-04-26

frame061

We hope you enjoyed it, and say thanks for watching.
Thank you for watching.

