Necrologium Lundense Online: A Digital Palaeographical Edition of Lund University Library, Medeltidshandskrift 6

Michael MacPherson
1/12/2015
4721-E15: Digital Diplomatics. Working with electronic texts
Instructor: Alex Speed Kjeldsen
## Contents

Section 1: Digital Palaeography, Project Overview, and Editorial Theory ................................................. 2  
  Section 1.1: Palaeography in Europe and Scandinavia .............................................................................. 2  
  Section 1.2: The Lund scriptorium and the *Necrologium Lundense* ...................................................... 4  
  Section 1.3: Digital palaeography, DigiPal’s formal model, and NLO’s interpretation of it ...................... 5  
  Section 1.4: Digital editorial theory in general. ...................................................................................... 11  
  Section 1.5: Concluding remarks and an overview of the structure of Sections 2 and 3 ....................... 12

Section 2: Pre-processing implementation .................................................................................................. 13
  Section 2.1: OCR, Post-correction, and transcription method ................................................................. 13  
  Section 2.2: Transforming the transcription to an array and the array structure. ............................... 14  
  Section 2.3: Text-image alignment and batch image cropping ............................................................... 17

Section 3: Front-End implementation ....................................................................................................... 18
  Section 3.1: Overall structure of the web page ....................................................................................... 18  
  Section 3.2: Functions relating to the Text tab ....................................................................................... 18  
  3.3: Functions relating to the Search tab ............................................................................................... 19  
  Section 3.4: Functions relating to the Collections tab ........................................................................... 20  
  Section 3.5: Functions relating to the Compare Hands tab ................................................................. 20  
  Section 3.6: Functions relating to the Statistics Tab .......................................................................... 21  
  Section 3.7: The annotation tool ........................................................................................................... 22

Appendix A – Allographs and Their Components ..................................................................................... 24

Bibliography ............................................................................................................................................... 32
Necrologium Lundense Online (hereafter, NLO) is an experimental “palaeographical edition” of the Necrologium Lundense (Medeltidshandskrift 6 held in the Lund University Library in Lund, Sweden). In particular, it focuses on the Memoriale fratrum (fols. 124v-173v), a necrology for the cathedral chapter containing 27 different scribal hands primarily from the years 1123-1170. The manuscript is written in Latin and is an excellent source for palaeographical study.

Section 1: Digital Palaeography, Project Overview, and Editorial Theory

Section 1.1: Palaeography in Europe and Scandinavia

NLO was conceived from the very beginning as a tool to study palaeography, bridging the gap between the early medieval development of Latin script with later (mostly vernacular) Scandinavian developments. The goal is to place the Necrologium Lundense in the narrative of Latin script as it moved from late antiquity, through the Caroline reforms, and into scriptoria across Europe through the creation of a database and web tools designed to research it. Scandinavia arrived late in this narrative. Gradually, Gothic script features would completely innovate the image of text-bearing objects. This history can be accessed by landmark works such as Bernhard Bischoff’s Latin Palaeography and Manuscripts and Libraries in the Age of Charlemagne, M.B. Parkes’ Their Hands Before Our Eyes, Albert Derolez’s Palaeography of Gothic Manuscript Books, Peter Stokes’ English Vernacular Miniscule, and others. This basic training also requires a good deal of pouring through the manuscript images themselves in an

1 Michael MacPherson, Necrologium Lundense Online (2015), available at <https://notendur.hi.is/mjm7>
effort to see changes as they progressed, and the *CMD* Catalogues of dated manuscripts are an invaluable resource along with the plates in the books mentioned above.³

For a student of Scandinavian scribes, scripts, and scriptoria, a lot of ground has been covered, though certainly the territory is less mapped out (particularly in East Norse). Important works include Hreinn Benediktsson’s *Early Icelandic Script*, *Palæografi A: Danmark og Sverige* edited by Brøndum-Nielsen, Seip’s *Palæografi B: Norge og Island*, Kålund’s palaeographic atlases, along with many others.⁴ My aim is to situate NLO in this tradition.

The main theme which dominates early Scandinavian script history is the question of European influence: where did Scandinavians learn to read and write and how is this reflected in their handwriting? Largely this is a matter of pointing out evidence of influence from either Insular features (from the British Isles) or Continental ones (mainly from France, the Low Countries, and – especially - Germany). Research to date has shown that West Norse script was influenced by Insular script more than East Norse, while East Norse script was influenced more by Continental scripts.⁵ But there is, of course, a great deal of gray area. Lots of Insular allographs and features still come up in East Norse manuscripts.⁶ Once this matter is dealt with, we then need to look at the uneven adoption of Gothic features during the period of transition

---

³ A list of the catalogues is available on the website of the *Comité international de paléographie latine*, <http://www.palaeographia.org/cipl/cmd.htm>


⁶ Consider, for instance, the use of Insular a by Scribe 16 of the *Necrologium Lundense*, <notendur.hi.is/mjm7/?76,74,71,66,63,62,61,56>
between Carolingian script and Gothic script during the twelfth- and thirteenth-centuries. Scandinavia did not simply wake up to a world dominated by the two-storey closed ‘a’. Derolez calls the ‘Pregothic’ transition script “hardly a script type in itself.” Rather, it is “in fact Carolingian script that displays to a greater or less extent one or more of the new features. These features would only be present all together in the fully developed Gothic Textualis.” The Scandinavian corpus is in a unique position: due to its late Christianization, it contains very few examples of full-fledged Carolingian. Thus, looking at Scandinavia as a whole, we have a script culture which varies on multiple axes: East versus West, Insular versus Continental, and Carolingian versus Gothic. There is also cursivity to consider.

Section 1.2: The Lund scriptorium and the *Necrologium Lundense*

Twelfth-century Lund is the perfect scriptorium to discover this fantastic mess of a brain teaser. Established in 1104, the manuscript output of the archdiocese of Lund is unmatched in Scandinavia during the early 12th century. If Lund is the perfect candidate for this period, then the *Necrologium Lundense* (in particular, the *Memoriale Fratrum*) is perhaps the best case study. Containing over eighty hands, the manuscript was used during the office in chapter after Prime to, among other things, pray for the souls of deceased brothers and sisters as well as for individuals who have financially supported the cathedral chapter. A Roman calendar structured as a martyrology stretches thirty-nine leaves. The dates for some entries are full to the brim and the writing extends far into the margins, while others are blank. The two principle studies of the manuscript are contained in Weibull’s 1923 diplomatic edition and Kroman’s 1960 facsimile

---

7 Derolez, 57.
edition.\(^8\) Weibull’s diplomatic edition was used as a base text for my digital edition, though I made a number of revisions to Weibull’s text to make it suitable for NLO: a different transcription scheme was employed, abbreviations on the diplomatic level were marked, and a facsimile level was added (see Section 2.1 below). Kroman, following Weibull, provides a chart of all the scribal hands on every page, which I used for my edition to mark hand shifts.\(^9\)

The necrology shows collaboration between many different scribes over a long period of time (mainly 1123-1170). Thus, the necrology could be considered a sort of microcosm of Scandinavian script culture at the time. The fact that the majority of the scribal hands span a period of about fifty years should also give us an image of script in transition. This is the intention of *Necrologium Lundense Online*.

**Section 1.3: Digital palaeography, DigiPal’s formal model, and NLO’s interpretation of it**

But why make a new edition? Why not just write a paleographical study of the manuscript using Weibull’s 1923 diplomatic edition and Kroman’s 1960 facsimile edition? If a more traditional path is pursued, the question then becomes: what kind of palaeographical study? The field of palaeography has changed considerably over the last years due to the convergence of two research trends: 1) the emergence of digital and statistical methods and 2) what has been called the ‘crisis of palaeography’. On the latter, Derolez puts it very succinctly:

> When an extremely experienced palaeographer declares that a manuscript was written in Northern France in the first half of the thirteenth century, but fails to indicate the criteria on which this

---


\(^9\) Kroman, XIX-XXVI.
statement is based, he may be a perfect connoisseur, but he is not being an effective teacher. What is more, he unconsciously contributes to the present-day crisis of palaeography as a discipline.¹⁰ Instead, digital means have provided palaeographers with the capability of linking their claims directly to manuscript images, and this is an important consideration when striving for reproducibility. The aim is to become more transparent, with an eye also to the pedagogical benefits that come with easier access to the inner workings of palaeographical argument.

For NLO, it just made sense to implement one or more of the new methodological paradigms in the field of digital palaeography. I drew a great deal of inspiration from the Digital Resource and Database of Palaeography, Manuscript Studies and Diplomatic (DigiPal),¹¹ a digital implementation of a palaeographical database which “set out to offer a visible and replicable methodology to allow people to explore palaeographical data and to communicate their evidence and accompanying argument to a wider audience.”¹² Its formal model attempts to represent in precise terms the relationships between scripts, manuscripts, scribes, graphs, and handwriting features. The formal model is represented by the Universal Modelling Language (UML) diagram in Figure 1. Starting at the far left, a manuscript contains any number of parts, which contain any number of pages. These pages contain any number of graphs, which are manifestations of a particular “idiograph” (a particular scribe’s representative rendering). An idiograph is an instance of an allograph, and multiple allographs can represent one character (for instance, ‘s’ can be manifested as straight ‘s’ or round ‘s’).¹³ Scribal hands are individual instances of a scribe’s writing (thus similar in concept to a scribal “stint”), and scribal hands are

¹⁰ Derolez, 2.
written in one or more scripts. The graphs on the page usually exhibit one on or more ‘features’ in one or more of their ‘components’. For instance, a ‘hook’ (component) of a straight ‘s’ (allograph) might be ‘trailing’ (feature). Alternatively, there may be ‘general features’ of a graph such as ‘angular’ or ‘rounded’.

**Figure 1 - UML Class Diagram of the DigiPal formal model**

The formal model developed by DigiPal could certainly be adapted to the Scandinavian context. In fact, it already has. In conjunction with the project, Matilda Watson developed a modified version of the DigiPal framework called ScandiPal to study early Latin manuscript fragments in Norway and Sweden thought to be written by Scandinavians and to contain evidence of insular influence.\(^5\) In the beginning, I intended to do something similar using the

---


\(^{15}\) Brookes, et al., 43.
DigiPal framework. I thus began by hosting a self-contained image of the framework with the Docker application and began uploading manuscript images and playing around with allograph creation in their convenient Mezzanine interface.\footnote{DigiPal: Digital Resource and Database of Manuscripts, Palaeography and Diplomatic (London, 2011–14). Software, available at, \url{https://hub.docker.com/r/gnoelddh/digipal/}.} In the end, however, I chose to write my own website, reproducing much of the functionality (albeit in a condensed manner and with less attention to writing production-quality code). For starters, I wanted to make my edition available on the web through the University of Iceland’s student hosting service. This would let me publish my edition and receive feedback immediately. A requirement of the hosting service was that all the functionality had to be client-side (purely HTML, JavaScript, and CSS). Since DigiPal runs a server-side database, I would not have been able to publish my edition without hosting (and paying for) a server. Secondly, coding it myself would give me the requisite front-end web development knowledge to eventually work more comfortably with server-side development. Lastly, I wanted to experiment with a modification of DigiPal’s formal model (though this is of course possible to do by an experienced developer editing DigiPal’s source code) which emphasizes statistical palaeographical analysis over the curation of idiographs.

Since this method of handwriting analysis is fairly new, it would benefit from some experimentation. I wanted to play around with the idea of a ‘palaeographical edition’, one which fully reproduces the text of the source and creates links to the manuscript on the word-level rather than the character-level. The aim is to be \textit{decidedly non-critical} about what graphs to select for feature annotation. For my own research, I chose to annotate all the tops of ascenders and minims and every right tip of the hook of ‘r’. There are also other scattered features annotated, meant to demonstrate the breadth of feature annotation. The user is then provided with an ‘Annotation Tool’ and the necessary instructions to annotate the manuscript as they choose.
based on ‘rules’ (see Section 3.7 below). It was essential for the DigiPal project, which looked at
574 scribes and 1675 manuscripts, to create a guideline of four graphs per scribe and thus to
curate representative graphs and annotate them with their particular features.\footnote{Brookes, et al., 38.}
\textit{NLO} makes no
curatorial decision about whether or not a graph is representative of that scribe’s typical
manifestation of an allograph. Thus, \textit{NLO} has no notion of an ‘idiograph’. This allows the
annotation to serve as a basis for quantitative palaeographic queries which take into account the
entire text, and a ‘Statistics’ tab was designed to demonstrate some possibilities for the types of
questions such as “what percentage of Scribe 13’s ascenders are forked?” The statistical
approach to palaeography was chosen in response to Derolez’ endorsement of it:

The quantitative method which should be applied to palaeography is a statistical one. It
consists of counting and measuring significant features of handwriting and charting the
results… By applying statistical methods to palaeography, we will, no doubt, arrive at
important new and objective statements. The method applied hitherto in palaeographical
handbooks has produced an authoritarian discipline, the pertinence of which depends on
the authority of the author and the faith of the reader.\footnote{Derolez, 8-9.}

Future statistical questions may range from “which features tend to occur together in the same
hand?” to “which feature collocations are the strongest, which are the weakest?”

I have also chosen to define more strictly the location of features and to add an additional
object. In DigiPal, features exist at either the allograph- or component-level. General features
such as angularity and pen-angle are captured at the allograph level, while more ‘locatable’
features such as wedges or forks occur at the component level. I prefer to see features such as
angularity and pen-angle as actually existing at certain points in each component. I did this with
the belief that, if a feature occurs in a graph, certainly it must take place at one or more points in a component of that graph (even if it happens to exist at all points in every component of that graph). This would also aid in the annotation process described above which targets only certain points in the Annotation Tool based on ‘rules’ (again, see Section 3.7 below). A set of ‘attachment points’ (borrowing the terminology from animation) were defined for each component. These are locations on an individual component which you can easily point out. For instance, an ascender might have a ‘top’, ‘body’, and ‘foot’. Originally, a large set of these points were defined, but over time they were condensed to very few: body, top, foot, joint, right tip, and left tip. Top, foot, right tip, and left tip correspond to the different ways strokes begin or end. Often these points display such features as ‘approach stroke’, ‘wedged’, ‘forked’, ‘angled up’, ‘trailing’, etc. A joint is a beginning of a component which overlaps with another component, such as the point where a scribe producing the arch of ‘h’ transitions from the shoulder of the arch to the downstroke of the arch. Joints contain features such as ‘blobby’ or ‘through’. The body holds any feature relating to the main part of the line which occurs between the initial and the final strokes or before and after a transition from another stroke. Bodies contain the largest variety of features. Examples of features in bodies are: ‘broken’, ‘angular’, ‘blobby’, ‘forward-leaning’, ‘shaded’, etc.

Furthermore, instead of describing subcomponents of components, I chose to make the ‘component’ level the most broken down level. For instance, in DigiPal, an ‘o’ contains a bowl which contains a south-west quadrant. In NLO, an ‘o’ contains a right bowl component and a left bowl component. A supercomponent is then described which matches at any of its constituents. So the supercomponent ‘bowl’ will match features at either a right or left bowl component. This allows for easy addition of new supercomponents. We can narrow the supercomponent ‘bowl’ to
‘right bowl’ for all rightward-facing bowls such as in ‘b’ and ‘p’ or ‘left bowl’ for all leftward-facing bowls such as in ‘q’ and ‘a’. We can also create supercomponents which apply only to a limited set of allographs. For instance, an ‘f’ contains a tongue and a hook while an ‘e’ and an ‘e caudata’ also each contain a tongue and a hook. A supercomponent ‘eye’ can be defined which matches features at either tongues or hooks only in the ‘e’ and ‘e caudata’ allographs, since the hook and tongue in the allograph ‘f’ do not constitute an ‘eye’. The flexibility allows us to create very broad supercomponents such as ‘horizontal strokes’ or ‘vertical strokes’ without tweaking the content of individual components.

A similar approach is taken to features. Features such as ‘forked’ and ‘wedged’ constitute a superfeature ‘decorated top’. In the course of my research it is my intention to group features using this functionality along the axes ‘Gothic’ or ‘Carolingian’ and ‘Insular’ or ‘Continental’, so that a corpus-wide search could be conducted instantly with a modifiable list of features.19

An additional tool was created in the ‘Compare Hands’ tab in order to create a web version of a common practice of palaeographers which involves comparing each allograph of scribes visually in order to determine the identity of the scribe. This is manually done by examining a number of examples of every allograph in a scribe’s inventory and visually analyzing them with other scribes. I implemented a tool to accomplish this with greater ease, based on Tom Davis’ description of a tool he had once implemented (see Section 3.5).20

Section 1.4: Digital editorial theory in general.

In contrast to standard print editions, digital editions are never published in a final form. They can be more or less freely modified to add additional levels of annotation, additional

19 Currently, no user interface has been implemented for the creation of new supercomponents or superfeatures, even though the interface exists to create new features. They must be edited in the source code.
20 Davis, 275.
searchability, or some other functionality. For instance, aside from the palaeographical interest of
the necrology, it is also a very important text for name and place research. While I have not
dedicated the time to implement it yet, these names and places could be marked up with
additional details including relationships between individuals and links to GPS coordinates.

Digital editions also have the capacity to be epistemologically more satisfying, too.
Editions of the type described here provide the user with the tools to explore the data themselves
in a non-linear manner, while also producing research which essentially provides newer users
with a guide of how experienced users are making use of the edition. All of this, while also being
able to produce links to reproducible collections linked to the original images, a durable and
easily-navigable concordance of the evidence of one’s argument. As Sperberg-McQueen writes:

We can distinguish, with the aid of the other languages, several types of knowledge, of which two
are relevant here: on the one hand, a knowledge of facts normally transmissible by language
(German Wissen), and on the other the knowledge of how to do something (German Können).\footnote{21}

With respect to palaeographical knowledge, “Wissen” would be the knowledge we obtain
through standard works of palaeography, while “Können” would be what we acquire over time
as we make our own studies. NLO tries to achieve a capacity for both types of knowledge to
flourish.

Section 1.5: Concluding remarks and an overview of the structure of Sections 2 and 3

It is my impression that scholars still have a lot to learn about how to do palaeography,
identify scribes, schools, and their relationships, and – especially – write script history. There are
higher levels of granularity still to be achieved in describing handwriting than has been

\footnote{21} C. M. Sperberg-McQueen, \textit{How to teach your edition how to swim} In \textit{Lit Linguist Computing} (2009) 24 (1), 30a.
attempted here. I hope that the formal model described above proves flexible enough to handle multiple scripts in their fully-realized manifestations and also transitional scripts. It must also be strict enough that the features are reproducible from one palaeographer to the next, from teacher to student, and from generation to generation. The durability and validity of palaeographical knowledge is at stake.

What follows is a technical description of the implementation in two sections. First, the pre-processing implementation is explained, and this includes the use of Optical Character Recognition (OCR), the transcription method, the transformation of the transcription into an array, the structure of the array, and the marking up of individual word coordinates for HTML image maps and for ImageMagick batch image cropping. Second, the front-end implementation is explained, and this includes the overall structure of the HTML and CSS, the functions relating to the Text tab, the functions relating to the Search tab, the functions relating to the Collections tab, the functions relating to the Compare Hands, the functions relating to the Statistics tab, and the functions relating to the Annotation Tool.

**Section 2: Pre-processing implementation**

**Section 2.1: OCR, Post-correction, and transcription method**

I began by taking scanning Weibull’s diplomatic edition and ran Google’s OCR software Tesseract to extract the diplomatic transcription from the images. The recognized text was around 75% accurate. Post-correction was performed manually in Apache OpenOffice. I devised my own shorthand to add information such as line breaks, page breaks, facsimile-level transcriptions, margin and superscript, and editorial interventions. The result is an XML-like document, except with no XML tags, which looks like this:
My transcription policy was to make the diplomatic level a ‘character-level’ transcription, whereas the facsimile transcription would be ‘allographic’, and additionally any abbreviations are expanded. Appendix A contains the mappings of allographs and their characters. Therefore, round ‘s’ and tall ‘s’ would be transcribed simply as ‘s’ on the diplomatic level, and insular ‘a’ and ‘a’ would be transcribed as ‘a’. This allows the user to search on the diplomatic level when they wish to search on the character-level rather than the allograph-level. A full list of the defined character-allograph mappings is available in the “globals.js” file in the source code, stored in the variable CHARACTER_MAP.

Section 2.2: Transforming the transcription to an array and the array structure.  

The intention here was to write a system which was less cumbersome to transcribe in than in pure XML and then transform it to XML. The XML could then be transformed via XSLT to an array structure. In the end, I abandoned the XML and XSLT method and wrote a script in Javascript to convert the shorthand directly to an array structure which closely resembles the

---

22 The raw transcription text is available online at <https://notendur.hi.is/mjm7/Weibull_Transcription_Post.odt>
23 The array transformation tool described here is available at <https://notendur.hi.is/mjm7/array_transformation_tool.zip>
24 Since XML is the most interoperable format, the transcription contains all the information needed to convert to valid MENOTA XML should the need arise.
array structure found in *Original Icelandic Charters Online*. The transcription described above in Section 2.1 was processed using Javascript into a series of nested arrays in an array called CORPUS. The array contains all the information related to the transcription (except for the abbreviation marker tags corresponding to `<am>` in MENOTA-XML which I did not incorporate into the web edition but which is still captured in the transcription to ensure compatibility).

The array CORPUS is a nested array consisting of all the information about the text. The highest level nested array is an array of two indices, the first index is the folio number, for instance “124v”. The second index is an array containing all the transcribed text on that page:

```
[[page1, [[element1,...][element2,...]]],[page2,... , [[elements...]],[...]]
```

The transcribed text is also structured as a series of nested array. The first index of these arrays indicates its “Type”. The varieties of ‘Type’ and their contents are as follows:

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>“w”</td>
<td>A word</td>
<td>Image Id, Facs-level transcription, dipl-level transcription, coordinates, folio and line number, scribe, scribal stint, feature array.</td>
</tr>
<tr>
<td>“wl”</td>
<td>A word that splits across a line or page</td>
<td>Image Id, Facs-level transcription, dipl-level</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>“c”</td>
<td>An editorial insert</td>
<td>Subtype “o”, “c”, or “w”. “O” and “w” open and close the insert, “w” is a word.</td>
</tr>
<tr>
<td>“p”</td>
<td>A punctuation mark</td>
<td>Image Id, facs-level transcription, dipl-level transcription, coordinates, folio and line number, scribe, scribal stint</td>
</tr>
<tr>
<td>“l”</td>
<td>A line break</td>
<td>Line number (generated automatically)</td>
</tr>
<tr>
<td>“s”</td>
<td>Supralineal text</td>
<td>Subtype “o” to open and “c” to close</td>
</tr>
<tr>
<td>“m”</td>
<td>Marginal text</td>
<td>Subtype “o” to open and “c” to close</td>
</tr>
</tbody>
</table>

The functionality of the website is built around looping over this array, finding the relevant data, and displaying it in some manner.
Section 2.3: Text-image alignment and batch image cropping\textsuperscript{26}

To study palaeography in the ‘Search’ and ‘Collections’ windows, it was necessary to include cut-out images of words. Text-image alignment was thus performed on the text. I used the simple free webtool \textit{Image Maps Tool}\textsuperscript{27} to draw the areas and manually entered the image IDs of each word and punctuation mark (see Figure 2). The output, an HTML image map, was parsed in the array transformation algorithm described in Section 2.2. The result was that the

\footnotesize{\textsuperscript{26} The batch image processing tool is available online at <https://notendur.hi.is/mjm7/image_magick_bat_tool.zip>}

\footnotesize{\textsuperscript{27} <image-maps.com>
coordinates found in the image map were inserted into the array. A second script was written to parse the HTML image-maps into .bat files which could batch crop the image files using the ImageMagick library.28

Section 3: Front-End implementation

Section 3.1: Overall structure of the web page

The website was implemented as a single Javascript page following Icelandic Original Charters Online (IOCO). Almost all of the CSS was taken directly from IOCO, and so was a lot of the HTML. Navigation between the ‘pages’ of the website is implemented using the display:none CSS rule. When the user navigates to another page, the <div> containing the previous page is hidden, and the destination <div> is shown. There are thus three main files, index.html, main.css, and nlo.js. There are also several libraries including JQuery. The Images folder contains the large manuscript images and the Images/small folder contains the individual word images.

Section 3.2: Functions relating to the Text tab

The Text tab is used to present the individual manuscript images alongside their transcriptions. All of the information about an individual word can be accessed by clicking on the word or on the image map area on the manuscript image. There are thus three main components to the Text tab: first pertaining to the layout and display of the image and text, second to the generation of the image map, and lastly to the display of the NoteBuffer which contains additional information about individual words.

When the user selects the folio number and transcription level, a function grabs the transcription element from the CORPUS array for that particular folio number. The image is displayed and the function proceeds through the transcription, adding the line numbers, words, words that break over lines, and appropriate editorial markers. An onclick event handler is added to each word so that when the user clicks the word, it displays a window with all the information about that word and a link to Perseus Latin Word Study Tool (though in many cases Perseus will not find the word, since Perseus requires normalized classical Latin). An identical onclick event handler is added to the &lt;area&gt; in the image map. The contents of this window are generated with a function which simply renders the contents of the word’s array as a table.

3.3: Functions relating to the Search tab

The Search tab is used to filter the text based on user input. Users can filter based on a text string (based on the diplomatic or the facsimile transcription, using RegEx or not, ignoring the expanded text or not if on the diplomatic level, and case insensitive or not), included or excluded scribes, by allograph, component, attachment point, feature, or any combination of the above. The results are displayed in a table. Each cut-out image is linked to the full-size image and each word contains a link to a window which contains the palaeographical annotation of that word.

There are some quirks to the search functionality which bear some explanation. First, the user cannot search for a text string on the diplomatic level and also specify an allograph, component, point, or feature. This is because, as explained above in Section 2.4, the allographs in the feature annotation are tied to the facsimile-level transcription, not the diplomatic-level transcription. So it is theoretically impossible for the engine to know if the letters which constituted the match correspond to the facsimile-level letters desired. This would be possible if
the diplomatic-level characters were mapped to the facsimile-level allographs, but this was beyond the scope of my project. Conversely, if the user searches for a text string on the facsimile level and also specifies an allograph, component, point, or feature, the search function applies the allograph filter *only to those indices which are part of the RegEx match*. For users who want to return multiple allographs of the same character (for instance want both Caroline and Insular ‘a’ to appear in their search results), this can be achieved on the facsimile level using a Regular Expression such as “[^qá].”

**Section 3.4: Functions relating to the Collections tab**

Following DigiPal, I chose to implement a ‘Collection’ functionality so that a user could collect together a number of graphs they think are interesting for some purpose and generate a link to them which they can then share, for instance in a footnote. Throughout the other tabs an ‘Add to Collection’ button exists for each word. This button adds the individual word and all of its information to a global variable. When the user navigates to the Collections tab, a table of all of these words is generated from the global variable. It is possible to select and remove individual words, to clear the entire collection, or to generate a link to this specific collection. The link is a URL with a query string added on with the individual IDs of each word in the collection. When the page is first loaded, the page checks if the user specified a query string in the URL. If they have, the global variable is populated from the query string and the page goes directly to the Collections tab and loads that collection.

**Section 3.5: Functions relating to the Compare Hands tab**
This tool is a more dynamic implementation of the tool described by Tom Davis.\textsuperscript{29} The user begins by specifying how many columns they want the tool to display and the maximum number of images for each allograph which they are interested in. Selection boxes are populated in the first row over the specified number of columns, containing a list of the scribes. When a scribe is selected, a selection of that scribe’s allograph inventory is generated. Below it, images up to the maximum number are displayed which contain that the chosen allograph.

Section 3.6: Functions relating to the Statistics Tab

The statistics tab utilizes the existing search function, executing a large number of searches and summarizing them in a series of tables for statistical research purposes. Statistics are currently only possible for the entire corpus or by individual scribe, though other ranges are certainly possible. Currently, six different statistics are implemented:

1) List of scribal stints and ranges: A dynamically generated list of all the scribe’s stints and their ranges in folio and line numbers.

2) Word, punctuation, and character count

3) Allograph frequency by character: Looping through the CHARACTER_MAP global variable, this statistic counts the number of instances of each allograph and calculates the percentage of how often that allograph is used to represent the character it is mapped to.

4) Feature frequency by allograph: Conducting a large number of searches, this statistic loops through every single allograph, their components, and their points and counts the number of instances of features which occur.

\textsuperscript{29} Davis, 275.
5) Feature frequency by component: Same as above, but with the component as the highest level category.

6) Feature frequency by point: Same as above, but with the point as the highest level category.

**Section 3.7: The annotation tool**

The annotation tool allows users to edit the palaeographical annotation of the corpus. Clicking on the tab brings the user to the ‘Annotation Tool Settings’ field set. Rules can be created, rules for which fields the user is interested in annotating. Existing rules are displayed and can be deleted. The features which the user wishes to mark up are displayed in the Enabled Features select box. By selecting a different point, the select boxes are repopulated to only show features which are pertinent to that point. Features can be enabled and disabled using the corresponding buttons. New features can also be created for individual points. These features will then be available for annotation and also displayed in the Search and Statistics tabs. Users can also edit the maximum number of images results here (discussed below). Finally, the “Backup Changes to Corpus” button creates a text field containing all the variables which the user edits in the course of using the annotation tool. To save and return to their version of the annotation, the user can download a full copy of the website to their local storage, open the “corpus.js” file, select all, and paste in the text generated by the “Backup Changes to Corpus” button.

When a user begins annotating, the annotation tool proper is displayed. The tool searches the corpus for the next match based on the rules provided, starting at the beginning of the corpus unless the user entered the annotation process via the “Enter Annotation Tool from here” option available in the Text tab. Previous and Next buttons are provided for navigation through the
corpus together with the image of the current word, its facsimile-level transcription, and the individual allograph bolded. A selection box is populated for each point containing the enabled features for that point. When the contents of this box are changed, the tool displays images and transcriptions with the appropriate highlighting which already match the selected feature. This is designed for the annotator to receive reminders of what he or she has already determined this feature looks like and can help the annotator determine whether, for instance, an ascender is ‘forked’ or ‘shallowly forked’ based on their own visual criteria. Finally, the features currently stored at this point are displayed, with the option to delete them.
Appendix A – Allographs and Their Components

The following table presents a list of the allographs used in NLO and their components. An image is provided which highlights the individual components. The image is, of course, only an abstract representation of an allograph, which only ever manifests itself as a graph on a page. Components are guidelines, and they do not always correspond so neatly to these abstract images. The images are offered strictly for reference. The red lines are the headline and baseline. I have only included minuscule allographs in the table. Small capitals, majuscules, abbreviation marks, and punctuation marks have been omitted.

<table>
<thead>
<tr>
<th>Image</th>
<th>Character</th>
<th>Allograph</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>A</td>
<td>a</td>
<td>Bowl left upper curve, bowl right upper curve, back, hook left</td>
</tr>
<tr>
<td>a</td>
<td>A</td>
<td>a</td>
<td>Bowl left upper curve, bowl right upper curve, back</td>
</tr>
<tr>
<td>b</td>
<td>B</td>
<td>b</td>
<td>Ascender, bowl right upper curve, bowl right lower curve</td>
</tr>
<tr>
<td>b</td>
<td>B</td>
<td>b̄</td>
<td>Ascender, bowl right upper curve, bowl right lower curve, tongue</td>
</tr>
<tr>
<td>Upper case</td>
<td>Lower case</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>c</td>
<td>Upper curve c, lower curve c</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>d</td>
<td>Bowl left upper curve, bowl left lower curve, ascender</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Ȗ</td>
<td>Bowl left upper curve, bowl left lower curve, ascending back d</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>e</td>
<td>Lower curve e, hook right, tongue</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>ɛ</td>
<td>Lower curve e, hook right, tongue, caudata</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>f</td>
<td>Ascender, hook right, tongue</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>g</td>
<td>Back, tail g, bowl left upper curve, bowl left lower curve</td>
<td></td>
</tr>
<tr>
<td>Uppercase</td>
<td>Lowercase</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>h</td>
<td>Ascender, shoulder, downstroke</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>h</td>
<td>Ascender, shoulder, downstroke</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>i</td>
<td>Minim</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>í</td>
<td>Minim, acute accent</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>j</td>
<td>Descender</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>k</td>
<td>Ascender, upper branch, lower branch k</td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>Ascender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>Ascender, tongue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>Minim, middle, shoulder m, middle, downstroke m, final, shoulder m, final, downstroke m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ω</td>
<td>Upstroke unc m, middle shoulder m, middle downstroke m, final shoulder m, final downstroke m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Minim, shoulder, downstroke</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>o</td>
<td>Bowl left curve o, bowl right curve o</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>oφ</td>
<td>Bowl left curve o, bowl right curve o, loop_oe</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>ø</td>
<td>Bowl left curve o, bowl right curve o, slash o</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>p</td>
<td>Descender, bowl right upper curve, bowl right lower curve</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>p</td>
<td>Descender, bowl right upper curve, bowl right lower curve</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>q</td>
<td>Descender, bowl left upper curve, bowl left lower curve</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>r</td>
<td>Minim, hook right</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>2</td>
<td>Upper curve r rotunda, bottom stroke r rotunda</td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>----</td>
<td>----------------------------------</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>s</td>
<td>Upper curve round s, lower curve round s</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>f</td>
<td>Hook right, downstroke</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>c</td>
<td>Topstroke, lower curve insular t</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>t</td>
<td>Topstroke, downstroke</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>u</td>
<td>Downstroke, shoulder u, minim</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>v</td>
<td>Downstroke v, right component v</td>
</tr>
<tr>
<td>W</td>
<td>W</td>
<td>w</td>
<td>Left component w, right component w</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>x</td>
<td>Northwest branch x, southwest branch x, northeast branch x, southeast branch x</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>y</td>
<td>Upper left branch y, right to left stroke y</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>ÿ</td>
<td>Upper left branch y, right to left stroke y, dot</td>
</tr>
<tr>
<td>J</td>
<td>J</td>
<td>J</td>
<td>Topstroke, downstroke et</td>
</tr>
<tr>
<td>J</td>
<td>J</td>
<td>z̄</td>
<td>Topstroke, downstroke et, tongue</td>
</tr>
<tr>
<td>J</td>
<td>J</td>
<td>&amp;</td>
<td>Left component amp, right component amp</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>Upper curve r rotunda, bottom stroke r rotunda, downstroke rum</td>
</tr>
</tbody>
</table>
Bibliography


Hreinn Benediktsson, *Early Icelandic script as illustrated in vernacular texts from the and thirteenth centuries.* Reykjavík, 1965.


— *Palæografisk atlas, dansk afdeling*. København: Gyldendal, 1907.


