



Open Library of Humanities



Part of the Ubiquity  
Partner Network



---

## Research

**How to Cite:** Dahllöf, Mats. 2018. "Automatic Scribe Attribution for Medieval Manuscripts." *Digital Medievalist* 11(1): 6, pp. 1–26. DOI: <https://doi.org/10.16995/dm.67>

**Published:** 24 December 2018

---

## Peer Review:

This is a peer-reviewed article in *Digital Medievalist*, a journal published by the Open Library of Humanities.

## Copyright:

© 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

## Open Access:

*Digital Medievalist* is a peer-reviewed open access journal.

## Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

---

## RESEARCH

# Automatic Scribe Attribution for Medieval Manuscripts

Mats Dahllöf

Uppsala Universitet Uppsala, SE

[mats.dahllof@lingfil.uu.se](mailto:mats.dahllof@lingfil.uu.se)

---

We propose an automatic method for attributing manuscript pages to scribes. The system uses digital images as published by libraries. The attribution process involves extracting from each query page approximately letter-size components. This is done by means of binarization (ink-background separation), connected component labelling, and further segmentation, guided by the estimated typical stroke width. Components are extracted in the same way from the pages of known scribal origin. This allows us to assign a scribe to each query component by means of nearest-neighbour classification. Distance (dissimilarity) between components is modelled by simple features capturing the distribution of ink in the bounding box defined by the component, together with Euclidean distance. The set of component-level scribe attributions, which typically includes hundreds of components for a page, is then used to predict the page scribe by means of a voting procedure. The scribe who receives the largest number of votes from the 120 strongest component attributions is proposed as its scribe. The scribe attribution process allows the argument behind an attribution to be visualized for a human reader. The writing components of the query page are exhibited along with the matching components of the known pages. This report is thus open to inspection and analysis using the methods and intuitions of traditional palaeography. The present system was evaluated on a data set covering 46 medieval scribes, writing in Carolingian minuscule, Bastarda, and a few other scripts. The system achieved a mean top-1 accuracy of 98.3% as regards the first scribe proposed for each page, when the labelled data comprised one randomly selected page from each scribe and nine unseen pages for each scribe were to be attributed in the validation procedure. The experiment was repeated 50 times to even out random variation effects.

---

**Keywords:** palaeography; manuscript; handwriting; scribe; scribe attribution; letter; classification; digital palaeography; digital humanities

---

## 1 Motivation and purpose

In 1970, workers involved in the restoration of a chapel in Speyer found a reliquary containing a very old manuscript leaf. Experts called in to examine the item were—we can imagine—excited to find writing in gold and silver ink on purple vellum using a somewhat odd alphabet. These details must soon have led their inquiries in the direction of the evangeliary known as Codex Argenteus, which, after dramatic travels, had ended up in Uppsala. Philologists could later definitely verify that the Speyer leaf was one of those missing from the codex. There are a number of circumstances, e.g. the Gothic language and alphabet, the extraordinary design, and the textual content, which strongly speak in favour of the conclusion that the solitary leaf belongs to the evangeliary.

If we are interested in finding the manuscript home of an odd medieval fragment of a more ordinary kind, say a piece of parchment with Latin text written in some cursive script typical of the 15th century, we face a difficult problem. Just to see that two pages are from the same codex, scribe, or cultural context, and to justify such a conclusion, requires the expertise of a palaeographer. Browsing thousands of 15th century manuscripts, one-by-one, library-by-library, to compare them with an enigmatic leaf, would necessitate enormous efforts, even if the most efficient of philologists would contribute their competence to the project.

Today many libraries are in the process of digitizing their historical collections. This gives us new opportunities to compare manuscripts and to find new connections among them. A modern expert trying to place an odd medieval page in its context of production would most likely use a computer, at least for viewing digitized manuscripts. This article will be concerned with using computational resources to compare parts of manuscripts. To be more specific, it will focus on automatic scribe attribution.

The main purpose of the system described here is to predict, by means of automatic analysis of digital images, which scribe has produced the writing on a manuscript sample. This is essentially a classification problem. In this context, each scribe (or we could say class) is identified as the hand behind a set of given manuscript

images. The system has access to a set of writing examples which constitutes a database of known scribes.

A secondary purpose of the present system is to produce arguments for scribe attributions which are comprehensible to a traditional palaeographer or even an ordinary human reader. This means that the classification procedure must follow a series of steps from which we can derive a presentation of the evidence which is compatible with this purpose. The central idea is that we can justify scribe attributions by highlighting similarities between the letters of the manuscript under examination and letters produced by scribes from the database. The system is consequently in the vein of “digital palaeography” (Ciula 2005) in its wish to contribute to methods in manuscript analysis which are quantitative and amenable to objective validation and, at the same time, support philologically meaningful reasoning and visualization.

In connection with this study, we have compiled and published open-source a data set comprising 46 medieval scribes writing in book hand scripts (see Appendix for details).

## 2 Previous studies

Knowing who has produced a manuscript is of obvious relevance in disciplines like history, literary studies, and philology. In traditional palaeography (as defined in e.g. Aussems and Brink 2009), scribe attribution has to a large extent relied on qualitative analysis. Fundamental properties include the morphology of the script and the execution of the writing, where ductus, speed, and care are three aspects. Examination of what is called the “graphical chain” (Stutzmann 2016) focuses on how characters appear in the context of writing, e.g. on how allographs are distributed, and how scribes connect letters. Linguistic features, such as spelling, including the use of abbreviations, are also relevant as qualitative evidence for scribe identification. Research in palaeography has increasingly come to rely on more formalized criteria and quantitative evidence, such as letter widths, heights, distances, and angles. This development of the field has been described as a move from an “art of seeing” to an “art of measurement” (see Stansbury 2009 for a discussion). Systematic and extensive measurement of script features is hardly practically possible without the use of

digital tools. This means that research on quantitative methods has clustered in a discipline of “digital palaeography” (Ciula 2005). In addition to palaeography, there is also a more recent area of expertise concerned with modern handwriting, which is strongly associated with forensic sciences. It finds its main application in criminal and civil cases, where the origin and authenticity of documents are to be verified. Forensic handwriting scholarship and palaeography have developed as two more or less independent academic fields.

The high costs of non-digital approaches to scribe attribution—or more commonly, “writer identification,” in technical contexts—have motivated researchers to study automatic scribe attribution for both historical and modern documents. The challenging nature of the problem from the point of view of image analysis has also stimulated academic attention. Computational research on modern handwriting overlaps with forensic science, whereas work on historical data belongs to the field of “digital palaeography.” Closely related problems which can also be assigned to this area are script classification (Stutzmann 2016; Cloppet et al. 2018), manuscript dating, and fragment rejoining (Wolf et al. 2010). They are of particular relevance for historical manuscripts and these tasks to a large extent face the same difficulties and have to use the same kinds of method as automatic scribe attribution.

Most scribe attribution systems for historical manuscripts are based on machine learning and make use of features which can be extracted independently of linguistically informed segmentation and labelling of the writing (Jain and Doerman 2014). Among such features we find, for instance, probability distributions for character fragment contours (Schomaker, Bulacu, and Franke 2004), character fragments as normalized bitmaps, distributions of the orientations of hinged edge fragments (Bulacu and Schomaker 2007), and distributions of stroke fragments (Tang, Wu, and Bu 2013). Another system (Brink 2011) used a “Quill” feature, which models the relation between the local width and direction of ink traces. He, Wiering, and Schomaker (2015), working in the same school, proposed features capturing the distribution of junctions (meetings of strokes). Mixing features relating to texture, shape, and curvature in writer attribution systems has led to improved results (Jain

and Doerman 2014). Feature engineering of this kind has been combined with machine learning techniques such as clustering for the generation of codebooks of recurring writing components, nearest-neighbour classification (Schomaker, Bulacu, and Franke 2004; Bulacu and Schomaker 2007; Brink 2011; Tang, Wu, and Bu 2013), and multi-layer perceptrons (De Stefano et al. 2011).

Feature models working in the fashions described above capture, on a document sample level, the distribution of image details much smaller than letters. This means that the models are difficult to visualize in terms comprehensible from a traditional palaeographic point of view. By contrast, Ciula (2005) and Dahllöf (2014) proposed systems for comparing scripts and scribes letter-by-letter by means of mathematical models of letter similarity. As both systems relied on manual extraction of letters, they did not provide fully automatic tools for manuscript classification. However, they do point in the direction of methods where “the traditional qualitative palaeographic paradigm can be strengthened and assisted by the creation of graphic models that are quantitative in nature,” to quote Ciula (2005). The current work aspires to implement these ideas in a fully automatic system.

Comparing the performance of scribe attribution systems is an intricate task, since different systems target different kinds of writing. Furthermore, evaluation scores for different systems are based on data with varying numbers of writers and different amounts of data available for each writer (Brink 2011). Modern data sets have typically been created in laboratory environments with standardized pens and writing supports. Medieval data on the other hand derive from physical manuscripts which have been created using writing supports, pens, and inks with varying properties. And the storage and use over the centuries have in most cases radically changed the appearance of the writing, or even damaged it. Additions of later writing are also common.

An important metric in validation of scribe attribution systems, and classification systems generally, is the top-1 accuracy score, which considers the highest-ranking prediction for each query item: It is the ratio between the number of true predictions and the total number of predictions. State-of-the-art systems for *modern* handwriting

reach higher performance scores than those reported for medieval data. For instance, He and Schomaker (2016) report a top-1 accuracy score of 93.2% for a data set with 650 hands writing in English. Their overview quotes similar scores for modern Greek, Arabic, and Chinese writing.

In the *ICDAR2017 Competition on Historical Document Writer Identification (Historical-WI)* (Fiel et al. 2017), the participating systems reached top-1 accuracy scores between 47.8% and 76.4% for 720 writers. The data set is said to cover the 13th to 20th century, but no details on the distribution of the documents over time are given.

In their work on medieval handwriting, Brink (2011) reported top-1 accuracy scores in the range 70%–92% for data sets comprising 10–18 scribes. Another approach (De Stefano et al. 2011), relying only on page layout features, with each writing sample consisting of four rows of writing, achieved 92% top-1 accuracy for 12 scribes, all producing Carolingian minuscule writing.

### **3 Scribe attribution procedure**

When given a query example, the current system predicts a scribe selected from a set of individuals, each one defined by labelled manuscript data. The scribe attribution procedure relies on a sequence of processing steps involving two fairly simple classification modules. One of the advantages of this is that the process will use evidence in a way that is comprehensible for palaeographer with a traditional understanding of the task. This means that predictions are reached in a way that corresponds to an argument that can be visualized for the user. Another gain is that the system can be applied without a potentially time-consuming training step, as would typically be necessary when models based on machine-learning are used. The system exists and was evaluated in the form of a Java implementation.

The operation of the system is guided by a set of parameters. Experiments made during the development phase suggested that the parameter setting described below leads to a good performance. It is the one which was used in the evaluation reported below. The parameter values can arguably also be explained and justified from the point of view of an a priori understanding of Latin book hand scripts, even if the

values, admittedly, to some extent are arbitrary. In work with new data, the system invites retuning of the parameter settings.

### ***3.1 Amount of labelled data and amount to be classified***

In each application of the system, the labelled data are a set of images sampling a certain amount of writing for each scribe. This amount can be just a part of an image, one full image, or several images. Different sizes of the query units (to be attributed to a scribe) are also possible. In the experimental rounds of the evaluation reported below, one manuscript image was in all cases the size both of the labelled samples and of the query units. As will be described below, the labelled data were randomly selected from the manuscript data set, and the remaining (unseen by the system) images were used to generate queries in the evaluation procedure. The images of the data set are in the high-resolution state-of-the-art quality forms provided by the libraries and correspond to one page or one spread. (The data set is published open-source, see below.)

### ***3.2 Extraction of image components, mainly letters***

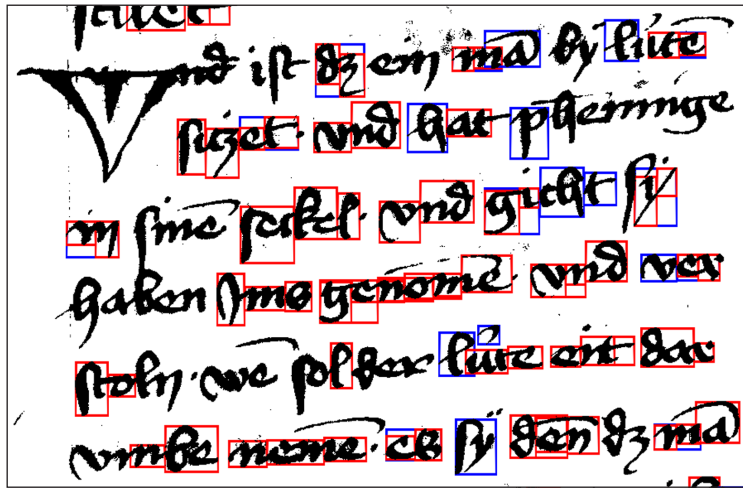
The first processing steps applied to the manuscript files are cropping, which removes the image margins, and scaling. After that, the system will operate on “binarized” versions of the manuscript images. In these, the pixels only carry a binary value indicating writing foreground (ink) versus background (parchment/paper). This is a considerable reduction of the information content of the images, as colour and greyscale information will not be available in the further processing. The binarization is executed by means of a version of the commonly employed Otsu (1979) algorithm. Using binarization is a common practice in handwriting analysis (Brink 2011; Jain and Doerman 2014; He, Wiering, and Schomaker 2015).

The binarized representation allows the system to perform connected component labelling for the purpose of extracting connected regions of ink pixels. These regions, defined as sets of foreground pixels, will typically cover letters and letter sequences. Some of the regions are then further segmented into smaller pieces. The idea behind this is that the segments and a subset of the connected components will correspond to single letters and pairs of connected letters. These image elements will be referred



to as “components”, and they form the primary objects of scribe attribution in the current system. The segmentation process is guided by the estimated typical stroke width,  $W_s$  for each manuscript image. The system estimates  $W_s$  by determining the most common width of sequences of continuous horizontal foreground pixels separated by at least two pixels of background.

Six parameters expressed as products of a constant and  $W_s$  constrain the segmentation process applied to the connected components. Vertical cuts are only proposed where the pixel column sum of ink is thinnest, but not thicker than,  $1.0W_s$  and not closer to another cut than  $3.0W_s$ . Segments between cuts are extracted if their width is between  $3.0W_s$  and  $9.0W_s$  and their height is in the same interval, i.e.  $[3.0W_s, 9.0W_s]$ . This parameter setting, i.e. the six real numbers, (1.0, 3.0, 3.0, 9.0, 3.0, 9.0), represents a heuristic and pragmatic assumption about the relevant script types and is assumed to filter out non-letter connected components, while admitting components useful for the present purpose. **Figure 1** shows an example. Note that the scheme excludes many instances of ⟨i⟩, which are narrower than  $3.0W_p$ . We guess

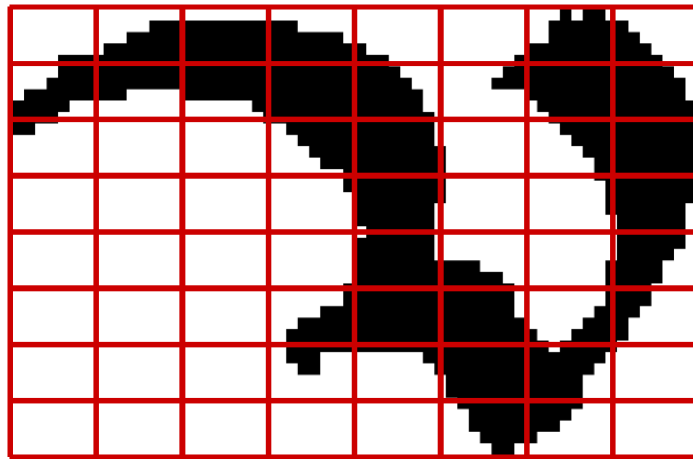


**Figure 1:** Extraction of writing components. This example shows a region from page 105 in Cod. Sang. 726 (hand csg0726B, here, from the St. Gallen Stiftsbibliothek). The page has been binarized and rectangles indicate which image components were extracted. Blue rectangles frame components which were produced directly by the connected component labelling, whereas the red ones were the result of further segmentation.

that (i) components are too “anonymous” to be useful for scribe attribution. The point of using the writing-relative  $W_s$  value in this fashion is to make the system less sensitive to image size and scale. If more than 500 components are retrieved for a scribe, only the 500 ones whose width is closest to the midpoint of the width interval (i.e.  $6.0 w_s$ ) are kept for the later steps of the attribution process.

### **3.3 Feature model and distance (dissimilarity) metric for component comparison**

The shape of the image components is represented by a sequence of numeric measurements (features). In other words, they form coordinates in a feature space. This allows similarity between components to be modelled in such a way that distance corresponds to dissimilarity. The features, which are computed with reference to the minimal bounding box enclosing the foreground pixels, characterize the component in terms of the distribution of foreground (ink) pixels as captured by a grid of  $8 \times 8$  equal subrectangles over the bounding box. This gives us 64 features, as illustrated by **Figure 2**. Each value is the ratio of the



**Figure 2:** The grid corresponding to the features which capture the distribution of foreground (ink). It consists of  $8 \times 8$  equal subrectangles defined in relation to the bounding box enclosing the image component (from Cod. Sang. 983, p. 69). Each value is the ratio of the number of foreground pixels to the subrectangle area. The feature vector would in this case look something like, showing the first eight and last eight values: (0.1, 0.5, 0.7, 0.5, 0.2, 0.1, 0.7, 0.3, ..., 0.0, 0.0, 0.0, 0.0, 0.0, 0.6, 0.3, 0.0), when the image is “read” top-down and left-right.

number of foreground pixels to the subrectangle area, i.e. belongs to the interval  $[0,0,1,0]$ . The concept of distance used is Euclidean distance (with the features given equal weight) computed by this formula (the generalized form of the Pythagorean theorem):  $distance(I, J) = \sqrt{\sum_{i=1}^n (I_i - J_i)^2}$ . So, the distance 0 means that the model does not record any difference between two images, whereas  $8 = \sqrt{64}$  is the maximal dissimilarity.

### **3.4 Scribe attribution for image components**


Using the components extracted from the labelled manuscript images, the system predicts a scribe for each component extracted from a query page or spread by means of “nearest neighbour” classification. This means that each component is assumed to have been produced by the scribe behind the most similar (least distant) labelled component. Each prediction has a strength which is inversely related to the distance between the two components, i.e. the shorter the distance between the query component and the closest labelled component the better. So, for each query image a set of component-level scribe attributions is generated, and these attributions are at the same time ranked on a scale of strength.

### **3.5 Scribe attribution for manuscript samples (pages or spreads)**

The second main module of the attribution process assigns a scribe to each query manuscript sample by means of a voting procedure. This is based on the arrangement of the component-level predictions in ascending order by the distance score, as described above. A scribe prediction for the query image is generated by voting in two steps: First, the (at most) five scribes who receive the largest number of votes from the top 120 component predictions (or all of them if their number is smaller than that) is determined. After that, the system repeats both the classification of image components and the voting with only the labelled components from these five scribes available, again with voting by the top 120 (or all) component predictions. Finally, the scribe who has received the largest number of votes is returned as the prediction for the query image.

### 3.6 Visualizing scribe attribution arguments for a human reader

As the component-level predictions are based on the pairwise similarity of image components, they can be visualized for a human reader in a straightforward way. The example in **Figure 3** shows the 56 best component hits for a query page based on the labelled manuscript data involved in a possible evaluation round (see Section 4). The system creates this overview in the form of an HTML page which can be viewed in any web browser. The scribe coded as csg0990B (see the Appendix for an overview of the scribes in the data set) is clearly getting the majority of the votes so far. The component pairs are arranged in tables, but logically they are only sequentially ranked. The examples which we exhibit here show the top 56 component predictions of the 120 component predictions used to reach a scribe attribution. In each cell, the component from the query example is placed to the left, and the matching labelled component to the right. The distance value (rounded to one decimal) appears below the two components. The foreground (ink according to the binarization) of

 csg0990B:0.8	 csg0990B:0.8	 csg0990B:0.8	 csg0990B:0.9	 csg0990B:0.9	 csg0990A:0.9	 csg0990B:0.9
 csg0990A:0.9	 csg0990B:0.9	 csg0990B:0.9	 csg0990B:0.9	 csg0990B:0.9	 csg0990B:1.0	 csg0990B:1.0
 csg0990B:1.0	 csg0990B:1.0	 csg0990A:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0
 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0
 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0
 csg0990B:1.0	 csg0990B:1.0	 csg0990A:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0
 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.0
 csg0990B:1.0	 csg0990B:1.0	 csg0990B:1.1	 csg0990B:1.1	 csg0990B:1.1	 csg0990B:1.1	 csg0990B:1.1

**Figure 3:** Matched image components. Query components appear to the left and the labelled ones to the right in the cells. Page 325 in the csg0990B sequence was the one under scrutiny. Predictions conforming to the most common decision for the query image, which were true here, are placed in yellow cells and other ones appear with blue background. This outcome consequently strongly spoke in favour of the hypothesis that csg0990B is the scribe.

the components is rendered in dark blue, whereas the rest of the bounding box (background and ink not belonging to the component) appears as in the original manuscript image. Predictions conforming to the most common decision for the query example, which can be true or false, are placed in yellow cells and other ones appear with blue background.

The example in **Figure 3** shows a very clear outcome for the scribe coded as csg0990B (in real life: Elisabeth Schaigenwiler) of the St. Gallen codex Cod. Sang. 990. Page 325 (the fourth) in the csg0990B sequence was queried against all the 16th pages in the scribe sequences, which, in other words, provided the source for the labelled data. The system was specifically asked to generate an attribution based on this data configuration. In the evaluation, the labelled data in each round were randomly selected. The system has proposed matches involving the letters ⟨t⟩, ⟨e⟩, ⟨a⟩, ⟨n⟩, ⟨m⟩, ⟨s⟩, and ⟨r⟩, along with four pairs of two-letter components, ⟨er⟩, ⟨or⟩, ⟨en⟩, and ⟨er⟩. All matches connect graphematically equivalent components. Also note that the four erroneous scribe hits shown in the table point to csg0990A, which is a very similar Bastarda hand, responsible for another unit in the same codex. This scribe, whose name was Dorothea von Hertenstein, worked in the same scriptorium at the same time.

## 4 Performance evaluation

We evaluated the scribe attribution system proposed here by applying it to a data set comprising 46 scribes. As mentioned above, each prediction was based on one image of labelled data for each scribe and one image to be classified. We report the mean top-1 accuracy score and give an overview of which incorrect predictions were made.

### 4.1 Data set

Each of the 46 scribes was represented by 10 manuscript images in the evaluation data set. When it comes to medieval documents, in particular books, scribes can often only be identified through instances of their work. In the present data set, only few of the scribes are known by name. The scribes were selected from digitized manuscripts published by a number of websites: *e-codices—Virtual Manuscript Library of Switzerland*, *ALVIN—Platform for digital collections and digitized cultural*

*heritage*, and the national libraries of Denmark and Sweden. The terms of use for the digitized manuscripts allow that the images be used and distributed for research purposes. (See Appendix for details on the data set, which is included in the one we have published open-source under doi: <https://doi.org/10.5281/zenodo.1202106>.) The images provided by the libraries correspond to one manuscript page in most cases, but some codices are digitized one spread on each image. The set of scribes is the union of three different subsets: The first subset comprises 18 9th century scribes of Carolingian minuscule (language: Latin) taken from the collection of the St. Gallen Stiftsbibliothek, which is the library which has contributed the largest number of manuscripts to *e-codices*. The second subset is also selected from manuscripts belonging to the Stiftsbibliothek and contains the same number of 15th–16th century scribes using scripts classified as Bastarda (languages: Latin, Alemannic, and German). The third set is a collection of 10 Scandinavian 13th–15th century scribes (languages: Old Swedish and Old Norse). Scribes and image sequences for these were selected with the aim of finding continuous sequences of pages filled with fairly clean and well-preserved writing. In many images there are additions of later writing. As can be expected, the amount of writing in each image varies considerably. Let us venture to say that the pages are fairly typical for medieval book manuscripts as regards density and size for the writing and layout. The images were downloaded in their highest resolution version. The files, in JPEG or TIFF format, are between 2MB and 90MB in size. The ground truth scribe attributions as well as information about scripts and dates were taken from statements published by the libraries and compiled from various palaeographical sources. (Details are given in the Appendix.) One disputed case will be discussed below.

During the development and tuning phase another, disjoint, set of pages from the 36 *e-codices* scribes had been used as data in experiments. The first 10 pages in the 36 Cod. Sang. page sequences defined in the Appendix (and included in the data set as published) were used during the development and tuning phase and the following 10 pages provided data for the final evaluation, e.g. pages 147–156 and 157–166, respectively, for the second scribe of Cod. Sang. 186 (csg186B). The

Scandinavian data had not at all been consulted during the system development phase. However, the data set as published contains 20 pages for each of these hands, but only the second half of these sequences were used in the evaluation.

#### **4.2 Experimental design and results**

We designed an experimental set-up to assess the performance of the system using the data set described above. This set-up corresponds to a scenario where images are compared one by one. In each experimental round, one image for each scribe was randomly selected to provide labelled data, i.e. labelled image components were extracted from these 46 images. The remaining  $9 \times 46 = 414$  images, unseen by the system, were used for evaluation. Each experimental round consequently produced 414 predictions, all based on the same labelled data. The experimental procedure was repeated 50 times in order to even out random variation effects.

The images were cropped in such a way that, if  $l$  is the smallest value of the image width,  $w$ , and height,  $h$ , the further processing was concerned with the image in the centred rectangle of size  $(w-0.05l) \times (h-0.05l)$ . Furthermore, each image was rescaled to be processed at a randomly chosen resolution  $\in \{10, 11, 12\}$  pixels/mm. The original resolutions of the images were estimated from images digitized with rulers on the pages or from codex size metadata. (The estimates are recorded with the published data set.) The rescaling—with few exceptions a downscaling—led to shorter processing times and was motivated by a wish to neutralize possible effects of the original resolution.

The 50 iterations of the experimental procedure of the evaluation required roughly 41 hours on a Windows laptop (processor: intel core i7-4600U @ 2.10 ghz, maximum heap for the Java Virtual Machine: 6.1GB). This corresponds to on average seven seconds for each image query. The current implementation of the system is an experimental one, which is far from optimized efficiency-wise.

This exercise produced 20358 true predictions (out of a total of  $414 \times 50$ ). The system consequently reached a mean top-1 accuracy of 98.3%. We can also look at the scribe attributions for single image components: During the 50 rounds of the experimental procedure, roughly 9.6 million component attributions were made,

on average 464 for each page. For the first step classifications (with all labelled component data available), 4.2 million of these predictions were correct. This gives us 44.0% as the top-1 accuracy for the component-level scribe attribution.

The system allows us to retrieve information about which false predictions were made. This makes it possible to see which images and thereby which hands lead the system to make mistakes. The erroneous predictions are shown in **Table 1**. We can note that 33 of the 46 hands were attributed with 100% top-1 accuracy. The three most often misclassified hands were csg0186B, csg0586, and csg0576. They gave rise to 90, 59, and 51 errors, respectively, in the evaluation rounds (for  $9 \times 50$  attributions). In other words, the system only reached 80%–89% top-1 accuracy for these hands, whereas the overall mean top-1 accuracy was 98.3%.

**Table 1:** The errors produced in the 50 rounds of experimental evaluation.  $9 \times 46 \times 50$  predictions were made, 98.3% of them were true. These are the remaining 342 incorrect ones. The total number of errors for each hand is recorded here, as are the number of specific erroneous predictions.

Hand	Errors	Erroneous predictions
csg0186B	90	csg0926: 30, csg0089: 16, csg0186A: 11, csg0557: 9, csg0078: 8, csg0861: 7, csg0053: 3, csg0569: 3, csg0902: 1, csg0562A: 1, csg0562B: 1
csg0586	59	csg0726B: 32, csg0990A: 13, csg0602: 8, uubC528: 4, csg0593: 1, csg0644: 1
csg0576	51	csg0077: 30, csg0088: 9, csg0078: 7, csg0053: 4, csg0089: 1
csg0562B	47	csg0053: 32, csg0557: 8, csg0562A: 5, csg0078: 2
csg0990A	34	csg0990B: 34
csg0112	26	csg0053: 25, csg0562A: 1
csg0186A	18	csg0569: 17, csg0088: 1
uubB68	9	csg0725: 8, csg0990A: 1
csg0726A	3	csg0726B: 3
csg0089	2	csg0562A: 1, csg0569: 1
csg0557	1	csg0926: 1
csg0562A	1	csg0053: 1
csg0565A	1	csg0077: 1



## 5 Discussion

The evaluation of the system showed that the system performed well on a data set containing both completely new manuscripts (the Scandinavian ones) and unseen images from the same codicological units as those consulted during the tuning of the system. As studies in medieval scribe attribution are few, and the data sets used in evaluations have had different properties, it is not possible to make a fully-fledged comparison of the present system with previous ones, as regards their performance as classifiers in scribe attribution. That said, we can however see that it delivered a mean accuracy score which is higher than the numbers which have been reported for previous experiments with medieval data, which covered smaller sets of scribes. We will also argue below that the errors of the system to a high extent are “reasonable”. An innovative component of the present system is the module that presents evidence for attributions in a way that invites qualitative inspection of the kind promoted by traditional palaeography.

### 5.1 Limitations

Some challenges for the present system should be mentioned: A basic difficulty is that manuscripts on which the binarization module would perform poorly could be difficult to process in the intended way. Defective binarization would interfere with the extraction of writing components. This situation could, for instance, arise for manuscript images with uneven contrast between background and ink, in particular in combination with damages. Low resolution would be a related kind of problem. As these are common and serious troubles for all work with historical manuscripts, they can hardly be seen as indicating specific flaws of the present approach.

Another possible obstacle is that densely connected forms of writing could make it difficult for the component extraction module to find a sufficient number of useful segmentable components. Furthermore, the system is sensitive to rotation of the writing in relation to the digital images. The text lines in the images which have been studied here are roughly parallel with the x-axis. In the evaluation of the system, rotation was consequently not a serious problem. However, some mechanism for correcting image orientation would make the system more robust.

Systems of this kind face many challenges on the path to becoming really useful tools for historians and philologists. One of the most important questions is what happens when the data sets become much larger. The “nearest neighbour” classification is an instance of linear search. The time it takes is proportional to the size of the set of labelled components. This means that some more efficient component classification method will be needed as the data sets grow. Given that the labelled data comprise hundreds of components for each scribe (and each page), it would be possible to estimate which shapes are most strongly distinctive for one or a few scribes, and which ones are more “commonplace”. After that, only the more distinctive shapes would be used as labelled data in the component classification step. This would reduce the time needed for the “nearest neighbour” step and could improve the ability of the system to deal with a larger number of scribes.

The decision to use a size-neutral feature model was guided by a wish to focus on the shape of letters rather than their actual size. (See the discussion of **Figure 5** below for an illustrative example.) This idea is based on the assumption that the personal characteristics of a scribe are likely to be preserved independently of the actual size of the writing. Admittedly, this is a complicated issue, since the size of the writing is likely to have a reciprocal impact on the execution of letters, both as a matter of design intentions and of motoric conditions influencing their shape.

























































### ***5.2 A case subject to different opinions***

There is a disputed case among the manuscripts studied here: In the e-codices “Standard description” for Cod. Sang. 603, Von Scarpatetti (2003) counts, with some hesitation, Hand 2 (csg0603B), “163a–443b, 500a–571b, frakturnahe Bastarda,” and Hand 3 (csg0603C), “446a–499b, sehr charakteristische, eckige Bastarda,” as two different scribes. Mengis (2013, 334) is of the opposite opinion: She holds that these page sequences are produced by one and the same hand (as Von Scarpatetti notes, being aware of Mengis’ then unpublished work). In the data for the evaluation of the current system, Hands 2 and Hands 3 were, following Von Scarpatetti, counted as two different ones. In the evaluation rounds, we saw that the instances of both hands were attributed with 100% top-1 accuracy. (Notice their absence from the

list of errors in **Table 1**.) This means that the system definitely can tell them apart. It also justifies the conclusion that they are “objectively” different. So, if Hand 2 and Hand 3 are from the same individual, she must have changed her writing characteristics in a systematic way from one unit to the next. Nevertheless, if we look at the component attributions we find many Hand 3–Hand 2 associations. This is, for instance, illustrated by **Figure 4**, where page 468 (the fourth) in the Hand 3 (csg0603C) sequence was queried against the 16th pages in the scribe sequences (as in **Figure 3** example). Nine out of the 13 erroneous component attributions we see here are to Hand 2 (csg0603B). Many details of the two hands are strikingly similar, but still the vast majority, 91 (43 of them exhibited in the table) of the 120 component attributions point correctly to Hand 3.

### 5.3 Errors

As mentioned above, **Table 1** gives an overview of the erroneous predictions generated in the evaluation rounds. We can see that the incorrect predictions in all cases attribute the images to hands producing the same script as the correct scribe.

 csg0603C:0.7	 csg0603C:0.7	 csg0603B:0.8	 csg0603C:0.8	 csg0603C:0.8	 csg0603C:0.8	 csg0603C:0.8
 csg0603C:0.8	 csg0603A:0.8	 csg0603B:0.9	 csg0603B:0.9	 csg0603B:0.9	 csg0603C:0.9	 csg0603C:0.9
 csg0603B:0.9	 csg0603C:0.9	 csg0603C:1.0	 csg0603C:1.0	 csg0603C:1.0	 csg0603C:1.0	 csg0603C:1.0
 csg0603C:1.0	 csg0603C:1.0	 csg0603B:1.0	 csg0603C:1.0	 csg0603A:1.0	 csg0603C:1.0	 csg0603C:1.0
 csg0603B:1.0	 csg0603C:1.0	 csg0603C:1.0	 csg0603C:1.0	 csg0603C:1.0	 csg0603C:1.0	 csg0603C:1.0
 csg0603C:1.0	 csg0603C:1.0	 lubbMH17:1.0	 csg0603C:1.0	 csg0603C:1.0	 csg0603C:1.1	 csg0603C:1.1
 csg0603C:1.1	 csg0603C:1.1	 csg0603C:1.1	 csg0603B:1.1	 csg0603C:1.1	 csg0603C:1.1	 csg0603C:1.1
 csg0603C:1.1	 csg0603C:1.1	 csg0603C:1.1	 csg0603B:1.1	 csg0603C:1.1	 csg0603C:1.1	 csg0644:1.1

**Figure 4:** Matched image components for page 468 in the csg0603C hand. The page was attributed correctly (91 component attributions out of 120 support that), but many components were matched with the csg0603B hand. (The table only shows the 56 strongest instances of the 120 component attributions which would decide the verdict on the page.)



**Figure 5:** Matched image components for page 148 in the csg0186B hand. The hand csg0186B is the one which was most often misclassified in the evaluation. For the components that we can see here, the associations stay in the Carolingian minuscule script, and in the same grapheme (sequence), but many point in the direction of an erroneous hand. The three top hands for this page as regarding number of votes received are csg0078 (33 votes, in yellow), csg0186B (29 votes), and csg0569 (22 votes), of a total of 120 votes. (The image only shows the 56 strongest ones.)

So, for instance, the most often misclassified hand, csg0186B, was associated with ten other Carolingian minuscule hands. Similar situations obtain for csg0576 and csg0562B, with 51 and 47 errors, respectively. Again these hands are in Carolingian minuscule and they were consistently attributed to scribes writing in the same script. **Figure 5** illustrates what happened when page 151 of the csg0186B sequence was queried against all the 16th pages in the scribe sequences. (This selection of data could have been part of an evaluation round.) The voting based on the 120 best component matches (of which 56 are shown) gave most support to csg0078 (33 votes) ranking the correct csg0186B in the second place (29 votes). However, the matching of the exhibited components stays within the current script and letter

category: We only see Carolingian minuscule letters of the right grapheme matching the csg0186B components, and one ⟨re⟩–⟨re⟩ coupling. In particular, instances of ⟨s⟩ dominate the picture. **Figure 5** also illustrates the fact that the feature model is neutral with respect to the size of components.

The Bastarda scribe csg0586, the second most often misclassified one, gave rise to 59 errors. The letters of this scribe are connected by thick lines in a way that seem to cause an unusually small number of components to be extracted. This probably contributed to the difficulties. However, we see again that these attributions are to scribes using the same kind of script, i.e. other varieties of Bastarda, and to uubC528, which, like csg0586, is characterized as a cursive script. This suggests that a method similar to the one proposed here could be used to address the task of script classification.

The most common specific incorrect attribution (34 cases) is pages from csg0990A being classified as csg0990B. As mentioned above, the two hands represent very similar Bastarda scripts, and worked in the same scriptorium at the same time. A similar situation can be seen as regards the hand csg0562B. It is striking that it was often, in 32 cases to be precise, attributed to csg0053. According to Von Euw (2008) the Cod. Sang. 562 scribe “gehören wohl zum Kreis um Sintram”, the scribe behind Cod. Sang. 53. The similarity that the system found between the two hands is consequently consistent with previous observations.

## 6 Conclusions

We have outlined and evaluated an automatic system for identifying the most plausible scribe responsible for the writing found in a manuscript image. The set of known scribes was defined by one manuscript image for each hand in the individual experiments we conducted. The central principle of the system is that scribe attribution is performed as a two-step bottom-up classification procedure. First, the system classifies roughly letter-size components by means of “nearest neighbour” classification, based on shape-related similarity. Secondly, the set of component-level attributions, which typically contains hundreds of elements for a page, is used to predict the page scribe by means of a voting procedure. Both the pairings of similar

components and the voting procedure are easy to understand for a user without knowledge about the computational details of the system. This makes it possible to instruct the system to generate a visualized presentation of the evidence for a proposed scribe attribution. This forms a kind of argument which highlights the pairwise similarities between the writing components which were taken to decide the issue. This innovative feature allows the system to provide input to qualitative palaeographic analysis.

The binarization step and the extraction of writing components are motivated by a wish to specifically focus on the writing as ink on the writing support. This idea goes hand in hand with the assumption that in general writing is a matter of a bichrome contrast between ink and background. Notwithstanding, the design of many medieval manuscripts, also several of those in the current data set, makes artful use of several colours. Colours and their distribution also have a lot to tell about the composition of the ink and the writing material, as well as about the way a manuscript has been handled during the centuries. It is certainly possible to exploit this information in a classification system associating pages with codicological units, and it would most likely be useful for the current data. This would however be another task, one of performing codicological unit attribution based on the full range of information available in manuscript images. This problem is worthwhile and interesting in its own right, but it is something else than scribe attribution based on the writing itself as the visible trace of the scribe's performance.

The basic principle of the present system, that of performing scribe attribution bottom-up, classifying details first and derive a verdict on the whole sample from the detail-level attributions, is compatible with further refinement of the modules involved. The binarization module, the component extraction and selection, the feature model, the component classification algorithm, and the voting procedure all invite experimentation with more sophisticated and context-sensitive mechanisms. In particular, we can note that the system treats all writing components in the same way. The examples in **Figures 3–5** illustrate how ⟨e⟩–⟨e⟩, ⟨r⟩–⟨r⟩, ⟨s⟩–⟨s⟩, and ⟨t⟩–⟨t⟩ matches dominate the pictures. This is obviously related to the fact that these letters

are frequent in the data, which reflects their distribution in the languages involved (in these examples mainly Latin and German). Moreover, the segmentation module and the feature model rank the matched components in a way that will promote certain components. So, for instance, instances of Carolingian ⟨s⟩ are typically isolated, easy to retrieve, and will fit the bounding box in a regular way. A traditional palaeographer, by contrast, would probably pay attention to different letter types, guided by assumptions about which letters tend to be the most distinctive ones for individual scribes. As **Figure 3** shows, two-letter components, like ⟨er⟩, ⟨or⟩, ⟨en⟩, and ⟨er⟩, also contribute to the attribution process. We can see this as a way in which the present system is able to capture aspects of the “graphical chain”. By modifying the parameter setting constraining the segmentation process it is possible to instruct the system to extract a relatively larger portion of two-letter, or even wider, components. It would also be possible to use clustering of the labelled components in order to enforce that different shapes (letters) are considered in a more controlled way in the attribution process.

The evaluation data do not present the more challenging task of identifying scribes across different codices, with possibly different scripts, let alone across different languages. Rather, in the data, each scribe is represented by one codex in one language. As illustrated by the examples discussed above, language influences which writing components are likely to be extracted and consequently how they can be matched. To explore the present system for cross-language scribe attribution is an interesting possibility for future research.

Considering that the present system is a simple and straightforward one, it works remarkably well. It attributes scribes to manuscript images with a high degree of correctness, and it has the ability to show us why it counts an image as the work of a known scribe. In order to create a really useful software tool from the ideas that we have exploited here, the system should be equipped with an interface that allows the user to experiment with different modules and parameter settings. Furthermore, as hinted above, systems of this kind should be implemented in a fashion that makes it possible to work with really large collections of manuscripts.



## Additional File

The additional file for this article can be found as follows:

- **Appendix.** The Medieval book hand data set. DOI: <https://doi.org/10.16995/dm.67.s1>

## Acknowledgements

This work has been carried out in two projects, supported by the Swedish Research Council (Vetenskapsrådet, Dnr 2012-5743) and Riksbankens Jubileumsfond (Dnr NHS14-2068:1), and led by Anders Brun and Lasse Mårtensson, respectively. The current paper is based on a presentation prepared for the conference *Parchment, Paper and Pixels Medieval Writing and Modern Technology*, Maastricht, February 2–3, 2017, organized by the working group “Writing and Writing Practices in the Medieval Low Countries” (Schrift en Schriftdragers in de Nederlanden in de Middeleeuwen–SSNM), in collaboration with the Huygens Institute for the History of the Netherlands, the Regional Historic Centre Limburg, and the Henri Pirenne Institute for Medieval Studies (Ghent University). We also want to acknowledge the helpful comments from the anonymous reviewers, who suggested several improvements of the presentation.

## Competing Interests

The author has no competing interests to declare.

## References

- Aussems, Mark**, and **Axel Arjan Brink**. 2009. “Digital Palaeography.” In: *Kodikologie und Paläographie im Digitalen Zeitalter – Codicology and Palaeography in the Digital Age*. Malte Rehbein, Patrick Sahle, and Torsten Schaßan, in cooperation with Bernhard Assmann, Franz Fischer, and Christiane Fritze (eds.), 293–308. Schriften des Instituts für Dokumentologie und Editorik 2. Norderstedt: Books on Demand.
- Brink, Axel Arjan**. 2011. *Robust and Applicable Handwriting Biometrics* (dissertation). Rijksuniversiteit Groningen. <http://www.ai.rug.nl/~axel/dissert/brink-handwritingbiom.pdf>.



- Bulacu, Marius, and Lambert Schomaker.** 2007. "Text-Independent Writer Identification and Verification Using Textural and Allographic Features." *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(4): 701–717. DOI: <https://doi.org/10.1109/TPAMI.2007.1009>
- Ciula, Arianna.** 2005. "Digital Palaeography: Using the Digital Representation of Medieval Script to Support Palaeographic Analysis." *Digital Medievalist*, 1. DOI: <https://doi.org/10.16995/dm.4>
- Cloppet, Florence, Veronique Eglin, Marlene Helias-Baron, Cuong Kieu, Nicole Vincent, and Dominique Stutzmann.** 2018. "ICDAR2017 Competition on the Classification of Medieval Handwritings in Latin Script." In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 1371–1376. DOI: <https://doi.org/10.1109/ICDAR.2017.224>
- Dahllöf, Mats.** 2014. "Scribe Attribution for Early Medieval Handwriting by Means of Letter Extraction and Classification and a Voting Procedure for Larger Pieces." In: *2nd International Conference on Pattern Recognition (ICPR)*, 1910–1915. DOI: <https://doi.org/10.1109/ICPR.2014.334>
- von Euw, Anton.** 2008. "St. Gallen, Stiftsbibliothek, Cod. Sang. 562." <http://www.e-codices.unifr.ch/en/description/csg/0562/vonEuw>. Referred to as: Euw, Anton von. 2008. *Die St. Galler Buchkunst vom 8. bis zum Ende des 11. Jahrhunderts, Band I: Textband*. St. Gallen: Monasterium Sancti Galli, Bd. 3, 439–440, Nr. 112.
- Fiel, Stefan, Florian Kleber, Markus Diem, Vincent Christlein, Georgios Louloudis, Stamatopoulos Nikos, and Basilis Gatos.** 2017. "ICDAR2017 Competition on Historical Document Writer Identification (Historical-WI)." In: *2017 14th IAPR International Conference on Document Analysis and Recognition*, 1377–1382. DOI: <https://doi.org/10.1109/ICDAR.2017.225>
- He, Sheng, and Lambert Schomaker.** 2016. "Co-Occurrence Features for Writer Identification." In: *2016 15th International Conference on Frontiers in Handwriting Recognition*, 78–83.
- He, Sheng, Marco Wiering, and Lambert Schomaker.** 2015. "Junction Detection in Handwritten Documents and Its Application to Writer Identification." *Pattern Recognition*, 48: 4036–4048. DOI: <https://doi.org/10.1016/j.patcog.2015.05.022>

- Jain, Rajiv,** and **David Doerman.** 2014. "Combining Local Features For Offline Writer Identification." In: *14th International Conference on Frontiers in Handwriting Recognition*, 583–588. DOI: <https://doi.org/10.1109/ICFHR.2014.103>
- Mengis, Simone.** 2013. *Schreibende Frauen um 1500: Scriptorium und Bibliothek des Dominikanerinnenklosters St. Katharina St. Gallen*. De Gruyter. DOI: <https://doi.org/10.1515/9783110220896>
- Otsu, Nobuyuki.** 1979. "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems, Man and Cybernetics*, 9(1): 62–66. DOI: <https://doi.org/10.1109/TSMC.1979.4310076>
- von Scarpatetti, Beat Matthias.** 2003. "St. Gallen, Stiftsbibliothek, Cod. Sang. 603." <http://www.e-codices.unifr.ch/en/description/csg/0603/>. Referred to as: Scarpatetti, Beat Matthias von. 2003. *Die Handschriften der Stiftsbibliothek St. Gallen, Bd. 1: Abt. IV: Codices 547–669*. Hagiographica, Historica, Geographica, 8–18. Jahrhundert. Wiesbaden 2003, 162–166.
- Schomaker, Lambert, Marius Bulacu,** and **Katrin Franke.** 2004. "Automatic Writer Identification Using Fragmented Connected-Component Contours." In: *Ninth International Workshop on Frontiers in Handwriting Recognition*, 185–190. DOI: <https://doi.org/10.1109/IWFHR.2004.22>
- Stansbury, Mark.** 2009. "The Computer and the Classification of Script." In: *Kodikologie und Paläographie im Digitalen Zeitalter – Codicology and Palaeography in the Digital Age*. Malte Rehbein, Patrick Sahle, and Torsten Schaßan, in cooperation with Bernhard Assmann, Franz Fischer, and Christiane Fritze (eds.), 237–249. Schriften des Instituts für Dokumentologie und Editorik 2. Norderstedt: Books on Demand.
- de Stefano, Claudio, Francesco Fontanella, Marilena Maniaci,** and **Alessandra Scotto di Freca.** 2011. "A Method for Scribe distinction in Medieval Manuscripts Using Page Layout Features." In: *Image Analysis and Processing—ICIAP 2011*. Springer-Verlag, Berlin, Heidelberg, 393–402. DOI: [https://doi.org/10.1007/978-3-642-24085-0\\_41](https://doi.org/10.1007/978-3-642-24085-0_41)

- Stutzmann, Dominique.** 2016. "Clustering of Medieval Scripts Through Computer Image Analysis: Towards an Evaluation Protocol." *Digital Medievalist*, 10. DOI: <https://doi.org/10.16995/dm.61>
- Tang, Youbao, Xiangqian Wu, and Wei Bu.** 2013. "Offline Text-independent Writer Identification Using Stroke Fragment and Contour Based Features." In: *2013 International Conference on Biometrics (ICB)*, 1–6. DOI: <https://doi.org/10.1109/ICB.2013.6612988>
- Wolf, Lior, Nachum Dershowitz, Liza Potikha, Tanya German, Roni Shweka, and Yaacov Choueka.** 2010. "Automatic Palaeographic Exploration of Genizah Manuscripts." In: *Kodikologie und Paläographie im Digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2*, Franz Fischer, Christiane Fritze, Georg Vogeler, in collaboration with Bernhard Assmann, Malte Rehbein and Patrick Sahle (eds.), 157–179. Schriften des Instituts für Dokumentologie und Editorik 3. Norderstedt: Books on Demand.

**How to cite this article:** Dahllöf, Mats. 2018. "Automatic Scribe Attribution for Medieval Manuscripts." *Digital Medievalist* 11(1): 6, pp. 1–26. DOI: <https://doi.org/10.16995/dm.67>

**Published:** 24 December 2018

**Copyright:** © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



*Digital Medievalist* is a peer-reviewed open access journal published by Open Library of Humanities.

**OPEN ACCESS** The Open Access icon, which is a stylized 'O' with a person inside, representing open access to knowledge.