

## Digital Hermeneutics, Medieval Texts, and Urban History: A Case Study from Aberdeen, Scotland

Wim Peters <w\_dot\_peters\_at\_textdimensions\_dot\_eu>, Text Dimensions (NL)

William Hepburn <william\_dot\_hepburn\_at\_abdn\_dot\_ac\_dot\_uk>, University of Aberdeen (UK)  <https://orcid.org/0009-0004-1725-6835>

DOI: <https://doi.org/10.63744/xwvq7n6xea5g>

### Abstract

This article presents an inquiry into the use of natural language processing (NLP) methods to enrich, rather than replace, hermeneutical workflows in historical research. Making use of digital technologies in the form of existing tools and custom computational processing, it advocates an approach that fosters deep text interpretation by historical scholars with the aim of incrementally addressing and expanding the range of research questions asked about a particular theme, using a particular textual corpus.

In general, this paper argues that success hinges on the possibility of incrementally and systematically unlocking new data for hermeneutical knowledge acquisition and integration without compromising the role of the human historical researcher and their core scholarly analysis methods. This entails that, just like in a traditional manual, close reading effort, the scholar should retain maximum control of research activity and strategy.

Our main finding is that the digital hermeneutical method applied in the described work provides relevant results for a 'gude compt and reking' (the late medieval concept of 'good account' described in this article) of the conceptual structure of our domain. The general conclusion we draw from this is that NLP brings possibilities for more focused and fine-grained qualitative text analysis in this domain, while allowing easy access to a global perspective on the texts under study. We contend that a combination of tailored quantitative and qualitative text analysis methods can be integrated into a flexible research workflow, which empowers the hermeneutical work of humanities researchers.

## Hermeneutics

Hermeneutics as a theory is concerned with the nature of interpretation, its subjectivity, historicity, multi-perspectivity, and its dependence on context. From classical times until now the hermeneutical approach has been a central tenet of scholarly research, informed by both philology and philosophy. With the origin of its name and definition rooted in Aristotelian terminology, its methodology evolved till modern times through bible exegesis [Marshall 2011]. Around 1900 the modern proponent Wilhelm Dilthey turned hermeneutics into a fully-fledged methodology for non-positivist scientific research [Mallery, Hurwitz, and Duffy 1986]; [Makkreel 2021].

Understanding, both the basis and goal of the hermeneutical approach, is incrementally augmented by the methodology of the hermeneutical circle (see figure 1). This involves an iterative, circular research workflow sequence, in which the scholar starts with their a priori knowledge accumulated from experience and understanding, then inspects textual fragments or details in order to acquire knowledge (contextualization), and finally integrates this new knowledge into the scholar's incrementally growing understanding through interpretation and explanation. Dilthey claimed that interpretations become more valid as they assimilate more knowledge about the author(s) of the text and their values instead of reflecting the interpreter's own values or sense of reality. The process eventually reaches stability because successive interpretations serve to constrain subsequent refinements in the background model of the author.

1

2

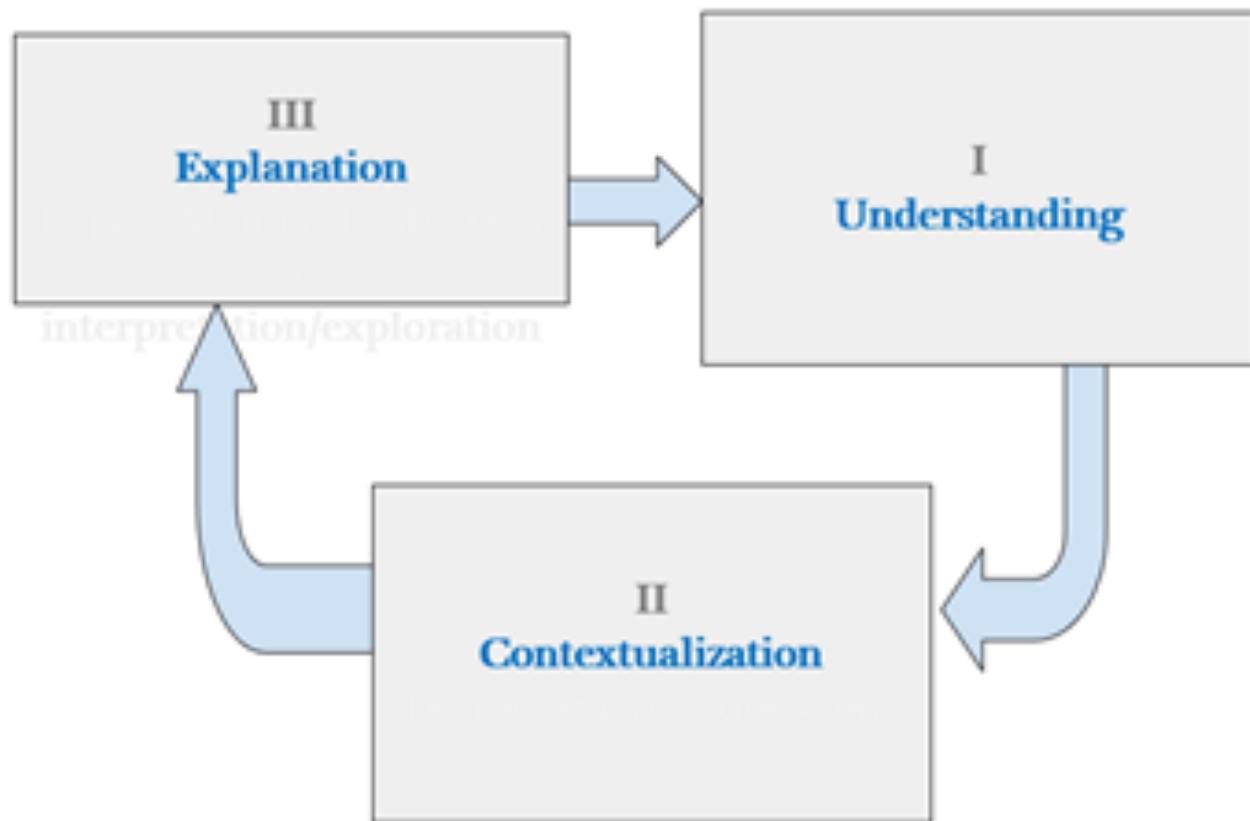


Figure 1. Methodological workflow: The hermeneutical Circle

## Digital Hermeneutics and Digital History

Computational support for this hermeneutical activity, where automated analysis supports the manual hermeneutical methodology of the expert, is considered part of digital history and digital hermeneutics. 3

A key motivation for applying forms of automated analysis lies in the availability of increasing amounts of digital textual data, which has posed a challenge to the traditionally manual scholarly research workflows. Recent decades have seen an increasing availability of historical texts in digital formats through the work of digital humanities projects using a variety of automated and non-automated methods. This data increase creates a significant bottleneck for an exhaustive scholarly understanding of the content of the textual source material and the historical domain it represents [Peters, Parks, & Lennan 2019]. 4

Given the quantity of textual material it is often not possible to obtain an adequate interpretation that applies to the whole corpus under study using the sole application of qualitative, manual techniques, which often are linear and time consuming. Scholars are working with datasets that are getting so large it has become difficult to assume that manual methods of traditional text analysis will ensure extensive scholarly understanding of them. For this reason many scholars in various humanities disciplines adopt mixed method approaches [Creswell 2009]. These combine the traditional close reading methodology with distant reading techniques involving quantitative analysis [Moretti 2005]. 5

Under the signifier of 'digital history', the area of our interest, historians experiment with tools, concepts and methods from other disciplines, such as, in our case, natural language processing (NLP), in order to enlarge the scope and coverage of their traditional research methods [Romein et al. 2020]. NLP is a subfield of linguistics, computer science, and artificial intelligence. Under its banner it combines a broad range of computational techniques, varying from linguistic approaches, such as rule-based modeling of human language, to statistical methods, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text and to derive meaning [Foote 2019]. Our interest lies in investigating the feasibility of adopting a subset of NLP techniques that provide potentially meaningful input into a specific digital historic interpretation workflow, while maintaining the leading role of the human scholar in the interpretation, evaluation and selections of results. 6

## Integrating NLP into the hermeneutical Circle

The aim of including NLP techniques into hermeneutical scholarly workflows should not entail the replacement of the historian's interpretive and hermeneutical work [Zaagsma 2013]. Automated text analysis results can add to the conceptual breadth, depth and granularity of hermeneutical scholarly understanding as a whole, but they cannot produce that understanding on their own. A digital hermeneutical workflow should establish a balanced integration of both traditional qualitative approaches and quantitative techniques, which can help address the limitations of each method alone [Gibbs and Owens 2013]. A synergetic interaction between close interpretative reading and algorithmic analysis requires an awareness from scholars of how they are using and interpreting data in order to incrementally find the answers they are looking for. How we can reach this level of awareness of the influences, benefits and pitfalls of computer analysis, and rely on the computer as a hermeneutical instrument, are questions we are trying to address here.

7

The challenge is to marry the powerful techniques brought in by NLP with the traditional scholarly close reading method. From an epistemological perspective, as we have also seen argued above [Zaagsma 2013], only applying fully automated NLP results does not satisfy scholarly hermeneutical requirements. An approach involving only NLP techniques is fragmented, conceptually incomplete, and offers only potentially useful suggestions derived from automatic text analysis. The results lack an *a priori* conceptual framework against which these results should be judged and integrated. Therefore, we cannot assume that the knowledge obtained by solely automatic means is sufficiently relevant to answer historical research questions.

8

However, we do not regard the contribution from NLP as essentially different from human interpretation, because both interpret text structure and content. Automated text analysis is rather to be regarded as complementary to human text processing, offering additional perspectives on textual meaning, which are generated through an algorithmic deformation of the text itself [Ramsay 2011]. This perspective cannot yield "hermeneutically neutral" results because each tool brings its own methodological assumptions, and therefore its own hermeneutical perspective [van Zundert 2016]. For instance, every statistical measure depends on a language model that assumes significance in distributional properties, which do not necessarily reflect meaningful semantic salience and relationships, and also potentially filter out important information by imposing thresholds for statistical significance. Also, concordancing or keyword in context (KWIC) results divide running text up into snippets of a predetermined length, which can lead to information loss if the context size is not large enough to reveal certain relational patterns between the keyword and its context elements.

9

For our purposes, the synergy between human and computer is best described by hermeneutical philosophers such as Hans-Georg Gadamer, who emphasizes the importance of intersubjectivity, a term that means agreement or the sharing of subjective states of comprehension by two or more individuals [Scheff 2015], and the notion of dialogue in the construction of meaning [Regan 2012]. According to hermeneutical theory, the inclusion of NLP techniques provides complementary analysis results for the hermeneutical process of interpretation, assuming intersubjectivity emerges through the dialectical process of interpretation by and dialogue between interpretation agents, who all contribute their own perspectives to the dialogue. In other words, we pose that, when working with both animate and inanimate interpretation agents, each NLP technique is to be treated as an additional dialectic participant in this interpretation process, whereas only the scholar is able to interpret and evaluate the contributions from each participant, and consolidate overall understanding by incrementally building a semantic schema of the knowledge domain (see section "Description of the Research and Assessment of Outcomes").

10

In order to operationalize a dialogue between human and computer within a hermeneutical scholarly workflow, where the human stands at the center of interpretation, we require a feedback procedure between NLP's distant reading results on the one hand and close reading manual interpretation on the other. From an automated distant reading perspective, NLP techniques should provide results that are customized to scholars' research questions [Parks and Peters 2022]; [Peters and Wyner 2016]; [Woolf and Silver 2018]. From a close reading perspective, tools should enable the manual exploration and evaluation of the distant reading results. This gives weight to flexible experimentation and ad hoc interpretation of observed patterns and outliers, rather than giving in to full reliance on a narrow quantified empirical method [Rockwell 2003], enabling a felicitous text exploration focused and informed by computer output [Ramsay 2010].

11

Within the scope of the hermeneutical cycle, the integration of NLP requires a constant cyclic interaction between automatic and manual analysis, whereby workflow cycles allow scholars to explore and interpret the text space in which distant reading results occur using close reading. After manual evaluation of the results they can then integrate them into their holistic understanding of the domain, and decide whether it is required to instantiate a new cycle of hermeneutical analysis. Figure 2 below shows a schematic representation of each cycle.

12

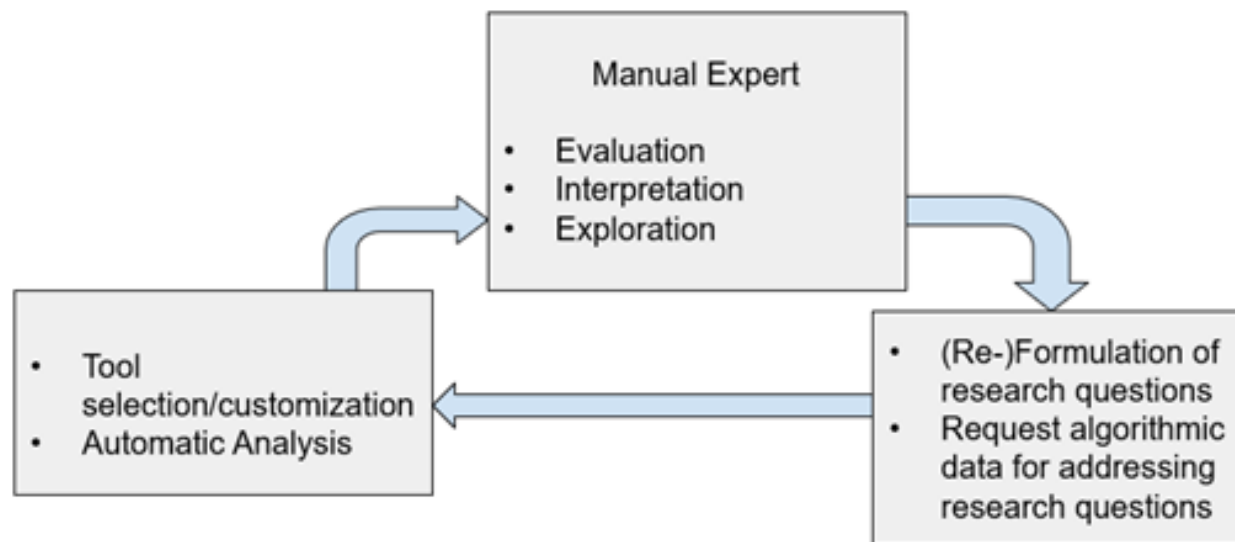


Figure 2. Workflow cycle involving the use of automated analysis.

## Our Historical Focus

The “Finance, Law and the Language of Governmental Practice in Late Medieval Towns: Aberdeen and Augsburg in Comparison” (FLAG) project <sup>[1]</sup> was run by the University of Aberdeen and the Johannes Gutenberg Universität Mainz. The project examined the practice of government and the associated ideas of government through the language of civic administrative records from two European towns. Aberdeen and Augsburg were very different towns in their political and economic settings, and there is no reason to posit any direct contact or connection between them in this period unlike, for instance, towns deemed to be in the same ‘family’ through the sharing of law codes [Bartlett 1994, 172–77]. These dissimilarities do not hinder, and arguably clarify, the project’s goal of exploring a deep-rooted culture of urban government in Europe in this period beyond the level of like-for-like comparisons and direct connections.

13

Aberdeen and Augsburg have one similarity that is crucial to the project: the extensive survival of their civic records from the period and the recent digitization of many of those records. For Aberdeen, the FLAG project uses the Aberdeen Council Registers covering the period 1398-1511, recently transcribed and digitized at the University of Aberdeen as part of the ‘Law in the Aberdeen Council Registers’ project (University of Aberdeen, 2016-2019, funded by the Leverhulme Trust). This is the Aberdeen Registers Online: 1398-1511 (ARO) corpus, <sup>[2]</sup>. The work of this project also produced a prototype online search tool for these records called ‘Search Aberdeen Registers’. <sup>[3]</sup> In the scholarly work described below this tool was used for quick navigation of the text during ‘manual’ research but does not form an integral part of the digital hermeneutical text analysis outlined by this article. The starting point for Augsburg was the digital transcription of the Augsburg Baumeisterbuecher (accounting) covering 1320-1466/70, though further Augsburg sources were transcribed and incorporated into the analysis as the project progressed.

14

The project identifies three broad domains for the study of governmental language shared by these towns: ‘Unity’, based on their communitarian politics; ‘Order’, covering the judicial functions of urban government; and ‘Budget’, covering financial administration and the role of urban government in economic activity. This framework represents the a priori knowledge accumulated from the experience and understanding of the project members and the starting point of the hermeneutical analysis process. In this article the example is drawn from work on ‘Budget’ in the Aberdeen records.

15

In the context of Aberdeen’s medieval records, the traditional hermeneutical method might involve the historian, drawing on their existing understanding of the period and the themes they were exploring, skimming the Aberdeen records for subjects of interest or reading them in order. These methods are both time-consuming and prone to error and omission. They also require considerable time investment within each hermeneutical loop.

16

Furthermore, a major challenge these records pose for manual analysis is the high volume of formulaic and repetitive language, making the selection of representative material for analysis difficult. Early historical studies avoided this problem by combing the records for exceptional ‘highlights’ and presenting them separately from the more ordinary material (e.g. Stuart, 1844). The opportunity for the application of digital hermeneutical methods to this corpus, therefore, was to identify trends within the corpus and allow rapid analytical movement between micro and macro levels of the text leading to qualitative conclusions about the conceptual language of

17

the corpus, which can, through quantitative evidence, be understood as prominent in terms of the corpus as a whole, prioritizing the language of the everyday over the exceptional. In this way the research was shaped at the level of its research questions by the digital methodologies of the project.

The feasibility of automated analysis was made substantially easier by recent work, which has increased the accessibility of the Aberdeen records by transcribing them out of archaic script into modern script and presenting these transcriptions in digital form. On top of this, the records can be word-searched or browsed in order using the Search Aberdeen Registers tool <sup>[4]</sup> While these developments have made it much easier to explore the corpus, there is considerable scope for NLP methods to enhance its hermeneutical interpretation. Without the availability of these digital resources and methods, the questions asked by the project would have been difficult to pursue to the extent that it is unlikely the choice would have been made from the outset to focus on the everyday language of the records as an entire corpus.

18

The historical material from Aberdeen addressed in this case study, with around 27,000 types (unique text corpus elements) and 1,500,000 tokens (total sum of text corpus elements), is, in the context of natural language processing, not a very large data set and therefore not an obvious candidate for the application of certain quantitative NLP techniques. Nevertheless, we postulated that automated analysis results, taken into account with the caveats above, could offer valuable information about the ways in which concepts are expressed and related to each other in these texts and a perspective that would be very difficult or impossible to achieve with non-computational methods. We investigated this in our project work. The goal of the analysis was to form a conceptual characterization of the budget domain in relation to town governance in medieval Aberdeen, using the digital tools to work through a hermeneutical process from that broad, *a priori* conceptualization to a more focused understanding of the specific language within this domain used in the records. We hoped that these findings, along with those from studying the domains of 'order' and 'unity', could lead to wider conclusions about civic government in Aberdeen and Scotland and, even more broadly, Europe, particularly in connection with the study of records from Augsburg also underway as part of the project. The NLP techniques used by the project helped to define these domains but also to establish connections between them because they allow language from one domain to be studied at scale in the context of other language, including from other domains, which appears near it in the text.

19

## Workflow Requirements and Specification

In order to enable a continuously growing holistic understanding of our domain of interest, we adopted a practical workflow, which instantiates the cyclic workflow illustrated in figure 2 above, and involves both distant and focused close reading of text passages, as discussed in section "Digital Hermeneutics and Digital History".

20

First, we selected techniques for the acquisition of relevant words, concepts and relations between them. Our criterion was to apply algorithms with increasing degrees of computational complexity, which we preferred to be as accessible and workable as possible for the scholar.

21

Our first two functionalities, the computations of frequency and concordancing, also known as Keyword in Context (KWIC, see above) meet that criterion. Frequency provides an intuitive first clue about the possible importance of a term, whereas concordancing provides an accessible overview of the occurrences of these terms and their immediate textual context.

22

For more complex techniques, we selected functionalities that would yield interpretable statistical correlations between groups of concepts. The results would suggest links between words or terms selected by means of frequency and concordancing, or produce additional text-derived candidates that had not yet been captured by the first two techniques mentioned above. In order to meet this functionality requirement, we chose distribution-based groupings in the form of collocations and topics.

23

Collocations or co-occurrence patterns capture pairs of word pairs whose members are distributionally associated with each other. Statistical measures such as Point-wise Mutual Information identify word pairs that co-occur in the text in a statistically significant manner [Manning and Schütze 1999].

24

Topic modelling is a method for finding distributional clusters of two or more words in larger amounts of text [Blei 2012]; [Brett 2012]. Successful topics may reflect thematic coherence, identify relevant new concepts and show possible connections between cluster members.

25

The tools that we selected serve the requirement that they cover as many techniques as possible from our list of four and give automated pointers towards the identification of relations between words, following an onomasiological approach, a word or token based strategy detecting important related concepts by grouping words into 'lexical clusters' [Steinmetz, Freedon, and Fernández Sebastián 2017]. The context-dependent nature of meaning, based on the distribution of words in the text, and the significance of meaning relations between words on the basis of the patterns with which they co-occur in text give us an insight into the conceptual structures that are at play in the domain. A famous quotation that illustrates this point is the following by J.R. Firth: "You shall know a

26

word by the company it keeps" [Firth 1957].

While inspecting and evaluating the suggestions from these quantitative distant reading tools, we assumed that the subject expert decides, at the appropriate stages within the workflow, which words, concepts and relations are relevant for the research and indicative of potential answers to the research questions. At any time, the scholar should therefore be able to shift the focus of attention to any (context) word and embark on alternative routes of investigation through the text. For this purpose we stipulated our final tool requirement of a flexible bridging functionality enabling a switch between suggestions from automated distant reading and manual, focused close reading of text fragments containing these suggestions.

27

Figure 3 below illustrates the various workflow cycles in which these NLP functionalities and tool requirements were operationalized (see box 2 and 3).

28

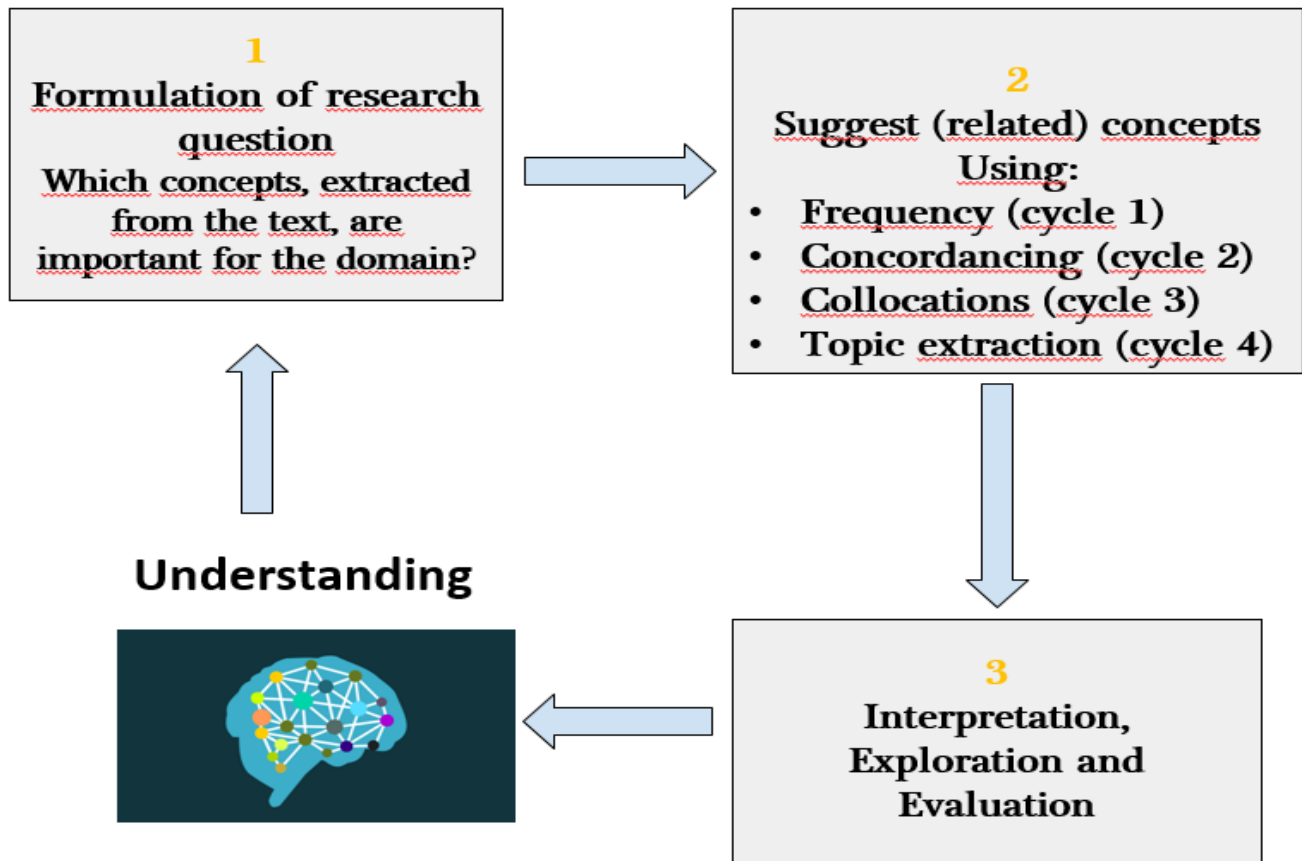


Figure 3. Four cycles of the workflow model in Figure 2

In our selection of NLP tools we concentrated on the tools that have the desired functionalities, their availability and their ease of deployment. The tools we chose are freely obtainable and relatively easy to use. Although they may not be state of the art, they have a well established history of development and adoption in the field of digital humanities, and serve as a good example of what historians can do with automated text analysis, which we assume to play an ancillary role in the human researcher driven analysis.

29

The primary tool we chose for the conceptual exploration and evaluation of the automatic results and the close inspection of texts was AntConc [5] [Anthony 2018], which we used for frequency information, concordancing and collocations. [6] This tool allows setting parameters for the automated identification of concepts of interest, such as minimum token frequency, the use of stopwords and a threshold for context window size when performing KWIC. Moreover, with AntConc the user can jump to text instances where the keywords under consideration occur, thus providing a flexible bridging functionality between distant and close reading. AntConc comes with detailed tutorials and is very intuitive in its use. Once scholars grasp the aspects of the user interface with respect to sorting, and the nature of the results from the statistical methods involved in the activities above, they can very easily extend their 'thinkering' [Fickers, Tatarinov, and van der Heijden 2022] or 'disciplined play' of corpus exploration [Ramsay 2010] by themselves by finding a purposely selective and subjective path through a large amount of information. This offers the researcher a clear sense of control and authorship while using the tool without the need for expert technical knowledge.

30

The key features of AntConc used in this case study were as follows. Firstly, AntConc's ability to sort the words in the corpus by frequency offered a simple but valuable tool to help the scholar stay focused on the common, quotidian aspects of the corpus and was used frequently throughout the process. Concordancing describes the various tools and settings within AntConc which allow words to be examined within their textual context. AntConc's KWIC (keyword in context) tool is particularly powerful in this regard, allowing the scholar to sort instances of a word and its context into a list arranged by frequency of different contextual words and phrases and quickly jump to the full textual context of each example for further examination. AntConc's collocation function allows the scholar to quickly produce statistically-determined lists of the terms which appear alongside a given term within a text window of a given size. While this sometimes served to simply corroborate the findings from concordancing, it opened up new avenues of exploration and allowed more fine-grained analysis of the relationship between words. These tools in conjunction were used to reach the conclusions described below.

31

Topic modelling is the only technique outside the scope of AntConc. In order for the humanities researcher to operationalize this in her research, several tools are freely available, such as the DARIAH-DE TopicsExplorer<sup>[7]</sup> and various Python libraries. Given our relative familiarity with Python, we chose the flexibility of the Gensim text analysis package<sup>[8]</sup> This enables the application of a widely used topic modelling technique, called Latent Dirichlet Allocation (Blei, 2012), which yields distributionally-based clusters of words, with variable settings for the number of word token members per topic.

32

Analysis types such as topic extraction are increasingly easy to apply for researchers who are not proficient in computer programming. In order to avoid most of the coding bottleneck, it is nowadays possible to harness AI for the purpose of writing a Python script. With a minimal investment into Python coding itself, a basic Python script can be readily created by prompting an AI engine such as Perplexity<sup>[9]</sup> with the following text: "Write a python script that ingests a text file, tokenizes it and computes topic clusters of 10 unique tokens each using Gensim latent dirichlet allocation." The result of this prompt is shown in Appendix A. For English and other modern languages it is possible to obtain better results by adding language-specific stopword filtering and word stemming. These functionalities are included in the Natural Language Toolkit (NLTK) package<sup>[10]</sup> By progressively including explicit requests for additional functionality in the prompt it will enable the researcher to adapt the result from the AI created Python script to his or her specific needs. Overall, we consider prompt editing an interesting way for scholars to boost expertise in Python programming.

33

## Description of the Research and Assessment of Outcomes

In line with the hermeneutical circle, the starting point of the work was the scholar's *a priori* understanding of the category of budget in relation to late-medieval urban government, which was that the term encompassed the administration of government finances and resources, and was likely to be found in terms relating to subjects such as accounting, taxation or other forms of raising revenue, and government spending. The first stage in the iterative process, using AntConc, was to test whether this *a priori* understanding was a good fit, or whether it needed to be adjusted.

34

The manual selection from a vocabulary set extracted from the ARO corpus and sorted by frequency (cycle 1 in figure 3) served as the source for the compilation of relevant conceptual vocabulary by the scholar. Each keyword/term candidate could, if required, be further inspected in AntConc by jumping to the actual text fragment in which the candidate terms occur. For instance, scanning down from the top of a frequency list of words across the whole corpus reveals the word 'americamento' - a form of the Latin word for a judicial fine – as a term potentially relevant to 'budget' near the top of the list. Clicking on it displays each of the 3,210 instances of the word in a new list, also showing their immediate textual context and thus allowing a very quick manual survey of the common contexts in which the term occurs. From this survey of the text, it became clear that the *a priori* definition of 'budget' would be too limited to represent the language in this area across the corpus. The majority of records in the corpus are records of the burgh court, in which the burgh government frequently intervenes in disputes concerning trade and transaction between private groups and individuals. Given the close overlap between the burgh court and other areas of burgh government, illustrated by the overlap of personnel between these areas and the shared record-keeping practices represented by the council registers, it seemed that a fitting definition of 'budget' to represent the whole corpus should take this into account. Thus, as a result of testing the initial definition of the category of 'budget' using this process, it was broadened to encompass the entirety of the burgh government's involvement in economic matters, including all language in the areas of accounting, finance, money and transaction, thus including the burgh government's oversight, primarily through the burgh courts, of a range of private transactions as well as its own budgets.

35

Using this broadened definition, the next step in the process was to produce a list of the most common language in this area and to sort that language into categories. Using AntConc, it was straightforward to produce a list of the most common terms in the ARO corpus. From this, the scholar worked down the list using his prior knowledge to assemble a list of potential terms. The ability to quickly jump to the context of every individual usage of each term allowed the scholar to determine if the words were indeed being used in a manner relevant to the domain under study. For instance, the Scots verb *chargit* (charged) was used in relevant financial contexts when someone was 'charged to pay' someone as well in non-relevant contexts where someone was 'charged' with doing

36

something else – thus the word was not being used in a way that directly related to the ‘budget’ domain even if it sometimes appeared in related contexts (i.e. it was not being used in the sense of ‘he was charged £5 for his shopping’). Thus it was possible to produce a list of the most common terms in this area which was, in our estimation, more interpretatively valuable than a list generated by human examination alone, or corpus analysis techniques alone. It benefited from the whole-corpus quantitative analysis achieved with AntConc, and the inclusion and exclusion of terms by the scholar based on prior knowledge of the source material.

The words on this list could be sorted into categories based on their intrinsic definitions as well as the contexts in which the words appear. Using the technique of concordancing, displaying keywords in their left and right context, the scholar was able to explore the uses of keywords in the text, guided by the frequency of context words if required. Broadly speaking, words in the corpus within the ‘budget’ domain fit into four categories, as outlined in Table 1.

37

Category 1 – currencies etc.	Category 2 – Transaction-related actions	Category 3 - Types of payment	Category 4 – Calculation, summation
solidorum	pay	amerciamento	price
money	debit	redditu	summam
monete	soluet	amerciament	compt
merkis	pament	annuatim	rekning
siluer	payment	maile	
lib	debito	expensis	
denariorum	solucionis	amerciamentis	
penny	regratacione	fee	
solidos	awand	annuell	
marcarum	soluto	expens	
schillingis	soluendis	annual	
libris	selling	taxt	

Table 1. Classification of concepts in the ‘budget’ domain

The first category consists of words describing money and denominations of money (Category 1). These are fairly straightforward terms, relevant to economic studies of subjects such as pricing [Gemmill and Mayhew 1995] but offering little within the scope of this study.

41

The second category relates to words describing economic actions such as paying or selling (Category 2). Of these, *pay* and its variants is by far the most common.

42

The third category consists of the many words for different types of payments in the records (Category 3). These include the most commonly-occurring word in the list, *amerciamento*, which describes a judicial fine and points to the predominance of records of the burgh court among the ARO corpus.

43

The last category encompasses words relating to calculation, accounting and summing up (Category 4). Words in this category are not the most common in the list but are, as will be discussed further below, the most important for understanding the ideals underlying all the economic language of urban government in the ARO.

44

In the following hermeneutical cycles (2-4 in Figure 3), we aimed to identify additional concepts and relations between concepts using collocations and topics (see section "Description of the Research and Assessment of Outcomes"). This analysis reveals that most ‘budget’ words appeared in a context in which they formed some part of a transactional arrangement, whether recording the completion of a transaction, recording that a transaction had not been completed, or making arrangements for a transaction to be completed. Everything within the three salient categories (Transaction-related actions; Types of payment; Calculation, summation) outlined in table 1 above is encompassed by, or closely connected to, the subject of transaction. And the most common side of

45

transactions mentioned is that of payment. Considering Category 2 from the table, the word *pay* alone, never mind its other forms and variants, appears 2,257 times in the corpus. The most common context in which the word appears is in phrases such as ‘the baillies chargit him to pay’ [11]. These phrases describe orders being issued by the burgh court for litigants to pay some debt or charge the court has found that they owe. The third category outlined above, covering different types of payment, demonstrates the range of transactions that took place within the orbit of Aberdeen’s civic government. The most common word describing a payment or transaction in the records is *amerciament* [amercement], which refers to fines issued by the burgh court. Rather than an order to settle a debt between individuals, amercements were fines imposed by the court as punishment for transgressions committed by individuals within the burgh against other individuals, burgh authorities or the burgh community. In addition to these most common types of payments, words in this category refer to a much wider range of transactional activity, such as the payment of taxes, rents for land and property, fees of employees and expenses for burgh officials. Figure 4 below illustrates some of the other concepts linked to this transactional language, based on the study of the terms identified in the lists above as well as the context in which they appear as revealed by the concordancing tool in AntConc (cycle 2 in figure 3).

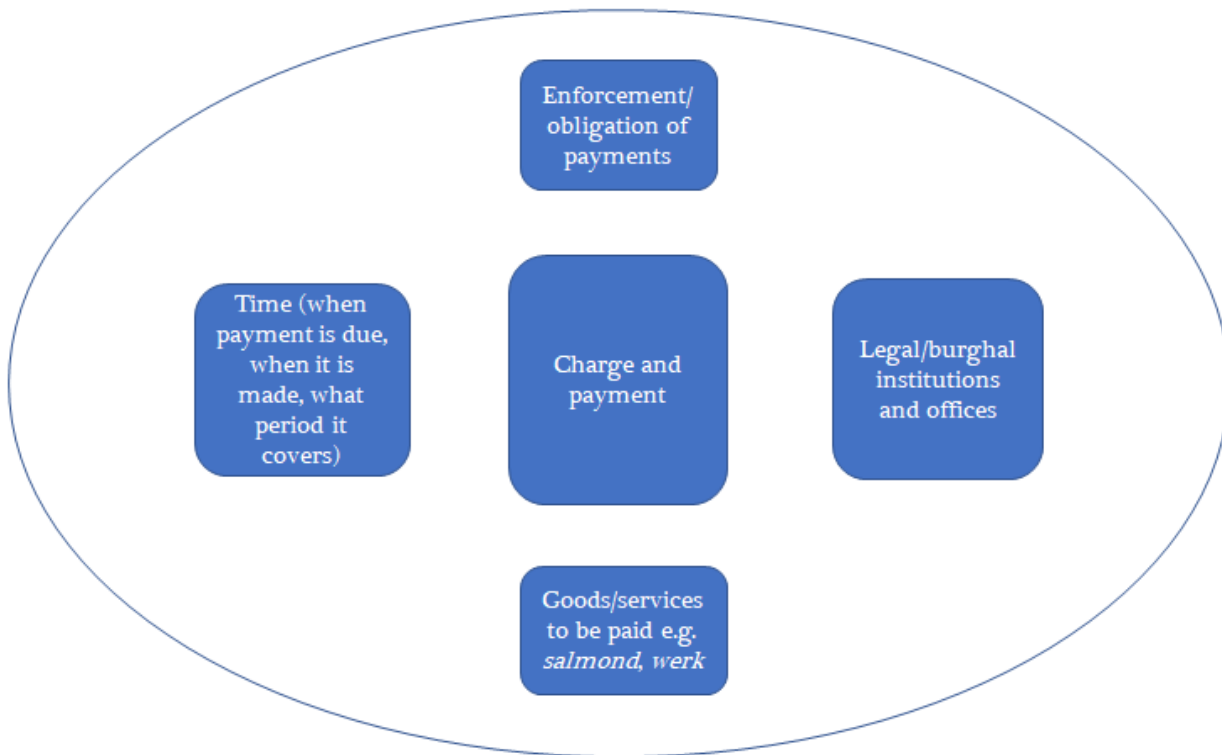


Figure 4. Charge and payment: relationship and contexts

Co-occurrence patterns or collocations (cycle 3 in figure 3) capture word pairs whose members are distributionally associated with each other. AntConc’s collocation functionality uses the Pointwise Mutual Information (PMI) measure to identify significant word pair co-occurrences within a context window (number of tokens to the left and right of the keyword) of pre-defined size. The user can then determine if the words are connected through a particular semantic relation that can be gleaned from the text fragments in which they both occur. Again, AntConc offers the functionality to jump from distant reading co-occurrences to a close reading of the textual fragments.

46

Analysis of collocation data benefited the scholar by allowing further structuring of the relationships between ‘Budget’ terms, particularly looking at how they formed part of a transactional structure. This can be illustrated by the example of ‘payment’, one of the most common, and most directly related to transactions, of the ‘Budget’ terms in the corpus. An examination of the words commonly located with payment allowed the creation of a semantic framework offering potential avenues for further research on the theme of transactions (See figure 4). This also links ‘Budget’ with the domain of ‘Order’, particularly through the subject of legal obligation and enforcement of payment. The example of *amerciament* links ‘Budget’ and ‘Order’ and applies to at least two of the concepts related to ‘charge and payment’ in figure 4, because it is an obligation to pay, resulting from legal a process, applied and enforced by a legal/burghal institution, the burgh court.

47

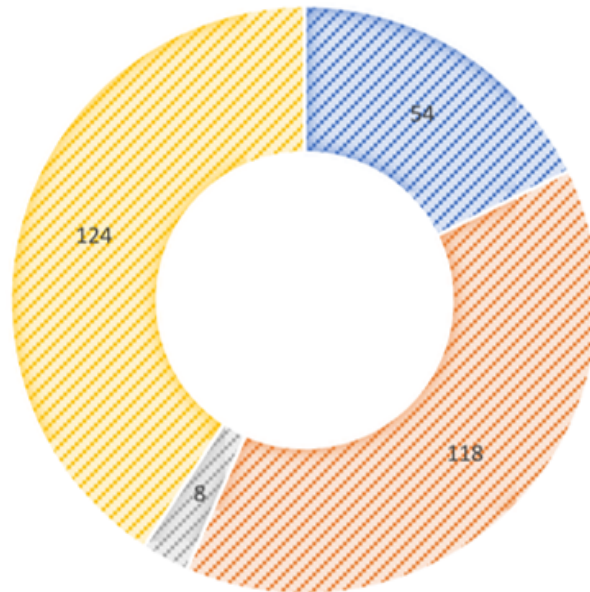
Concordancing and collocation also reveal links between elements from the language of payment. In Category 4 of the most commonly-occurring words in the domain of economic transactions (Table 1), those referring to calculation, accounting and summing up, the linked words *compt* and *rekning* stand out. The phrase *compt and rekning* appeared with the word ‘payment’ in a substantial

48

proportion of cases. Using the collocation tool on AntConc, which calculates the frequency of words appearing together, *pament* is the third most collocated word with both *compt* and *rekning*, with another spelling, *payment*, also coming in fifth place for both terms. An analysis of each of the three words and their relationships to the others individually (see Figure 5 below) shows that *payment* appears alongside *compt* around 20% of the time while it appears with *rekning* in roughly one third of cases. Even where the phrases are not explicitly linked, *compt and rekning*, meaning an account or the process of accounting, have a clear conceptual link to *payment*.

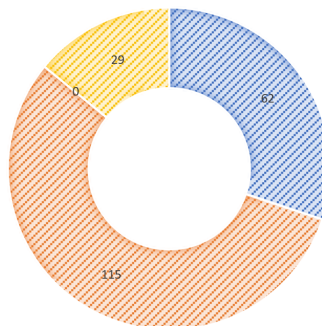
## COMPT

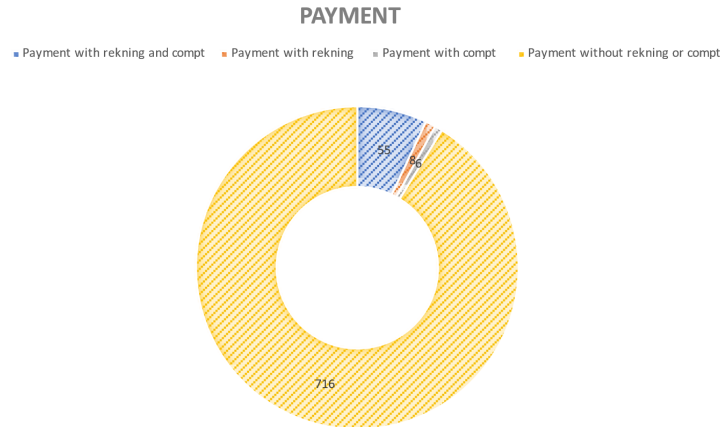
■ Compt with rekning and payment ■ Compt with rekning ■ Compt with payment ■ Compt without rekning or payment



## REKNING

■ Rekning with compt and payment ■ Rekning with compt ■ Rekning with payment ■ Rekning without compt and payment





**Figure 5.** Collocations of *compt*, *rekning* and payment

While the charts above show a link between *payment* and both *compt* and *rekning*, the link between *compt* and *rekning* is stronger. Using the collocation tool, the number one collocated word for each word is the other. Also, looking at the Figure 5 shows that each appears with the other in the majority of cases where they appear, overwhelmingly so in the case of *rekning*. The phrase *compt and rekning* is a binomial, which were common in Scots legal texts of this period. Joanna Kopaczyk [Kopaczyk 2009] describes 'binomials proper' as those in which the two words involve least some semantic repetition rather than complementarity: 'Their appearance in the text is not motivated by the need to hone the meaning, make it more precise, extend it, or make it less obscure. They are rather a stylistic device... the creation of binomials was governed by prosodic, etymological or other factors, and not by meaning.'. Therefore, while the link between *compt* and *rekning* represents a single semantic unit, the link between this unit and word 'payment' represents a meaningful link between two different concepts.

49

Some of the instances of records where the types of payments (Category 3 in Table 1) described above appear also include the phrase *compt and rekning*. An entry of uncertain date, but almost certainly from 1503, concerns an investigation into whether costs and expenses made upon land inhabited by William Rolland were less than, or more than, the mail he owed to Davy Colp for the land. In the former circumstance he would have to vacate his house on the land, in the latter he could stay there. To determine the case, Rolland was ordered to present his 'bil of expens compt and rekning made Apon the lande that he now Inhabitis' to Colp and they were to choose auditors to examine the accounts [12]. In this case *compt and rekning* appears in a phrase describing physical accounts made by Rolland and covering his expenses on the land he leased. The term, even when used in this context, means more than just the physical accounts. It describes a process that has been followed, one of collation, calculation as well as, in this case at least, recording. Moreover, it also signals the act of taking responsibility for one's conduct in office and the process whereby this was achieved, i.e. some form of auditing [Epurescu-Pascovic 2020]. This process is linked directly to payment in some cases. In 1491, the alderman and baillies (types of councillors) said that they had received 'be the handis of' Andrew Branch 'compt rekning and pament' of his part of the king's tax, with which they, 'in the nayme of the town', were well content and thus discharged Branch of his obligation [13].

50

*Compt and rekning* are sometimes accompanied by adjectives such as *good*, which had moral connotations. One of the most ubiquitous concepts in late medieval Christianity, the Day of Judgement, was a ready metaphor for, or even direct motivation to achieve, *gude compt and rekning*. The use of *compt and rekning* to describe both actual financial accounting practices and spiritual accountability to god, sometimes merged with judicial references, was broadly recognised in late-medieval Scotland, including in Aberdeen, as demonstrated by the copying of a poem by William Dunbar using financial and judicial accounting as moral metaphor into Aberdeen's sasine register in the early sixteenth century (Bawcutt, 1998). Against this background, the records of the ARO show the phrase used in a simple technical sense of balancing an account but also with moral connotations which must have been

51

heightened by the relatively small number of people regularly involved in transactions in Aberdeen and the familiarity between them, blurring the lines between good accounting in a technical sense and moral accounting to one's community [14]. Thus the Digital Hermeneutical methodologies deployed in this work allowed, through several interpretative steps, the subject scholar to carry out an analysis identifying the most common, quotidian language of budget across the corpus and to link this with an underpinning ideal with moral as well as legal and administrative connotations. While it is unsurprising that the research revealed much economic and transactional language given the well-established commercial role of a medieval burgh such as Aberdeen, the research identified a moral concept underpinning this activity and verified its link to the economic and transactional language that proliferates across the corpus. The concept is most directly expressed in the phrase 'compt and rekning', which was identified through a survey of frequent terms in the corpus identified using NLP tools. However, identifying the broader moral and social significance of this term required going beyond the results from NLP tools. The moral quality of this term was established through subject knowledge on subjects such as binomials and medieval Christianity as well as its proximity in the text to terms such as 'good' identified via NLP. The relationship of this term, and the concept underpinning it, to the majority of economic or transactional language in the corpus – particularly language about payment - was then established using the NLP tools.

Furthermore, as the full work of the project encompassing the domains of 'order' and 'unity' will illustrate, the concept of *gude compt and rekning* was the foundational principle of political culture within Aberdeen, and the political culture of Scotland's burghs fed into the political culture of the kingdom as a whole. Its social significance was determined with the assistance of NLP tools by examining the wider textual context in which the term appeared, such as injunctions for merchants to treat each other fairly, as well as broader historical research considering the range and variety of transactional activity within Aberdeen. The relationship of this term, and the concept underpinning it, to the majority of economic or transactional language in the corpus – particularly language about payment - was then established using the NLP functionalities, which also helped to link it to concepts beyond the initial 'budget' domain. This formed the basis of a wider analysis going further beyond the results of the NLP to form a working argument that the concept of *gude compt and rekning* was a foundational principle of good government within Aberdeen, and a distinctively urban principle within wider discourse on good government in Scotland.

Thus, the NLP tools enabled an initial overview based on our a priori understandings of the domain, largely through the use of word frequency tables. This provided a solid basis on which to carry out manual analysis of the terms identified, leading to a more nuanced and relevant approach to the research which incorporated economic as well as other types of transaction. Concordancing and collocation then allowed the relationships between terms to be defined more precisely, leading to the identification of the key phrase 'gude compt and rekning'. Manual research thus enriched these results by identifying the moral potency of this term, thus allowing a return to the texts with a focussed approach to the context of this phrase, in order to corroborate the findings from manual research, and establish the relationship of this term to the broader web of transactional language.

Our final cycle using topic modelling (cycle 4 in figure 3) did not suggest any new concepts or point at any thematic coherence other than that cluster members can be clearly linked to recognizable recurring formulae within the text, many of which are closely linked to the 'budget' domain. Indeed, at least three of the topics identified are closely related to commonly-occurring formulaic record types within the corpus (Croft Dickinson, 1957), more specifically entries describing the burgh court's issuing of judicial fines (*amerciamento/amercament*). This confirms our a priori understanding of the heavily formulaic character of much of the text as well as the prominence of 'Budget' as a domain.

The lack of new results from topic modelling seems to indicate that, with a smaller-sized corpus such as ours, simpler techniques (cycle 1-3) are possibly able to capture an adequate semantic representation of the domain insofar this can be captured by means of distributional techniques.

However, even if some topics could not be easily identified as statistical representations of manually-identifiable formulae, further analysis of these topics may reveal themes, related to 'Budget' as well as the other domains of the project, within extended text embedding the recognizable formal features of commonly-occurring record types as well as other textual sources. In this way topic modelling functions well within a hermeneutical framework, the computational methods providing a way to go back to the original records with new questions and hypotheses, which would not otherwise have occurred.

The economic and transactional conceptual structure obtained from the domain by means of our approach is illustrated in figure 6 below.

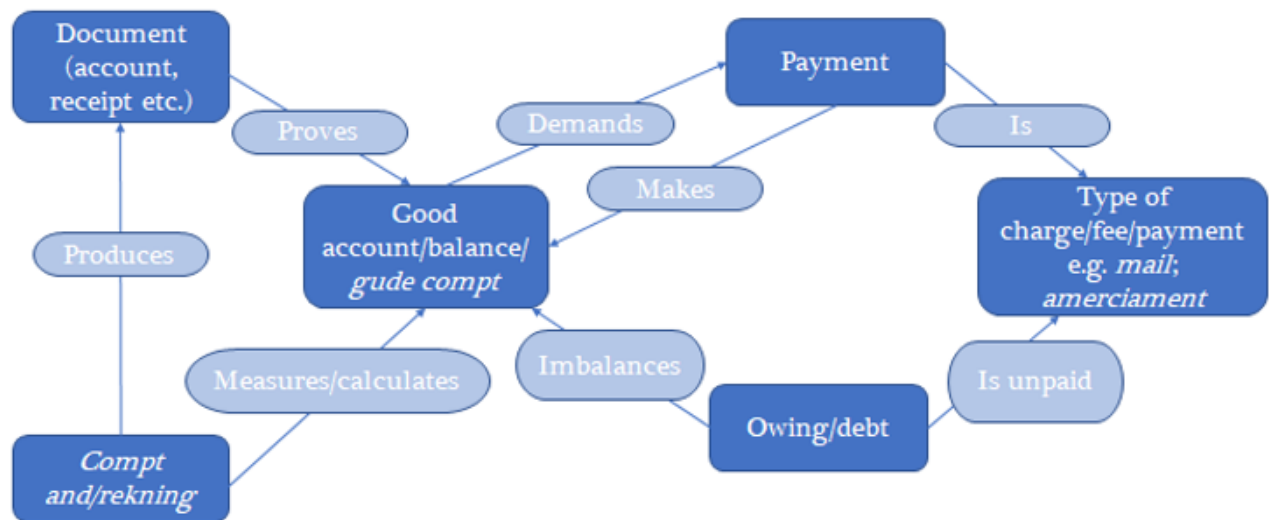


Figure 6. Conceptual context of 'good account'

Until now, the conceptual knowledge gathered in our hermeneutical exploration of the Aberdeen records was stored in various semi-structured forms, such as lists, tables, schemas such as the figure above, and, most importantly, the scholarly head. In order to make this knowledge explicit in an incremental fashion, we have formalized and integrated relevant concepts into the first version of a SKOS repository. This repository models the knowledge in a principled way, and will allow us to consolidate, fine-tune and extend our conceptualization of the domain in future hermeneutical analysis cycles. It will also provide a basis for comparison with the conceptual structure of the Augsburg records. The repository is available online: <https://www.textdimensions.eu/FLAG/ARO/FLAG1.rdf>.

58

## Summary of historical research outcomes

The broad conclusion drawn in relation to the domain of 'Budget' as found in the Aberdeen records, based on results from NLP incorporated into, and enhanced by, broader subject knowledge and interpretation, is as follows. While most of the records in the text are not expressly financial or budgetary (except for, arguably, the few financial accounts that appear or the statutes relating to the urban economy), within the bulk of the records whose express purpose relates more directly to the domain of 'Order' lies a dense substratum of transactions, debts and payments due, bound together with an ideology of 'good account' tied to the foundational identity of the burgh (see figure 6 above). The foundation of burghs could be described as a transaction between those who inhabited the burgh and enjoyed its privileges – the burgesses – and the burgh overlord, in Aberdeen's case the king. It was also a transaction aimed at stimulating further transactions, whether in the form of international, national or regional trade. The civic government that emerged from Aberdeen's foundation as a burgh presided over trade transactions as well as a range of other transactions, such as conveyancing and the payment of employees. It acted as both a party to those transactions and an arbiter of transactions between others, and by these means was involved in a dense and diverse web of transactional activity. *Gude compt and rekning* emerges from the records as the phrase which best captures the ideal result of all these transactions. Not just a technical term, it had a moral weight, because of the nature of the environment in which transactions took place in Aberdeen, where people were likely to have interpersonal ties besides the transaction, and because of the spiritual connotations and relationship to wider ideas of accountability of the term. The term and ideal underpinning it were not exclusive to urban settings, but it is likely to have had a particular association with, and resonance for, urban dwellers, and particularly those in urban government, based on the extent of transactional activity within burghs and the encouragement to trade and transaction inherent in the foundation of burghs. That burghs had a special association with this ideal but still shared it with wider society provided a context in which burgesses could use their understanding of the concept, and the practical experience on which it was based, to feed into administrative discourses through the many administrative connections burghs had with the rural aristocracy, the church and royal government.

59

## Conclusion

In this paper we explored the potential of four NLP techniques by integrating them into a digital hermeneutical workflow within the field of historic enquiry. Our overall experience shows that NLP techniques offer the scholar different perspectives on the content of the text under consideration. These perspectives provide, in comparison with manual analysis, a richer amount of potentially relevant contextual information, and help focus scholarly attention on text elements that warrant further manual evaluation and analysis, in

60

order to arrive at a more precise conceptualization of the domain. When used appropriately, the outcomes of NLP tools provide useful material for scholarly exploration and conceptualization.

In contrast to the results from exclusively manual research methods, we find that digital hermeneutical methods can claim an accelerated coverage of the whole text under study, and a greater repeatability of the methodological workflow. In our case study, as well as discovering these findings in relation to the budget domain, the approach outlined here offered many avenues for further research, including lines of enquiry that link 'Budget' to the other domains identified in the project.

Our results also show that historical records that could be considered – to a certain degree - familiar or exhausted can be rapidly reassessed by these methods and potentially produce new conclusions about higher concepts underpinning repetitive and - at face value - mundane language.

From a methodological point of view, the key benefit of the described approaches was the ability for the subject scholar to work with the grain rather than against, that is to say, the ability to work in a hermeneutical manner with a large body of repetitive material and find prevailing patterns, rather than looking past that material to find unusual material which is less representative of the text as a whole. For instance, questions about the context of the most commonly-occurring terms could be asked rapidly without a large time investment, allowing the experimental pursuit of lines of research which normally would not be asked by manual research because of the imbalance between the time investment required and the chance of producing meaningful results. Here the risk of not producing results was minimal as, for instance, searches using AntConc could be carried out quickly, leaving plenty of time for iteration or moving on to other lines of enquiry.

Our observations indicate that automatic analysis on its own, fully relying on distant reading, will never produce required scholarly results. Automatic techniques only offer a reductionist, and often statistically biased hermeneutical perspective of the text based on word distribution. They can, as agents within a dialectic workflow, only provide suggestions and influence research questions in each workflow cycle.

Our mixed methods inclusion of both automatic and manual analysis involves tools that enable a flexible shift between distant and close reading. The cyclical nature of the sequential stages of analysis shows the complementarity of scholarly manual qualitative analysis on the one hand, and more quantitatively oriented NLP analysis results on the other. This will, in resonance with the knowledge already accumulated by the scholarly expert, minimize the effects of the biases inherent in results from distant reading, and maximize the quality and coverage of research results.

The method described in this contribution explicitly assumes that qualitative scholarly interpretation is the most fundamentally important method to be applied for answering research questions, while computational approaches can only make potentially useful contributions to this process in a subservient role. In this day and age of AI, we tend to increasingly lose sight of the role of the computer in our lives. For a hermeneutical approach to text interpretation, it must remain clear that the scholar is king and the computer is strictly restricted to its assisting function. This advisory, ancillary aspect of computer analysis fosters a meaningful methodological integration of man and computer. A clear division of labour ensures maximum scholarly control of research activity and strategy, and forces automatic computational text analysis in the humanities to be minimally intrusive and overbearing. We hope to have demonstrated that in this way scholar and computer can collaborate successfully, with the scholar firmly in the driving seat.

## Acknowledgment

This work was supported by the FLAG project (<https://flag-project.uni-mainz.de/home/>), an international AHRC and DFG collaborative project involving the University of Aberdeen (UK) and the Johannes Gutenberg-Universität Mainz (DE). The authors wish to thank Prof. Jackson Armstrong, Prof. Dr. Joerg Rogge and Dr. Regina Schaefer for their comments and feedback.

## Appendix A: AI generated Python script for topic extraction

Perplexity (<https://www.perplexity.ai/>) prompt: "Write a python script that ingests a text file, tokenizes it and computes 10 topic clusters of 10 unique tokens each using Gensim latent dirichlet allocation."

Be aware that AI engines do not always return the same answers if you pose the same question more than once.

```
# start script
```

```
from gensim import corpora
```

```
from gensim.models import LdaModel
```

```
import re
```

```
# Step 1: Read and preprocess the text file
```

```
def preprocess(text):
```

```

# Lowercase, remove non-alphabetic characters, and split
text = text.lower()
text = re.sub(r'^a-z\s', '', text)
return [word for word in text.split() if len(word) > 2]
with open('your_text_file', 'r', encoding='utf-8') as f:
documents = [preprocess(line.strip()) for line in f if line.strip()]
# Step 2: Create dictionary and corpus
dictionary = corpora.Dictionary(documents)
corpus = [dictionary.doc2bow(doc) for doc in documents]
# Step 3: Train LDA model
lda_model = LdaModel(corpus=corpus, id2word=dictionary, num_topics=10, passes=10, random_state=42)
# Step 4: Print topics with 10 unique tokens each
for idx, topic in lda_model.show_topics(num_topics=10, num_words=10, formatted=False)
print(f"Topic #{idx+1}:")
print(", ".join([word for word, _ in topic]))

```

## Notes

- [1] FLAG Project: see <https://flag-project.uni-mainz.de/> and <https://aberdeenregisters.org/flag-project/>
- [2] Aberdeen Registers Online: 1398-1511 (<https://www.abdn.ac.uk/aro>) [hereafter ARO]
- [3] Search Aberdeen Registers (<https://sar.abdn.ac.uk/>)
- [4] Search Aberdeen Registers (<https://sar.abdn.ac.uk/>).
- [5] Laurence Anthony, 2018, *AntConc 3.5.7*, Tokyo, Japan, Wasdea University (<https://www.laurenceanthony.net/software/antconc/>)
- [6] A useful guide to AntConc for historians by Heather Froehlich is available at Programming Historian (<https://programminghistorian.org/en/lessons/corpus-analysis-with-antconc>).
- [7] <https://pypi.org/project/gensim/>.
- [8] <https://pypi.org/project/gensim/>.
- [9] <https://www.perplexity.ai/>
- [10] <https://www.nltk.org/>.
- [11] E.g. ARO-6-0976-02
- [12] ARO-8-0230-03
- [13] ARO-7-0285-04
- [14] ARO-8-0230-01; ARO-8-0230-02

## Works Cited

- Anthony 2018** Anthony, L. (2018) *AntConc*. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software>
- Bartlett 1994** Bartlett, R. (1994) *The making of Europe: Conquest, colonization and cultural change, 950-1350*. London, Penguin Books.
- Bawcutt 1998** Bawcutt, P. (1998) *The poems of William Dunbar, 2 vols*. Glasgow: Association for Scottish Literary Studies.
- Blei 2012** Blei, D. (2012) "Probabilistic Topic Models", *Communications of the ACM*, 55 (4): 77–84. doi: 10.1145/2133806.2133826. Available at <http://www.cs.princeton.edu/~blei/papers/Blei2012.pdf>.
- Brett 2012** Brett, M. R. (2012) "Topic modeling: A basic introduction", *Journal of Digital Humanities*, Vol. 2, No. 1. Available at: <http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/>.
- Carrier 2021** Carrier, J.G. (2021) *Economic Anthropology* (Newcastle upon Tyne).
- Creswell 2009** Creswell, J.W. (2009) *Research design: Qualitative, quantitative, and mixed methods approaches (3rd edition)*. Thousand Oaks, CA: Sage.
- Dickinson 1957** Dickinson, W.C. (1957) "Early Records of the Burgh of Aberdeen, Edinburgh: Scottish History Society".
- Epurescu-Pascovic 2020** Epurescu-Pascovic, I. (2020) "From the auditing of accounts to institutional accountability in late Medieval Europe", in Epurescu-Pascovic, I. (ed.) *Accounts and accountability in late Medieval Europe: Records, procedures, and socio-political impact*, 1-19. Available at: <https://www.brepolonline.net/doi/abs/10.1484/M.USML-EB.5.120735>.

- Fickers, Tatarinov, and van der Heijden 2022** Fickers, A., Tatarinov, J. and van der Heijden, T. (2022) "Digital history and hermeneutics – between theory and practice: An introduction", in Fickers, A. and Tatarinov, J. (eds.), *Digital history and hermeneutics: Between theory and practice*. Berlin, Boston: De Gruyter 2022. Available at: <https://doi.org/10.1515/9783110723991>.
- Firth 1957** Firth, J.R. (1957) *Papers in linguistics 1934–1951*. London: Oxford University Press.
- Foote 2019** Foote, K.D. (2019) A brief history of natural language processing (NLP). Available at: <https://www.dataversity.net/a-brief-history-of-natural-language-processing-nlp/>. Accessed November 03, 2023.
- Gemmill and Mayhew 1995** Gemmill, E. and Mayhew, N. (1995) *Changing values in Medieval Scotland: A study of prices, money, and weights and measures*. Cambridge Press.
- Gibbs and Owens 2013** Gibbs and Owens (2013) "The hermeneutics of data and historical writing", in Dougherty, J. and Nawrotzki, K. (eds.) *Writing history in the digital age*. E-book, Ann Arbor, MI: University of Michigan Press. DOI: 10.3998/dh.12230987.0001.001. Available at: <https://www.digitalculture.org/books/writing-history-in-the-digital-age/>.
- Hayles 2012** Hayles, K.N. (2012) *How we think: Digital media and contemporary technogenesis*. Chicago: University of Chicago Press.
- Kopaczyk 2009** Kopaczyk, J. (2009) "Multi-word units of meaning in 16th-century legal scots", in R. W. McConchie, Alpo Honkapohja, and Jukka Tyrkkö (eds) *Selected proceedings of the 2008 Symposium on New Approaches in English Historical Lexis*. Somerville, 2009, 88-95.
- Makkreel 2021** Makkreel, R. (2021) "Wilhelm Dilthey", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.). Available at:
- Mallery, Hurwitz, and Duffy 1986** Mallery, J.C., Hurwitz, R., and Duffy, G. (1986). "Hermeneutics: from textual explanation to computer understanding?" *AI Memos*, Memo AIM-871. <http://hdl.handle.net/1721.1/6438>.
- Manning and Schütze 1999** Manning, C. and Schütze, H. (1999). *Foundations of statistical natural language processing*. Stanford University and Xerox PARC. Cambridge, MA: The MIT Press.
- Marshall 2011** Marshall, A. (2011). "Augustine's Hermeneutics in a Modern Context". Society of Biblical Literature New England Regional Meeting, April 2011, Boston, Massachusetts. Available at: <https://apmarshall.files.wordpress.com/2012/11/augustine-hermeneutics-sbl.pdf>.
- Moretti 2005** Moretti, F. (2005). *Graphs, maps, trees: Abstract models for a literary history*. London: Verso.
- Parks and Peters 2022** Parks, L. and Peters, W. (2022) "Natural language processing in mixed-methods text analysis: A workflow approach", *International Journal of Social Research Methodology*, 26. 1-13. doi: 10.1080/13645579.2021.2018905.
- Peters and Wyner 2016** Peters, W. and Wyner, A. (2016). *Legal text interpretation: Identifying Hoffeldian relations from text*, *Proceedings of LREC 2016*.
- Peters, Parks, & Lennan 2019** Peters, W., Parks, L. and Lennan, M. (2019) "Integrating language technology into scholarly research Workflows". In Lana Pitcher and Michael Pidd (eds.) *Proceedings of the Digital Humanities Congress 2018, Studies in the Digital Humanities*. Sheffield: The Digital Humanities Institute. Available at: <https://www.dhi.ac.uk/openbook/chapter/dhc2018-peters>>.
- Ramsay 2010** Ramsay, S. (2010) "The hermeneutics of screwing around: or what you do with a million books". Available at: <https://libraries.uh.edu/wp-content/uploads/Ramsay-The-Hermeneutics-of-Screwing-Around.pdf>.
- Ramsay 2011** Ramsay, S. (2011) *Reading machines: Toward an algorithmic criticism*. Urbana: University of Illinois Press.
- Regan 2012** Regan, P.J. (2012) "Hans-Georg Gadamer's philosophical hermeneutics: Concepts of reading, understanding and interpretation", in: *Meta: Research in Hermeneutics, Phenomenology, and Practical Philosophy vol. IV. No 2*. Available at: [www.metajournal.org](http://www.metajournal.org).
- Rockwell 2003** Rockwell, G. (2003). "What is text analysis, really? Literary and linguistic computing", 18 (2), 209–19.
- Romein et al. 2020** Romein C.A., Kemman M, Birkholz J.M., Baker J., de Grijter M., Meroño-Peñuela A., (2020) "State of the field: digital history", *HISTORY*, 105(365): 291–312. Available at: <https://doi.org/10.1111/1468-229X.12969>
- Romele, Severo, and Furia 2020** Romele, A., Severo, M. and Furia, P. (2020) "Digital hermeneutics: From interpreting with machines to interpretational machines", *AI & SOCIETY*, 35. Available at: [https://www.researchgate.net/publication/326091804\\_Digital\\_Hermeneutics\\_From\\_Interpreting\\_with\\_Machines\\_to\\_Interpretational\\_Machines](https://www.researchgate.net/publication/326091804_Digital_Hermeneutics_From_Interpreting_with_Machines_to_Interpretational_Machines).
- Scheff 2015** Scheff, Thomas J. (2015) *Goffman unbound!: A new paradigm for social science*. New York: Routledge.
- Shadrova 2021** Shadrova, A. (2021) "Topic models do not model topics: epistemological remarks and steps towards best practices", *Journal of Data Mining & Digital Humanities*. Available at: <https://doi.org/10.46298/jdmdh.7595>
- Steinmetz, Freeden, and Fernández Sebastián 2017** Steinmetz, W., Freeden, M., & Fernández Sebastián, J. (eds.) (2017) *Conceptual history in the European space*. New York: Berghahn.
- Stuart 1844** Stuart, J. (ed.) (1844) *Extracts from the Council Register of the Burgh of Aberdeen, 1398-1570*. Aberdeen: Spaldin Club.
- Woolf and Silver 2018** Woolf, N., Silver, C. (2018) *Qualitative analysis using ATLAS.ti*. New York: Routledge. Available at:

**Zaagsma 2013** Zaagsma, G., (2013) "On digital history", *BMGN - Low Countries Historical Review*, 128(4), pp.3–29.  
<http://doi.org/10.18352/bmgn-ichr.9344>.

**van Zundert 2016** v. Zundert, J. J. (2016) "Screwmenetics and hermeneumericals: The computability of hermeneutics", in S. Schreibman, R. Siemens, & J. Unsworth (eds.) *A New Companion to Digital Humanities*, pp. 331-347. Wiley-Blackwell.

## Recommendations

DHQ is testing out three new article recommendation methods! Please explore the links below to find articles that are related in different ways to the one you just read. We are interested in how these methods work for readers—if you would like to share feedback with us, please complete our short evaluation survey. You can also visit our documentation for these recommendation methods to learn more.

### SPECTER Recommendations

Below are article recommendations generated by the SPECTER model:

1. The First Steps of Digital Humanities: A Review of Tara L. Andrews and Caroline Macé's Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches, 2021, Giulia Freni, Classics Department, University of Siena
2. Developing Geographically Oriented NLP Approaches to Sixteenth–Century Historical Documents: Digging into Early Colonial Mexico, 2020, Diego Jiménez–Badillo, Museo del Templo Mayor, Instituto Nacional de Antropología e Historia; Patricia Murrieta–Flores, Digital Humanities Hub–History Department, Lancaster University; Bruno Martins, The Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa, INESC–ID, University of Lisbon; Ian Gregory, Digital Humanities Hub–History Department, Lancaster University; Mariana Favila–Vázquez, Museo del Templo Mayor, Instituto Nacional de Antropología e Historia; Raquel Liceras–Garrido, Digital Humanities Hub–History Department, Lancaster University
3. Introduction: History and Data Science: Data-driven Inquiries into the Past, 2026, Gabor Mihaly Toth, Center for Contemporary and Digital History, University of Luxembourg
4. A Critical Collection History of Nineteenth-century Women's Letters: Overcoming the Occluded Archive with Data-Driven Methods, 2026, Ilona Pikkanen, The Finnish Literature Society; Matti La Mela, Uppsala University; Hanna-Leena Paloposki, Independent Scholar; Jouni Tuominen, University of Helsinki and Aalto University
5. Slow, Painful and Expensive: Current Challenges in Text-Mining Corpus Construction for the Digital Humanities, 2025, Matt Warner, Stanford University; Nichole Nomura, University of Wyoming; Carmen Thong, Stanford University; Alix Keener, Stanford University; Alexander Sherman, University of Texas, Austin; Gabi Birch, Stanford University; Maciej Kurzynski, Lingnan University; Mark Algee-Hewitt, Stanford University

### DHQ Keyword Recommendations

Below are article recommendations generated by DHQ Keywords:

1. Where's the ML in DH? And Where's the DH in ML? The Relationship between Modern Languages and Digital Humanities, and an Argument for a Critical DHML, 2017, Thea Pitman, University of Leeds; Claire Taylor, University of Liverpool
2. The Boundless Book: A Conversation between the Pre-modern and Posthuman, 2013, Alison Tara Walker, Saint Louis University
3. A Genealogy of Distant Reading, 2017, Ted Underwood, University of Illinois, Urbana-Champaign
4. Cyberinfrastructure for Classical Philology, 2009, Gregory Crane, Tufts University; Brent Seales, University of Kentucky; Melissa Terras, University College London
5. "Open" or "Close" Research Instruments? Conflicting Rationales in the Organization of Early Digital Medieval History in Europe (1960–1990)., 2024, Edgar Lejeune, Vossius Center for the History of Humanities and Sciences (University of Amsterdam)

### TF-IDF Recommendations

Below are article recommendations generated by the TF-IDF Model:

1. Tool criticism in practice. On methods, tools and aims of computational literary studies, 2023, J. Berenike Herrmann, Universität Bielefeld; Anne-Sophie Bories, University of Basel; Francesca Frontini, CNR - Istituto di Linguistica Computazionale A. Zampolli; Clémence Jacquot; Steffen Pielström, University of Würzburg; Simone Reborá, Johannes Gutenberg University Mainz; Geoffrey Rockwell, University of Alberta; Stéfan Sinclair
2. In a test bed with Kafka. Introducing a mixed-method approach to digital stylistics, 2018, J. Berenike Herrmann, Basel University
3. Categorising Legal Records – Deductive, Pragmatic, and Computational Strategies, 2023, Marlene Ernst, University of Passau; Sebastian Gassner, University of Passau; Markus Gerstmeier, University of Passau; Malte Rehbein, University of Passau
4. Facets of Friction: Investigating epistemological friction between computing and the humanities to support Digital Humanities computing education, 2026, Anna Sollazzo, University of Glasgow
5. Visualization of Categorization: How to see the wood and the trees, 2023, Ophir Münz-Manor, The Open University of Israel; Itay Marienberg-Milikowsky, Ben-Gurion University of the Negev

