Qualitative Space in Digital Diplomatics

Mark Bell, Eirini Goudarouli, John Stell, Jane Winters

Introduction

The digitisation of collections held by cultural heritage institutions has transformed research in the late 20th and early 21st century. It has enabled the exploration and study of collections in new ways. However, the variety of formats, scale, complexity and heterogeneity of digital collections have raised a number of challenges in the existing modes of access to these collections.

Providing novel modes of access to digital collections has been a central strategic priority for cultural heritage institutions around the world as they reimagine the ways that digital collections can be used by the public[1]. At the same time, there has been a growing emphasis within the academic and research communities on the exploration, complex analysis and interpretation of large-scale digital heritage collections, using advanced computational techniques (e.g. text mining, machine learning, network analysis).

Existing access techniques are unable to deliver the views of digital collections that vast and complex new digital archives require. For example, catalogues and keyword search do not readily support analysis of the spatial and temporal aspects of primary source materials or the relationships present within and between collections. Additionally, keyword searching allows only the study of words in isolation (or with some limited context), making it difficult to view large-scale collections as connected wholes, rather than as multiplicities of separate fragments.

This paper argues that being able to go beyond keyword searching of digital collections will enable the exploration and identification of evidence based on their qualitative spatial and temporal relationships, offering exciting new avenues to researchers in the humanities and social sciences. Moving past keyword searching also

1 For example, https://www.nationalarchives.gov.uk/about/our-role/plans-policies-performance-and-projects/our-plans/our-digital-cataloguing-practices/project-omega/.

https://doi.org/10.7788/9783412530921 CC-BY-NC 4.0

has the potential to open up new directions and generate new perspectives for researchers working in digital fields (such as computer scientists, interface designers, user experience researchers and digital archivists).

It also suggests new ways of exploring heritage collections as traces of dynamic processes and approaching documents as processual narratives, focusing on their spatial structure and interconnections. This approach allows researchers to reconstruct how documents came into being; to tell stories about when, how and by whom they were created; and to uncover the complex relationships both between the constitutive elements of an individual document and between documents within an archive. The spatial relationships in documents, such as where pieces of text are placed on a page, and the temporal relationships, such as knowing that one annotation was added after another, contribute significantly to meaning and interpretation. Relationships between multiple documents form complex networks, which shed light on how they have been created, used and reused. These networks include relationships such as the spatial (within a building or collection) and the temporal (between authors, versions or dates of accession to a collection).

The paper aims at using all these relationships as an aid to understanding and navigation. To do so, it will introduce Qualitative Spatial Representation (QSR) as a method to explore the spatial and temporal dimensions of digital document collections, focusing on the structure of the documents themselves and how they may have developed over time. The spatial arrangement of different parts of a document is often a key aspect of both original significance and the history of changes through time.

The layout of a document often conveys structure and meaning; consider how space is used to separate paragraphs, data is presented in regular columns, or the elements of an item of correspondence (address, salutation, signature) are positioned on the page. Werner analysed the relationship between visual aspects of papal documents and their intended recipients[2]. Spatial features examined included percentage of the page covered by text, relative distances of components of the documents, margin widths, and the size of capital letters. The interpretation of annotations depends on their placement in relation to both prior and subsequent writing. The positioning of individual signatures can denote relationships of importance. While experts can make judgements about the visual elements of an individual page, digital humanists are applying computational methods to large-scale collections to extract and analyse content. The field of document layout analysis has provided many methods for analysing document structure. Using computational techniques such

2 Judith WERNER, Gestaltung der Urkunde, Gestaltung Europas – Urkundenlayout zwischen europäischem Empfängereinfluss und päpstlicher Vereinheitlichung, in: Papstgeschichte im digitalen Zeitalter. Neue Zugangsweisen zu einer Kulturgeschichte Europas, ed. Klaus HERBERS/ Viktoria TRENKLE (Beihefte zum AKG 85), Cologne 2018, p. 109–134.

as machine learning and computer vision, vast amounts of data can be extracted from archival collections of documents, but these techniques can sometimes leave a large gap between essentially numerical data and the ways in which human experts describe documents in academic discourse. One way to bridge this gap is through the use of QSR which concerns relationships between entities in space. It is an important building block of Commonsense AI[3], combining mathematical logic and commonly understood concepts and language. This area, within Artificial Intelligence and Knowledge Representation, links digital representation with the kinds of spatial relationships that humans use in informal descriptions.

Children from the age of five in the UK are taught 'positional language' and 'repeating patterns'[4], both of which are used to understand spatial relations in documents. Examples include: *overlapping* (of some text and an annotation); *being to the left of* (of one block of text and another); *being underneath* (of the vertical arrangement of text and image, or a line beneath a word); *sequence of evenly spaced lines* (a possible definition of a block of text). Qualitative relationships have been used to represent humans' knowledge of space[5], and reasoning with spatial relationships is part of QSR (the R often signifying both representation and reasoning). Temporal relationships can be included; conceptualising time spatially as a line[6] or more complex branching structures with uncertain relationships between some events. A physical document is often both evidence of a process (accumulating annotations, additions, etc.) and a component in other processes (creating new versions, generating responses, political interactions, etc.). Exploring archival collections as traces of dynamic processes is limited without the ability to describe the spatial structure of successive changes. Machine learning can help to identify that a piece of text is an annotation, but it is QSR which can help us tie that annotation spatially to the text it addresses. These spatial connections can be essential for reading documents as narratives of processes.

While machine learning can achieve remarkable results, a well-known drawback is the lack of explanation associated with how those results are derived. This has led to the search for 'explainable AI' (XAI) as a way to allow decisions to be questioned and reduce the likelihood of bias due to unrepresentative training data. There is an increasing focus in archives on the importance of XAI for maintaining the trans-

3 Ernest DAVIS/Gary MARCUS, Commonsense reasoning and commonsense knowledge in artificial intelligence, in: Communications of the ACM 58/9 (2015) p. 92–103.

4 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335158/PRIMARY_national_curriculum_-_Mathematics_220714.pdf.

5 Anthony COHN/Jochen RENZ, Qualitative Spatial Representation and Reasoning, in: Handbook of knowledge representation, Elsevier, eds. Frank VAN HARMELEN, Vladimir LIFSCHITZ and Bruce PORTER, Amsterdam 2008, p. 551–596.

6 James F. ALLEN, Maintaining Knowledge about Temporal Intervals, in: Communications of the ACM 26 (1983) p. 832–843.

parency of process that underpins trust in the documentary record. In their consideration of AI in the archive, Rolan et al. 2019 distinguish between rule-based systems, statistical models and deep learning models, noting that the two former offer greater transparency and are more susceptible to inspection. Bunn[7] takes this framework as a starting point for considering the significance of XAI for recordkeeping and archival practice, highlighting 'the vital importance of transparency to the proper functioning of societies'. These considerations have influenced the choice of method explored in this chapter.

The research field of 'neuro-symbolic AI'[8] aims to combine the advantages of machine learning, the neural part, with logic and knowledge representation, the symbolic part. To adapt a simple example from an account of DARPA's XAI program[9] consider the task of classifying a portion of a document as either a later annotation or part of the original document. A purely machine learning approach might learn a classification function and be able to report 'later annotation, probability 0.87' based on the examples used in learning. An ideal XAI approach would still learn from examples but, instead of just learning a classification function, would learn a structured model of the process of classification. This model could be used to explain that 'later annotation' was chosen because (for example) the portion was overlapping a larger area of text on the page and was at an angle to it. This explanation uses the relationships: 'larger than', 'overlapping' and 'at an angle to'. To deliver explainable AI like this it is necessary to determine the appropriate relationships. In addition, the logical properties of these relationships are important to support deductive reasoning within explanations. Working with documents, spatial facts such as 'If two things are next to each other then they have nothing between them', can be used to build explanations.

So far, there is existing use of spatial relationships in document processing and document understanding. However, these are essentially restricted to identifying document genre and reading-order for text blocks[10]. The use of structure and layout can improve the quality of textual search results[11], but this does not treat the visual

7 Jenny Bunn, Working in contexts for which transparency is important: a recordkeeping view of explainable artificial intelligence (XAI), in: Records Management Journal 30 (2020) p. 143-153.

8 Pascal Hitzler et al., Neuro-symbolic approaches in artificial intelligence, in: National Science Review 9/6 (2022), https://doi.org/10.1093/nsr/nwac035; Vaishak Belle, Logic meets Learning: From Aristotle to Neural Networks, in: Neuro-Symbolic Artificial Intelligence: The State of the Art, ed. Pascal Hitzler/Md Kamruzzaman Sarker, Amsterdam/Berlin Washington D.C. 2022, Chapter 3, p. 78–102.

9 David Gunning/David W. Aha, Darpa's Explainable AI (XAI) Program, in: AI Magazine 40/2 (2019) p. 44–58, fig 3.

10 Floriana Esposito et al., Machine learning for digital document processing, in: Machine Learning in Document Analysis and Recognition, ed. Simone Marinai/Hiromichi Fujisawa, Berlin/Heidelberg 2007, p. 105–138.

11 Marco Aiello, Document understanding for a broad class of documents, in: International Journal of Document Analysis and Recognition 5 (2002) p. 1.

aspects of a document as features of interest which encode meaning. They do not provide a way to help users locate documents based on criteria such as 'the spatial structure of these elements should be similar to this example', where spatial structure is specified qualitatively (e.g., 'I want this to appear in between these other elements'). QSR can be used as an aid to understanding and navigation of document collections but it is almost entirely unexplored in this context.

This paper will introduce the ideas of QSR and demonstrate the relevance of these techniques to digital diplomatics through examples drawn from ongoing work at The National Archives UK in which we have been evaluating the role of spatial relationships in the analysis of documents. It will also explore how the spatial (e.g., relationships between blocks of text) and temporal (e.g., visible editorial and transmission histories) aspects of documents and collections may be used to enhance navigation through and interpretation of digital archival materials.

Document Layout Analysis

Document Layout Analysis (DLA) is among the first operations in Document Image Analysis[12]. DLA systems are primarily aimed at analysing digital documents, such as academic papers in PDF format, but a further sub-field of DLA focuses on the analysis of historical documents.

A survey of papers published at ICDAR, the bi-annual International Conference for Document Analysis and Recognition, held since 1991, identified three main tasks being addressed by researchers: pre-processing for image quality, identification of regions with homogeneous content, and textual content recognition[13].

The second of these is most relevant to QSR, and within this category several sub-tasks were identified: text line segmentation, baseline detection, text detection, symbol detection. The authors further summarised the application of deep learning technology as mapping inputs (e.g., a digitised page) to outputs (page, region, table, text-line, bounding box, transcribed text, image patch, image), with the algorithms generally classifying at the pixel level. Eskenazi et al. separate identification into physical (typed, text, images, decorations) and logical/functional (header, footer, main body), and define a typology of segmentation algorithms with three classes: (1) algorithm constrained, i.e., segmentation of specific layouts without needing

12 Galal M. Binmakhashen/Sabri A. Mahmoud, Document Layout Analysis: A Comprehensive Survey, in: ACM Computing Survey 52/6 (2020), Article 109.

13 Francesco Lombardi/Simone Marinal, Deep Learning for Historical Document Analysis and Recognition – A Survey, in: Journal of Imaging 6/10 (2020) p. 110.

training; (2) parameter constrained (e.g., font size, connected components size); (3) potentially unconstrained (e.g. neural networks)[14].

Current state of the art approaches use deep learning to segment images. Mechi et al. used a U-Net, previously applied to medical image segmentation, for text line identification in historical documents[15]. Vogtlin et al. presented DIVA-VAF, a deep learning framework for separating text and side notes[16], while Pandey and Harit used top-down visual saliency models to identify four types of annotation: underlines, encirclements, inline text, marginal text[17]. Li et al. applied self-supervised learning to segmentation, using a cross-modal approach which incorporated textual content with semantic components (titles, text blocks)[18]. Wei et al. evaluated three machine learning classifiers for pixel level labelling of periphery, background, text, and decoration[19]. Machine learning models require datasets to train them, and Nikolaidou et al.[20] have surveyed 65 such datasets, classifying them by task. Of those related to document structure, 25 contain annotations for layout components, 27 for text-lines, 7 for tables, and 11 for graphics or images. Four include reading order. Examples of layout components include text regions, main body text, footnotes, page numbers, table of contents, marginalia, paragraphs. One dataset includes text features, such as italics or capitalisation, and separates handwritten from typed text. The dataset of the Oficio de Hipotecas de Girona includes a notarial typology, which can be identified by entries in the left margin aligned with the beginning of a

14 Sébastien Eskenazi/Petra Gomez-Kramer/Jean-Marc Ogier, A comprehensive survey of mostly textual document segmentation algorithms since 2008, in: Pattern Recognition 64 (2017) p. 1–14.

15 Olfa Mechi/Maroua Mehri/Rolf Ingold/Najoua Essoukri ben Amara, Text Line Segmentation in Historical Document Images Using an Adaptive U-Net Architecture, in: International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia (2019) p. 369–374.

16 Lars Vogtlin/Paul Maergner/Rolf Ingold, DIVA-DAF: A Deep Learning Framework for Historical Document Image Analysis; State of the art results on separating text and side notes, 2022. (Accessed at https://arxiv.org/abs/2201.08295.

17 Shilpa Pandey/Gaurav Harit, Handwritten Annotation Spotting in Printed Documents Using Top-Down Visual Saliency Models, in: ACM Transactions on Asian and Low-Resource Language Information Processing 21/3 (2022) Article 58.

18 Peizhao Li/Jiuxiang Gu/Jason Kuen/Vlad I. Morariu/Handong Zhao/Rajiv Jain/Varun Manjunatha/Hongfu Liu, SelfDoc: Self-Supervised Document Representation Learning, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA (2021) p. 5648–5656.

19 Hao Wei/Micheal Baechler/Fouad Slimane/Rolf Ingold, Evaluation of SVM, MLP and GMM Classifiers for Layout Analysis of Historical Documents, in: International Conference on Document Analysis and Recognition, Washington D.C. 2013, p. 1220–1224.

20 Konstantina Nikolaidou/Mathias Seuret/Hama Mokayed, A survey of historical document image datasets, in: International Journal on Document Analysis Recognition 25 (2022) p. 305–338.

paragraph[21]. This is an example of semantic meaning of text being inferred from its spatial relationship with other text.

Earlier attempts at document layout analysis did use logic-based approaches. Esposito used spatial relations (top and left alignments, and touching), and qualitative size (small, medium, large), within a rule-based algorithm, combining with machine learning to optimally define the size categories. Aiello extended Allen Intervals to define Thick Boundary Rectangle Relations[22] (TBRR) in order to smooth the unevenness of text lines in historical documents. As an example, take two parallel text lines, A and B, where A is above B. Line A has x-coordinates of (100, 200) and B has x-coordinates (101, 201). Using Allen Intervals, the lines will be classified as overlapping, while to the human eye they will appear identical, since they differ by a single pixel at each end. TBRR provides a parameter, T, which defines how close two points can be to be considered equal (e.g. with T=2; 98–102 are all equal to 100).

Ishitani combined geometrical and logical layout analysis, using Gestalt principles of proximity, adjacency, and contextual similarity, to group objects on the page[23]. QSR has also been explored by Stell in relation to poetry where the layout can contribute key aspects of the meaning of the poem[24].

Controlled vocabularies of document features are available. Codicologia has a detailed vocabulary built from three databases (Vocabulaire codicologique, Le Lexicon[25], Arabic codicological glossary)[26]. Themes range from page layouts (long text, columns), and text regions (paragraphs, marginalia, lines) to signs such as brackets, underlines, and carets. SegmOnto aims to present a compromise between the detail of Codicologia and the limited range of features provided by PAGE XML (a standard schema for document layout), categorising layout features as zones and lines. Zone types include main text, margins, stamps, and tables, while line types include headings, drop capitals, and default text. Nikolaidou's[27] survey of datasets, and the

21 Lorenzo Quiros/Lluís Serrano/Vicente Bosch/Alejandro H. Toselli/Rosa Congost/Enric Saguer/Enrique Vidal, Oficio de Hipotecas de Girona. A dataset of Spanish notarial deeds (18th Century) for Handwritten Text Recognition and Layout Analysis of historical documents (1.0) [Data set] (2018); https://doi.org/10.5281/zenodo.1322666

22 Aiello, Document understanding p. 1.

23 Yasuto Ishitani, Logical structure analysis of document images based on emergent computation, in: Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR '99 (Cat. No.PR00318), Bangalore, India (1999) p. 189–192.

24 John G. Stell, Qualitative Spatial Representation for the Humanities, in: International Journal of Humanities and Arts Computing 13/1–2 (2019) p. 2–27.

25 Philippe Bobichon, Mise en page et mise en texte des manuscrits hébreux, grecs, latins, romans et arabes (Publications pédagogiques 5), Paris 2008.

26 http://codicologia.irht.cnrs.fr/theme/liste_theme/422.

27 Nikolaidou et al., A survey of historical document image datasets. https://doi.org/10.48550/arXiv.2203.08504.

literature on algorithmic approaches to DLA, suggests that the field has so far been focused on identification of major features of documents – paragraphs, lines, images – rather than the larger range of philological features included in Codicologia. Although some approaches identified have used spatial logic, they have used it as a means to an end–identifying region order, or classifying documents by type–not as a representation of the document itself.

Qualitative Spatial Representation (QSR)

QSR can be traced back to philosophical objections to a mathematical theory of space based on abstract points and uninformed by human experience. Whitehead notes that geometry as a deductive science is not grounded in the physical world, and asks 'How is space rooted in experience?'[28]. Whitehead was not the first to raise the mismatch between mathematical space and the space of experience. However, the theory of extensive connection[29] developed by Whitehead provided the basis for the development of the Region-Connection Calculus (RCC)[30]. The RCC consists of axioms for spatial regions, based on relationships between regions. It is generally applied to two-dimensional regions in the plane with relationships such as overlapping, touching at their boundaries, etc. Regions in this sense can be areas of text on a page, not necessarily rectangular nor of a single piece. The most commonly used form, RCC-8, includes eight relationships. Studies by Renz, Rauh and Knauff provide evidence that topological relations are important in human understanding of space, and 'the level of granularity given by the RCC-8 relations is particularly important'[31]. Each relationship consists of a precise mathematical definition, and a human interpretable, qualitative, description. For example, the relationship 'partially overlapping' is defined as 'there exists a part of a region X which is also a part of region Y and there is also some part of X which is not a part of Y'. The term 'part' also has a precise definition.

The relationship can also be more intuitively understood using circles to indicate two regions [Fig. 1]:

28 Alfred North WHITEHEAD, An Enquiry Concerning the Principles of Natural Knowledge, Cambridge [2]1925, p V.

29 Alfred North WHITEHEAD, Process and Reality. An Essay in Cosmology, Cambridge 1929.

30 David A. RANDELL/Zhan CUI/Anthony G. COHN, A spatial logic based on regions and connection, in: Knowledge Representation 92 (1992) p. 165–176.

31 Jochen RENZ/Reinhold RAUH/ Markus KNAUFF, Towards Cognitive adequacy of topological spatial relations, in: ed. Christian FRESKA et al., Spatial Cognition II, Lecture notes in Artifical Intelligence 1849 (2000) p. 184–197, p 194.

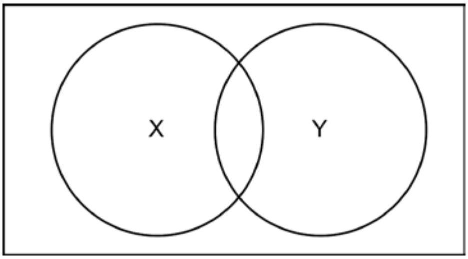


Fig. 1: Region X partially overlaps region Y.

For relationships between one-dimensional time intervals, the algebra of Allen[32] pre-dates the RCC and the development of computational QSR, but can now be understood as another system of relationships in the QSR mould. In this case the definition of two overlapping time intervals X and Y is that 'the beginning of Y occurs after interval X starts, but before it ends, and Y ends after X ends'. Again, a pictorial representation, this time using lines, is more intuitive [Fig. 2]:

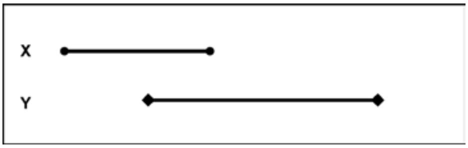


Fig. 2: Interval X overlaps interval Y.

The inverse of the 'overlaps' relationship is 'overlapped by'. In this case Y would begin before X and end during the interval X. The Allen algebra comprises 13 relationships: 6 (before, meets, starts, finishes, during, overlaps) plus their 6 inverses (after, met by, started by, finished by, contains, overlapped by), and the 'Equals' relationship.

Several extensions to Allen's original relations have been proposed, and we will make use of Thick Boundary Rectangle Relations[33] (TBRR) and Extended Allen Intervals[34] (EAI) in the experimentation section. TBRR recognises the imprecision of historical documents. While a paragraph may be left justified, the alignment of the beginnings of each line is generally approximate, especially at the pixel level, and may be further affected by skew from the digitisation process. To counter this imprecision the extremes of two intervals are considered equal if they are closer than a fixed distance T (setting T=0 is the equivalent of standard Allen intervals).

32 James F. Allen, Maintaining Knowledge about Temporal Intervals, in: Communications of the ACM 26 (1983) p. 832–843, p. 832.

33 Aiello, Document understanding p. 1

34 Renz, Towards Cognitive adequacy.

Returning to Fig. 2, if the beginning of Y was within distance T of the beginning of X, then the relationship under TBRR would be 'X starts Y'. The best value of T for an application may be identified through experimentation or derived using statistical or machine learning methods.

One criticism of Allen intervals is that they do not capture the degree of overlap; there is no distinction between two-one minute intervals overlapping by 59 seconds or by a single second. EAI addresses this issue by using the midpoint of each interval to make 27 relations. For example, the 'overlaps' relation becomes four relations which express the degree of overlap (o) by introducing the terms 'most part' (m) and 'least part' (l). These relations are: most part of X overlaps most part of Y (mom), least part of X overlaps least part of Y (lol), most part of X overlaps least part of Y (mol), least part of X overlaps most part of Y (lom). Fig. 3 shows the four overlapping relationships.

We also propose two simplified versions which will be referred to as Allen-2 ('overlap', 'no overlap') and Allen-3 ('before', 'overlap', 'after'). Following this naming convention, the standard Allen algebra will be referred to as Allen-13, the Extended version as Allen-27, and TBRR as Allen-13T.

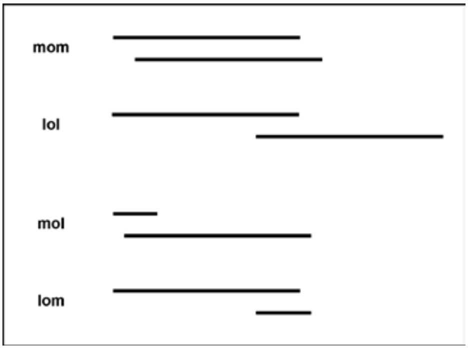


Fig. 3: Overlapping relations in EAI.

Allen intervals can also be applied to two-dimensional space. The Rectangle Algebra (RA) uses Allen-13 in both the x and y directions, which results in 169 possible relations between two rectangles. While this provides more nuance than RCC-8, it may make reasoning overly complex. For document features such as text lines, using Allen-2 (or 3) should be sufficient for one of the dimensions (vertical for text written across the page), reducing the algebra to 26 (or 39) relations. However, using Allen-27 for the other dimension increases would create up to 81 relations. The famous aphorism, 'All models are wrong, some are useful'[35], seems particularly appropriate when applying a qualitative representation. The algebra

35 George E. P. Box, Robustness in the strategy of scientific model building, in: Robustness in Sta-

used will always be a compromise between the nature of the problem being solved and the computational complexity of the solution.

Qualitative relationships have been analysed for directions such as 'left of', 'north of', etc, and also for relationships of distance such as 'near' and 'far'. These were not used in the experimentation reported here, but for a general overview of many spatial calculi see Cohn and Renz[36].

Applying QSR to Historical Documents

We have experimented with a number of documents to explore the value of QSR for identifying, categorising and perhaps even interpreting historical documents. Some are from the collections of the National Archives (TNA), and others come from the cBAD[37] document layout dataset. They have been selected to demonstrate the concepts of QSR but do not exhaustively represent the range of spatial features we may find in historical documents. The documents from TNA include medieval charters, which show how spatial relationships demarcate sections of a document, and incorporate visual features which group information together. Approaching charters in this way draws on the long tradition of diplomatics, a sub-discipline of archival science and legal theory that was developed to establish the authenticity of precisely such documents. Diplomatics plays an important role in the understanding of the form of individual documents, of what should and should not be present in a typical charter or contract from a particular time, region, or jurisdiction.

As Duranti notes, 'Any written document … contains information transmitted or described by means of rules of representation, which are themselves evidence of the intent to convey information: formulas, bureaucratic or literary style, specialized language … and so on'[38]. The present research is not concerned with authenticity, but it does seek to explore whether the same features that allow experts in diplomatics to assess whether a document is genuine can also shed light on document type at scale.

tistics, ed. Robert L. Launer/Graham N. Wilkinson, New York/San Francisco/London 1979, p. 201–236.

36 Anthony Cohn/Jochen Renz, Qualitative Spatial Representation and Reasoning, in: Handbook of knowledge representation, Elsevier, eds. Frank van Harmelen, Vladimir Lifschitz and Bruce Porter, Amsterdam 2008, p. 551–596.

37 Markus Diem/Florian Kleber/ Robert Sablatnig/Basillis Gatos, cBAD: ICDAR2019 Competition on Baseline Detection, in: International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia (2019) p. 1494–1498.

38 Luciana Duranti, Diplomatics. New Uses for an Old Science, Part I, in: Archivaria 28 (1989) p. 7–27, p. 15.

In addition to the medieval charters held at the National Archives, more modern documents have been examined, for example the drafts and final version of a speech made by Prime Minister Margaret Thatcher to President Ronald Reagan in 1982[39]. These pages are used to show how QSR can be deployed to establish and represent the evolution of a document over time. From the cBAD dataset we have chosen letters and documents with columnar layouts. This application of a common method over documents of widely differing types is again grounded in existing diplomatic practice. Writing in 1989, Duranti argued that 'notwithstanding the technical problems presented by some contemporary documents, the different structure of their text and the specific procedures governing their creation, maintenance, and use, the basic diplomatic principles and methodology formulated for the evaluation of medieval diplomas are still valid today'[40]. A charter might begin with an invocation and end with a dating clause and witness list, but an email is equally identifiable through the presence of a 'To' field containing at least one email address, CC/BCC fields, and a subject line (among other elements). Using form as a means of categorising, grouping, and connecting digitised and born-digital documents in heterogeneous collections would help historians to gain insights into the scope and shape of those collections and allow them to move between close, distant, and mid-scale reading.

QSR does not identify objects on a page it is solely a method for representing relations between them. Therefore we need other methods to identify the objects. The field of Document Layout Analysis has advanced significantly since the advent of deep learning and has produced many models for segmenting lines in digitised documents. We have applied one such model, ARU-Net[41], to identify baselines in our documents. Figs. 4(a)–(f) show the output of running ARU-Net on a selection of the documents we used. In many cases features of interest may be subtle, or beyond the capabilities of the current generation of machine learning models (or more likely, sufficiently large datasets of visual features to train them are unavailable). In preparing the examples for this paper we have also used the manual annotation tool VIA[42] to label lines and other features in documents. We found, for our purposes, that it was most pragmatic to draw baselines for text lines, and draw rectangles or polygons around symbols, underlinings, and other non-textual

39 https://discovery.nationalarchives.gov.uk/details/r/C13317582

40 Duranti, Diplomatics p. 23.

41 Olfa Mechi/Maroua Mehri/Rolf Ingold/Najoua Essoukri ben Amara, Text Line Segmentation in Historical Document Images Using an Adaptive U-Net Architecture, in: International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia (2019) p. 369–374.

42 Abhishek Dutta/Andrew Zisserman, The VIA Annotation Software for Images, Audio and Video, in: Proceedings of the 27th ACM International Conference on Multimedia (MM '19), October 21–25, Nice, France. ACM, New York, NY, USA, 4 pages, 2019.

annotations. We also used metadata labels to separate typed text, handwriting, and handwriting from different hands. From baselines we can generate rectangles, by naively filling the space vertically until encountering the next baseline, or by deciding on a standard height for each rectangle. A more sophisticated approach would use computer vision techniques to identify connected components above each baseline and derive a bounding box around them. Some segmentation models produce complex polygons, and jagged baselines, which closely follow the outline of the text. While these may benefit Handwritten Text Recognition (HTR) models, they introduce unnecessary computational complexity to deriving qualitative relationships. We therefore recommend simplifying the objects to rectangles and straight lines as much as possible.

Fig. 4(a)-(f) clockwise from top left: (a) Page from a 19th century Will; (b) 3 column page from the cBAD dataset; (c) Paragraphs with headings in left margin; (d) Letter from cBAD dataset; (e) poem in the form of a tree; (f) first page of a speech.

QS Representation of Documents

Working with rectangles suggests the Region Connection Calculus as an appropriate starting point for documents. However, the majority of rectangles would be non-overlapping, or may overlap but not in a meaningful way. To create an interval-based representation of a document, we firstly derive baselines which as data consist of a y-axis position (y-pos), and x-axis start (x-start) and end (x-end) positions. Each interval is compared to its nearest neighbours, based on vertical distance, and a relation is calculated between the first overlapping neighbour below, and to the right. If there is no overlapping neighbour below, then the first non-overlapping neighbour is used instead. A configurable limit can be set on vertical distance, so an interval may have no close neighbours.

In Table 1, we have classified the 27 relations of Allen-27 according to their directional tendency ('before' implies a large shift to the right from one text line to the next; 'equals', no horizontal shift; 'momi', a small shift leftwards), and the implied change in size ('cdi' implies the first line is longer than the next; 'cd' implies the first is shorter; 'mom' does not explicitly imply a difference in length). Table 2 shows the equivalent table of Allen-13 relations (TBRR shares these relations).

Table 1: Allen-27 relations by size effect and directional tendency.

stretch	lom	lf	rd	mf	cd	ms	ld	ls	mol
	after	met by	loli	momi	equal	mom	lol	meets	before
shrink	lomi	lsi	ldi	msi	cdi	mfi	rdi	lfi	moli
	← **Left**					**Right** →			

Table 2: Allen-13 relations by size effect and directional tendency.

stretch				f	d	s			
	pi	mi		oi	e	o		m	p
shrink				si	di	fi			
	← **Left**					**Right** →			

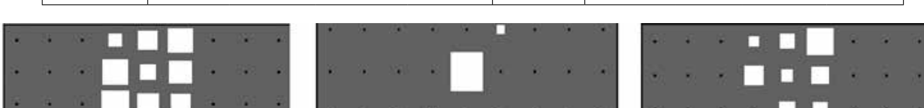


Fig. 5: Hinton diagrams of relations for Fig. 4(a), (page of continuous, even length lines of text). From left to right, Allen-13, Allen-13T (T=8), Allen-27.

A Hinton diagram is used to visualise the relative values in a 2-D array. Fig. 5 shows three such diagrams. The positions in each 3x9 grid map to the relations in Tables 1 and 2, and the more times a relation occurs in a document the bigger the white square in the diagram. For example, the fourth square in row two of the left-hand Hinton diagram represents the number of 'overlapped by' (oi) Allen Intervals in a document. Each diagram in Fig. 5 depicts the central tendency of the document, with no blocks to the extreme left or right. However, both Allen-13 (left-hand diagram) and Allen-27 (on the right) have a lot of noise, with a high percentage of relations indicating horizontal shifts as the text descends down the page. The middle diagram shows the behaviour of Allen-13T in accounting for small differences in the beginnings and ends of lines (in this case up to 8 pixels). This is a much more accurate representation of the document.

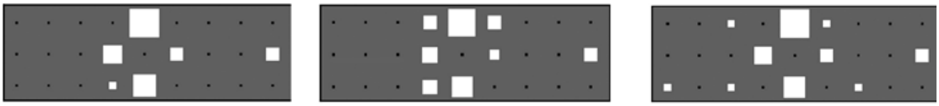


Fig. 6: Hinton diagrams of relations for Fig. 4(e), (poem in form of tree). From left to right, Allen-13, Allen-13T (T=8), Allen-27.

All three diagrams in Fig. 6 capture the expansion and contraction of the lines as they go down the page. This is evidenced by the top and bottom boxes in the middle column of each diagram. Allen-13T also captures the left alignments of the lower lines which represent the trunk of the tree. Allen-27 also shows relations which are not captured by the other two representations. For example, there is a single 'ldi' relation which occurs when interval X is near the beginning and wholly during interval Y. This may not be an important distinction but it does capture some of the unevenness of the document.

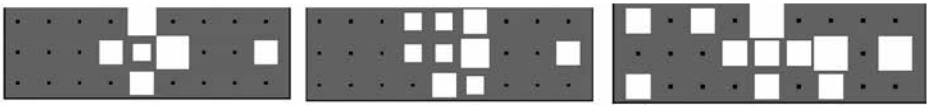


Fig. 7: Hinton diagrams of relations for Fig. 4(d), (informal letter). From left to right, Allen-13, Allen-13T (T=8), Allen-27.

In Fig. 7, each diagram provides a very different picture of the document. There is not much difference between the Allen-13T diagram and the left-hand image in Fig. 5, even though the documents are very different in layout. Allen-13 produces a distribution which is similar to that of the poem in Fig. 6. Only Allen-27, with its broader range of relations, captures the very uneven nature of the document. The way in which a collection of qualitative relationships is able to detect different characteristics in a document is similar to the use of QSR in identification of events from video data[43]. In that case, different types of events are identified partly by making use of the patterns of change to spatial relationships between areas in each video frame occupied by certain objects. Conceptually, the pattern of relationships in a particular example is like a spectrum detected by a measuring instrument in analysing the example.

The most natural data structure for storing these relations is a network graph. The nodes are objects on the page (text lines, symbols, annotations), and the edges are relations. The graph is necessarily directed, meaning that an edge with relation 'before' from nodes A to B does not also imply that B is before A. To capture the inverse 'after' relation, an edge could be created from B to A. We may also create alternative graph representations of a page. A subgraph (a subset of nodes and edges from the graph) of closely related nodes could be combined into a single node. For example, the rows of each column of text in Fig. 4(b) could be represented by a single node, resulting in a graph with 3 nodes (say, L,M,R). Working from left to right we have the relations: 'L before M', 'M before R'. A distinction should be made between text in columns, and data in tabular form. The latter could be represented in two ways, equally valid. The first representation consists of a node for each row of data, the second consists of a node for each column. This is an important aspect of QSR; it is not a prescriptive set of rules but a language and framework for representing spatial relationships, and the representation chosen will vary by application.

The R of QSR can also stand for Reasoning. The mathematical definitions of each relation enable logical assertions to be made. For example, using the fact that each Allen relation has an inverse, if A is before B, it is implied that B is after A. If L is before R and R is before M, then we infer that L is also before M without explicitly identifying the relationship between L and M. The given relations thus provide a basis from which others can be deduced without having to store every possible relationship that might be useful. A query about whether a certain relationship holds can often then be answered by checking whether it follows from the given ones.

Fig. 4(c) shows a commonly seen document structure consisting of paragraphs of text, the start of each being vertically aligned to a shorter line or two in the left margin. One way of representing this structure is to merge the left margin text into the rows of text immediately to their right. Assuming Allen-13T has been used, each line of paragraph text has the same x-start and x-end values, other than the shorter lines at the end of the paragraphs. We will also assume the margin text lines have the same x-start values. With these assumptions, we can describe the documents as sequences of relations (the first being relative to the top edge of the page). Two such sequences are shown in Table 3:

43 Krishna S. Dubba et al., Learning relational event models from video, in: Journal of Artificial Intelligence Research 53 (2015) p. 41–90.

Table 3: Sequences of relations between text lines and their immediate vertical neighbour.

TBRR: Fig. 4(c)	TBRR: Fig. 4(e)
Finished by	Contains
Equals	During x 3
Finished by	Finishes
Equals x 9	Contains
Started by	During x 3
During	Finishes
Equals	Contains
Finished by	Finishes
Equals x 14	During
Started by	etc.

These are examples of repeating patterns. In column one, the pattern 'finished by', followed by multiple 'equals' relations, is the general pattern of a line of text followed by a block of indented text. The indentation is reversed by a 'started by' relation. In the right-hand column the distinctive tree pattern of the poem consists of repetitions of 'contains', 'during' (multiple times), and 'overlapped by'. The pattern breaks down at the third tier due to the inherent messiness of a handwritten document, but resumes further down the page. This representation lends itself well to searching for documents based on their layout. Possible approaches include regular expression pattern matching, vectorised representation (as used in large language models), or sketch based image retrieval.

The example documents presented so far all read from left to right, and down the page. Fig. 8 shows a document containing text written at three different orientations, with lines coloured by orientation. The vertical lines on the left of the page read from bottom to top, while on the right of the page the horizontal lines read from left to right, and the vertical lines from top to bottom. What this example highlights is that orientation is an important attribute of each object on the page. For comparisons between intervals it makes sense to make them relative to the expected reading order of the language.

For example, the vertical line furthest to the left of the page is left of the next line along, but from the perspective of the reader it is above. The entire block of text itself would still be considered to be 'before' the block of text on the right (in the horizontal direction).

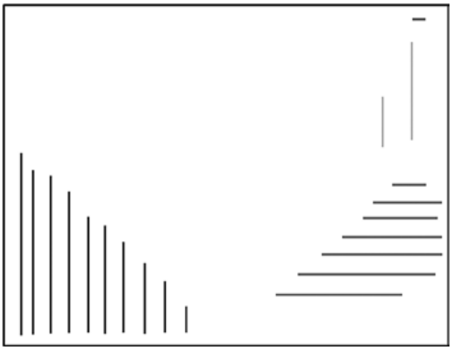


Fig. 8: Document with multiple text orientations.

Representing Change over Time

Fig. 9 shows a document with a Public Record Office (PRO) stamp. Without the contextual knowledge of when the document was written (c.1220), and when the PRO was formed (1838), a reasoning engine would not be able to decide whether the shape of the lines formed around the stamp, or whether it coincidentally formed a handy space to put a stamp. The spatial representation therefore requires a temporal dimension.

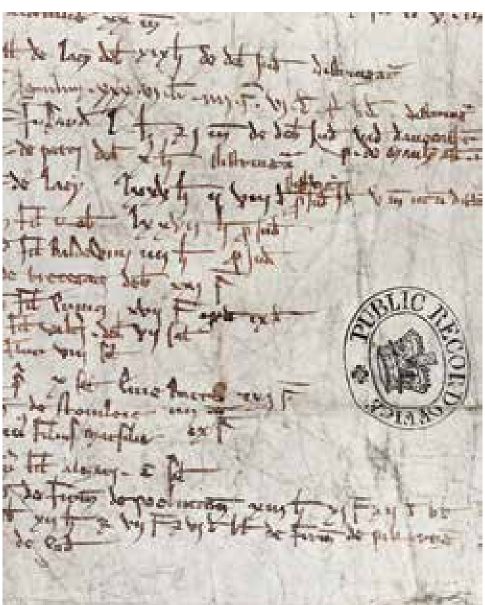


Fig. 9: Stamped document.

There are two draft versions of the speech made by Prime Minister Margaret Thatcher to welcome President Reagan on his visit to the United Kingdom in 1982, which will be used as an example of representing change over time. The layout of the first draft is that of a typed essay, several left justified paragraphs, with an indented first line. It is heavily annotated in black pen, including crossings out, encirclements, notes and replacement text between the lines and in the margins, and arrows. The sequence of events (typed speech, handwritten annotations) can be trivially represented on a timeline. Qualitative representation means we don't need to know the actual timeline. It is enough to state that the annotations came after the typed text. If there are multiple annotations, e.g. a blue pen, and a black pen, we will not necessarily know which came first but we know they are both 'after' the typed text. This can be modelled as a period of review 'after' the document was written, with each annotation occurring 'during' this period. No relation is derived between the annotations themselves. The temporal representation is enhanced when combined with the spatial arrangement of the annotations, and their relationship to the text.

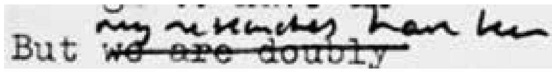

But ~~we are doubly~~

Fig. 10a: Crossed out text with replacement.

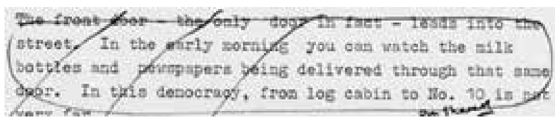
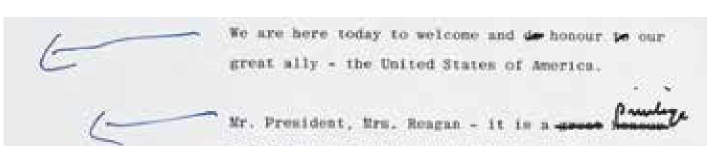

~~The front door – the only door in fact – leads into the~~ street. In the early morning you can watch the milk bottles and newspapers being delivered through that same door. In this democracy, from log cabin to No. 10 is not

Fig. 10b: Crossed out and enclosed text

Fig. 10(a) shows three words of text crossed out, with a suggested replacement written above the line. Without reference to a specific spatial calculus, the handwritten text line is above the typed line, and is also between two typed lines. The hand-drawn line shares the same space as the text it is crossing out; it overlaps in both dimensions. In Fig. 10(b) the crossing out lines only overlap partially with a small subset of the text. The line circling the section of text encloses most of the words but only overlaps some of them. The handwritten-text line can be represented by Allen Intervals, with the temporal dimension separating it from the flow of the typed text. The hand-drawn annotations are better represented by a 2-d calculus, such as RCC-8. This is another case where approximation is most likely better than exactness. While we could represent line 2, say, as being contained within (or a non-tangential proper part of, in the language of RCC-8) the ellipse, and line 1 as partially overlapping it, that overcomplicates the representation and misses the intention of the annotation. Specifying that the entire text is a tangential proper part (contained and touching the edges) of the ellipse would be more pragmatic and better capture the purpose of what has been drawn.


We are here today to welcome and ~~do~~ honour ~~to~~ our great ally – the United States of America.

Mr. President, Mrs. Reagan – it is a ~~great honour~~ and pleasure to have you both here with us.

Fig. 11: Crossed out text with replacement

Moving to the second draft, there are fewer amendments to the text but the layout has subtly changed. The left margin is wider, which has the effect of stretching the text (the bottom line appears half way down the page in the first draft). Every line is left justified, and there is a large gap between the end of one sentence and the next - presumably to aid reading in public. Fig. 11 shows the drawing of two arrows which are requests to unindent the first lines of each paragraph. The effect of this instruction is seen in Fig. 4(f). Each new sentence starts on a new line, the font size is increased, as is the space between lines, and the length of the third paragraph means it gets a page to itself. The net effect is that the first page of the speech contains only the top fifth (6 lines) of the original draft. This can be modelled by including the page itself in the spatial representation, with every line and annotation being fully contained within it.

Further investigation is needed to understand how QSR could aid reasoning about the influence of available space on the construction of a document. Elsewhere in the draft of the speech a long annotation is fitted into available whitespace and connected via an arrow to the relevant typed text, the location of the stamp in Fig. 9 is unrelated to the adjacent text, and other documents which have informed our thinking include some lines that were tightly squeezed into a space set aside for one author amongst many.

Concluding Remarks

This paper approaches heritage collections as traces of dynamic processes and processual narratives. The key component of this work is to examine documents based on their spatial structure (such as where pieces of text are placed on a page), temporal relationships (such as knowing that one annotation was added after another) and spatiotemporal interconnections. Focusing on relationships between multiple documents that form complex networks revealing knowledge about how the documents have been created, used and reused, the paper puts these relationships at the forefront.

The examples presented have demonstrated how existing spatial relation algebras can be applied to historical documents and can be approached as an aid to understanding and navigation of large heritage collections. Allen Intervals, in various forms, have been used to evaluate the alignment of adjacent lines of text in the horizontal and vertical directions, and to track changes to a document over time. The Region Connection Calculus has been used to relate two-dimensional areas of text and annotation to each other. While the logical statements describing each relationship generally work well in this context, the qualitative descriptions of the relations are not always appropriate for describing documents. The term 'Non-tangential proper part' from RCC-8, meaning 'region X is wholly inside region Y', is unlikely to be understood by historians. The terminology of Allen Intervals is rooted in temporal language - saying line X occurs 'during' line Y makes little sense. This suggests that a spatial calculus using language which is meaningful to humanists is needed.

Each variety of Allen Intervals that we tested had advantages and disadvantages. The TBRR variety is certainly superior to Allen-13 when working with historical manuscripts as it smooths out the unevenness in handwritten documents. Setting the T parameter to 0 is also equivalent to Allen-13. The extended version of Allen provides more nuance in that it provides an indication of magnitude when intervals partially overlap, and direction when they fully overlap. A thick bordered variety of Allen-27 would improve its application to historical manuscripts. The downside of Allen-27 is that it consists of more than double the possible relations of Allen-13, which will have a computational cost when reasoning. The number of relationships can be reduced. In some applications, the inverse relations may not be applicable, particularly the 'most-most' relations such as 'mom' (most of X overlaps most of Y). In a paragraph of text the difference between 'mostly started' and 'mostly started by' could be down to whether line X is slightly shorter than Y, or vice versa. It may not be a meaningful difference. Equally, there may be applications where we wish to increase the number of relations. The 'during' relations ('ld','cd','rd') and their inverses offer no reference to relative length beyond that one interval is shorter than the other. One possibility is to incorporate the delta calculus in the QSR which would compare the lengths of two intervals. For example, a relation such as 'X centrally during and more than half length of Y' could be defined. In two dimensions, the eight relations of RCC may seem limited in applications to document understanding, but the Rectangle Algebra could range from 26 relations (Allen-13 x Allen-2) to 729 (Allen-27 x Allen-27), which probably makes too many distinctions for a useful qualitative relation. RCC-8 can be extended to incorporate degree of connection. For instance, 'partially overlapping' could be split into 'just' and 'well' overlapping; 'non-tangential proper-part' (NTPP) becomes 'just inside' and 'well inside'. RCC also lacks any notion of direction. Combining it with a direction calculus could give rise to relations such as 'overlapping from left' or 'overlapping from above'. An RCC with added direction and degree of overlap would consist of 23 relations: disconnected x {left, right, above, below}, externally connected x {4 directions}, partially overlapping x {4 directions} x {well, just}, NTPP (and its inverse) x {well, just}, TPP (and its inverse), and equal.

The examples presented in this paper are QSRepresentation. The derived relations only state that, say, two objects on the page happen to overlap. In order to attain QSReasoning human knowledge is needed. Take for example, a hand-drawn line. If it goes through a word it is an instruction to remove that word, below the word it may be interpreted as emphasis. If it connects an annotation to a line of typed text, it could be an instruction to insert or append the annotation text. The margin text in Fig. 4 is interpreted as the beginning of a section. The authors of that document also drew vertical lines down the left side of each paragraph to explicitly group those lines together. Other documents use symbols for the same purpose, while spacing between lines or sub-headings may also have the same meaning.

Figs. 12(a)–(c)[44] show three ways of drawing brackets. Visually they differ considerably but semantically they are the same. Indentation and spacing can be used to the same effect. In 12(a) the author of the document has used this bracketing only in the cases where the left margin text does not align with the top line of the section; they show an awareness of how spatial relationships are interpreted.

Fig. 12a–c: Brackets to group lines of text.

This paper has introduced Qualitative Spatial Representation (QSR) as a method to explore the spatial and temporal dimensions of digital document collections. It is a complementary technology to machine learning for technology-based document layout analysis and understanding, but it can be applied separately to automated techniques. We have identified several applications for the technology, including: representing spatial relationships between objects on a page, and how they evolve over time; enhanced search capabilities such as searching text 'near to' an annotation, or finding documents similar to a certain layout; human understandable explanations of the results of machine learning based layout analysis; and a language for communicating human knowledge of document layout to algorithms. In order to progress these applications, experts in history and diplomatics need to work with computer scientists to develop qualitative representations which comprise relationships that are applicable to the study of historical documents, and to curate datasets which will encourage further exploration of this technology.

44 The National Archives, UK: a) E 368/3 b) E 368/5 c) E 372/179