Christoph Sander

Traces of Attention

A Computational Approach to the Study of Readers' Marks in Early-Modern Printed Books

Introduction

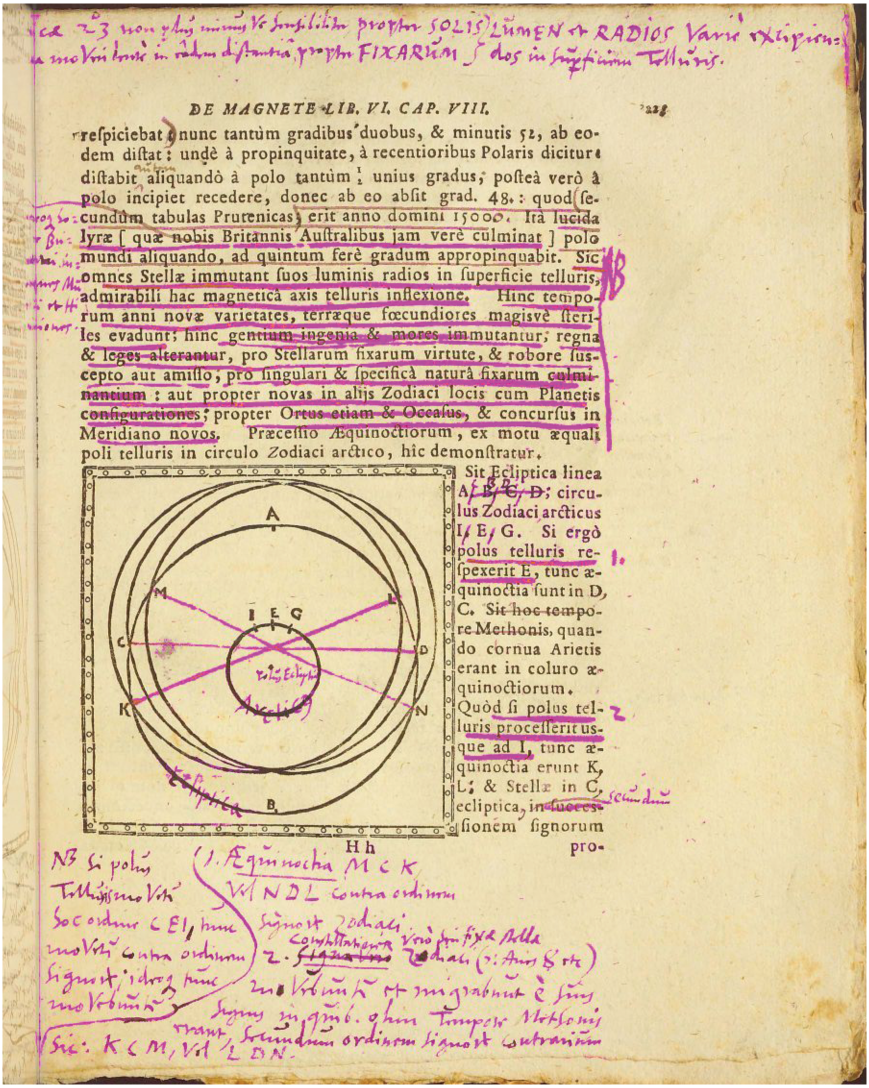
Anyone who has ever purchased or borrowed used books knows that the print can be enriched or soiled by underlinings, crossings out, marginal notes, corrections, and sketches made by previous readers. Reading a book may leave traces, and these traces can also be found in large numbers in very old copies of books (Figure 1). The material carriers of texts, namely the individual printed copies of the book, were, and continue to be, immediate points of contact with the ideas of an author and may testify to readers' written reactions to these ideas.

Surviving readers' annotations are material traces of past readings with significance for the history of ideas.[1] They can be used to research the past reception of scholarly publications, instead of investigating how one publication reacts to another one. This approach consequently yields into an original and promising synthesis of material history and the history of ideas.[2] The underlining of text, for example, is

[1] See, for example, Lisa Jardine and Anthony Grafton, "Studied for Action": How Gabriel Harvey Read His Livy, *Past & Present*, no. 129 (1990): 30–78; Danielle Jacquart and Charles S. F. Burnett (eds.), *Scientia in margine: Études sur les marginalia dans les manuscrits scientifiques du moyen âge à la renaissance*, Geneva: Droz, 2005; Heather Joanna Jackson, *Marginalia: Readers Writing in Books*, New Haven (CT): Yale University Press, 2001; William H. Sherman, *Used Books: Marking Readers in Renaissance England*, Philadelphia: University of Pennsylvania Press, 2010; Stephen Orgel, *The Reader in the Book: A Study of Spaces and Traces*, Oxford: Oxford University Press, 2015; Katherine O. Acheson (ed.), *Early Modern English Marginalia*, New York: Routledge, 2019; Anthony Grafton, *Inky Fingers: The Making of Books in Early Modern Europe*, Cambridge (MA): Harvard University Press, 2020; Ku-ming (Kevin) Chang, Anthony Grafton, and Glenn Warren Most (eds.), *Impagination – Layout and Materiality of Writing and Publication: Interdisciplinary Approaches from East and West*, Berlin: De Gruyter, 2021, <https://doi.org/10.1515/9783110698756>, accessed February 22, 2025; Sylvia Brockstieger and Rebecca Hirt (eds.), *Handschrift im Druck (ca. 1500–1800): Annotieren, Korrigieren, Weiterschreiben*, Berlin: De Gruyter, 2023, <https://doi.org/10.1515/9783111191560>, accessed February 22, 2025.

[2] This project does not subscribe to an old-fashioned division between the history of ideas on the one hand and bibliography and book history on the other, or, more specifically, between the study of content and its medium, such as Walter Wilson Greg, Bibliography: An Apologia, *The Library*, 4th series, 13, no. 2 (1932): 121–122, <https://doi.org/10.1093/library/s4-XIII.2.113>, accessed February 22, 2025.

Open Access. © 2026 with the author(s), published by De Gruyter. cc BY This work is licensed under the Creative Commons Attribution 4.0 International License.
https://doi.org/10.1515/9783111560229-012



ca. 2°3 non plus minus ve sensibiliter propter SOLIS LUMEN et RADIOS varie excipien[illegible] [illegible] in eadem distantia, propter FIXARUM [illegible] dos in superficiem Telluris.

DE MAGNETE LIB. VI. CAP. VIII. 228

reſpiciebat) nunc tantùm gradibus duobus, & minutis 52, ab eodem diſtat: undè à propinquitate, à recentioribus Polaris dicitur. diſtabit aliquandò à polo tantùm ½ unius gradus; poſteà verò à polo incipiet recedere, donec ab eo abſit grad. 48.: quod (ſecundùm tabulas Prutenicas) erit anno domini 15000. Ità lucida lyræ [quæ nobis Britannis Auſtralibus jam verè culminat] polo mundi aliquando, ad quintum ferè gradum appropinquabit. Sic omnes Stellæ immutant ſuos luminis radios in ſuperficie telluris, admirabili hac magneticâ axis telluris inflexione. Hinc temporum anni novæ varietates, terræque fœcundiores magisvè ſteriles evadunt; hinc gentium ingenia & mores immutantur; regna & leges alterantur, pro Stellarum fixarum virtute, & robore ſuſcepto aut amiſſo; pro ſingulari & ſpecificâ naturâ fixarum culminantium: aut propter novas in alijs Zodiaci locis cum Planetis configurationes; propter Ortus etiam & Occaſus, & concurſus in Meridiano novos. Præceſſio Æquinoctiorum, ex motu æquali poli telluris in circulo Zodiaci arctico, hîc demonſtratur.

NB

Sit Ecliptica linea A, ~~B, C, D~~; circulus Zodiaci arcticus I, E, G. Si ergò polus telluris reſpexerit E, tunc æquinoctia ſunt in D, C. ~~Sit hoc tempore Methonis~~, quando cornua Arietis erant in coluro æquinoctiorum. Quòd ſi polus telluris proceſſerit usque ad I, tunc æquinoctia erunt K, L; & Stellæ in C, ecliptica, ~~in ſucceſ~~ſecundum ſionem ſignorum pro-

1.

2.

H h

NB Si polus Telluris moveri[?] sec. ordinem C E I, tunc moveri contra ordinem signorum, idcirco tunc movebuntur

(1. Æquinoctia M C K vel N D L contra ordinem signorum zodiaci.

2. ~~signa~~ constellationes vero seu fixae stellae Zodiaci (v.: Aries & etc) movebuntur et migrabunt e suis signis in quibus olim tempore Methonis erant, secundum ordinem signorum contrarium

Sic: K C M, vel L D N.

Figure 1: A page from William Gilbert, *De Magnete* (1628), annotated by Andreas Granius. Everything considered a readers’ mark is digitally highlighted in pink (Source: HAB Wolfenbüttel: M: Nc 4° 46 <https://diglib.hab.de/drucke/nc-4f-46/start.htm>, accessed February 22, 2025).

presumed to indicate that a reader was particularly interested in this part of the text. Dog-earing helps to quickly find an important page again, while the deliberate tearing out of a page may indicate that this page contains something that should not be received. The attitude towards and interest in the ideas formulated in the text is expressed by the material handling of the book pages on which the text was printed.

For historians, handwritten annotations provide an undisguised, almost intimate, but often not easy to interpret insight into the reception of specific works.[3] Significant reception tendencies can be determined if similar comments or markings are found in several copies of a specific work.[4] Or they can be discovered if similar topics are marked by readers across multiple editions and copies of different works.[5] These research perspectives are not novel; however, they can be advanced and executed in a novel manner through the use of digital resources and methodologies.[6]

[3] See Robert S. Westmann, The Reception of Galileo's "Dialogue": A Partial World Census of Extant Copies, in: *Novità celesti e crisi del sapere: atti del convegno internazionale di studi galileiani*, Paolo Galluzzi (ed.), 329–371, Florence: Giunti Barbèra, 1984; Owen Gingerich, *An Annotated Census of Copernicus' "De Revolutionibus" (Nuremberg, 1543 and Basel, 1566)*, Leiden: Brill, 2002; Thomas F. Mayer, An Interim Report on a Census of Galileo's Sunspot Letters, *History of Science* 50, no. 2 (2012): 155–196, <https://doi.org/10.1177/007327531205000202>, accessed February 22, 2025; Dániel Margócsy, Mark Somos, and Stephen N. Joffe, *The Fabrica of Andreas Vesalius: A Worldwide Descriptive Census, Ownership, and Annotations of the 1543 and 1555 Editions*, Leiden: Brill, 2018; Mordechai Feingold and Andrej Svorenčík, A Preliminary Census of Copies of the First Edition of Newton's "Principia" (1687), *Annals of Science* 77, no. 3 (2020): 253–348, <https://doi.org/10.1080/00033790.2020.1808700>, accessed February 22, 2025.

[4] Sherman, *Used Books: Marking Readers in Renaissance England*, 25–52.

[5] Cf., for example, Gingerich, *An Annotated Census of Copernicus' "De Revolutionibus" (Nuremberg, 1543 and Basel, 1566)*, xvi: "In fact, a quite unexpected result has been the discovery of patterns of annotations that are found in multiple copies of the book." Similarly in Margócsy, Somos, and Joffe, *The Fabrica of Andreas Vesalius*, 56: "The availability of such large corpora of annotated books has allowed scholars to identify the reading patterns of even those annotators who left only one or two marginalia here and there." Cf. also Margócsy, Somos, and Joffe, *The Fabrica of Andreas Vesalius*, 61: "How can one use such a database that contains information on hundreds of annotators? A few years ago, newspapers were abuzz with the concern that e-book providers may carefully track what percentage of readers finishes the book they purchase, and that it is possible to tell at what page these readers stop reading. One can perform similar calculations on consumers of the *Fabrica*, using their annotations as an imperfect proxy for reading activity. By visually tracking how frequently each page of the *Fabrica* was annotated in our corpus, we are able to see where readers tended to turn their attention, and we can combine this quantitative analysis with a careful consideration of the variety of marginalia that decorate these pages." However, none of these projects ever published actual databases.

[6] Readers' marks have also already featured in digital humanities projects, for example as library subcollections of annotated copies (e.g. UCLA's Clark Library, <https://calisphere.org/collec-

For instance, federated library catalogs, comprehensive digitization of books, authority files, and standardized vocabularies facilitate the rapid identification, examination, and description of a multitude of copies of an edition.[7] The recognition of full texts through OCR/HTR provides machine readable access to their content. This content can itself become subject to explorative natural language processing techniques, such as Named Entity Recognition that looks out for persons, places, and other recognizable entities in the text.[8] Topic Modeling and the use of (Large) Language Models give reliable insights into the actual content beyond the literal by representing the semantic dimension as embedding. These powerful tools allow to track down if two texts address similar subjects or have a similar tonality. On the purely visual level, the identification of modifications of the printed pages by later hands or circumstances can increasingly reliably be recognized by Computer Vision.[9] Multimodal Visual Language Models are able to relate visual content, e.g. a readers' mark,

tions/26771/>, accessed April 8, 2024) or cross-sections of marginalia relating to certain topics (e.g. Early Modern Women's Marginalia, <https://earlymodernwomensmarginalia.cems.anu.edu.au>, accessed April 8, 2024). Most importantly, the large-scale project The Archaeology of Reading in Early Modern Europe (<https://archaeologyofreading.org>, accessed April 8, 2024) has created a sophisticated website to explore the marks made by two famous sixteenth-century readers, John Dee and Gabriel Harvey. Digital databases have also repeatedly recorded provenance information as structured data, especially with a focus on artworks (e.g., Boehler re:search, <https://boehler.zikg.eu>, accessed April 8, 2024, and Provenia, <https://www.proveana.de/de/start>, accessed April 8, 2024) but also on book history (e.g. Book Owners Online, <https://bookowners.online>, accessed April 8, 2024, Owners of Incunabula, <https://data.cerl.org/owners>, accessed April 8, 2024, and Archivio dei Possessori, <https://archiviopossessori.it>, accessed April 8, 2024. These impressive projects and their platforms are, however, neither intended nor equipped to analyze readers' marks with a quantitative methodology, and hence do not qualify to transparently discover large-scale patterns of book use and readers' engagements with topics beyond genres and very small social groups.

[7] For authority files and vocabularies, see esp. <https://data.cerl.org/mei>, <https://www.getty.edu/publications/categories-description-works-art/categories/object-architecture-group/8/>, <https://15cbooktrade.ox.ac.uk/reading-practices/>, <https://provenienz.gbv.de/T-PRO_Thesaurus_der_Provenienzbegriffe>, and <https://viaf.org>, accessed April 8, 2024.

[8] Megan R. Brett, Topic Modeling: A Basic Introduction, *Journal of Digital Humanities* 2, no. 1 (2012). <https://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/>, accessed February 22, 2025; Yida Mu, Chun Dong, Kalina Bontcheva, and Xingyi Song, Large Language Models Offer an Alternative to the Traditional Approach of Topic Modelling, *arXiv:2403.16248. Preprint, arXiv*, March 26, 2024. <https://doi.org/10.48550/arXiv.2403.16248>, accessed February 22, 2025.

[9] Jochen Büttner, Julius Martinetz, Hassan El-Hajj, and Matteo Valleriani, CorDeep and the Sacrobosco Dataset: Detection of Visual Elements in Historical Documents, *Journal of Imaging* 8, no. 10 (2022): 285, <https://doi.org/10.3390/jimaging8100285>, accessed February 22, 2025; Hassan El-Hajj, Oliver Eberle, Anika Merklein, Anna Siebold, Noga Shlomi, Jochen Büttner, Julius Martinetz, Klaus-Robert Müller, Grégoire Montavon, and Matteo Valleriani, Explainability and Transparency in the Realm

to textual content by way of embeddings.[10] Unsupervised machine learning algorithms as well as foundation models present out-of-the box instruments to explore this immense amount of information to arrive at a deeper understanding of emerging patterns and general structures.[11] Last but not least, the modeling of all of this as knowledge graphs provides a complex and powerful infrastructure to retrieve answers to actual research questions addressed at the digital data.[12]

All of this, and the many ramifications not outlined here, demand for a methodological, epistemological and ontological framework.[13] The present chapter

of Digital Humanities: Toward a Historian XAI, *International Journal of Digital Humanities* 5, no. 2 (2023): 299–331, <https://doi.org/10.1007/s42803-023-00070-1>, accessed February 22, 2025.

[10] Research is moving very quickly, several models for the task are available, e.g.: LLaVa, *Hugging Face*, May 7, 2024. <https://huggingface.co/docs/transformers/model_doc/llava>, accessed February 22, 2025. As a starting point, cf. also Leo Impett and Fabian Offert, There Is a Digital Art History, *Visual Resources* 38, no. 2: 186–209, <https://doi.org/10.1080/01973762.2024.2362466>, accessed February 22, 2025.

[11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay, Scikit-Learn: Machine Learning in Python, *Journal of Machine Learning Research* 12, no. 85 (2011): 2825–2830; Jakob Jünger and Chantal Gärtner, *Computational Methods für die Sozial- und Geisteswissenschaften*, Wiesbaden: Springer, 2023; Lorella Viola, *The Humanities in the Digital: Beyond Critical Digital Humanities*, Heidelberg: Springer, 2023. For some case studies in the domain of early modern intellectual history, see Matteo Valleriani, Florian Kräutli, Maryam Zamani, Alejandro Tejedor, Christoph Sander, Malte Vogl, Sabine Bertram, Gesa Funke, and Holger Kantz, The Emergence of Epistemic Communities in the Sphaera Corpus, *Journal of Historical Network Research* 3, no. 1 (2019): 50–91, <https://doi.org/10.25517/jhnr.v3i1.63>, accessed February 22, 2025; Matteo Valleriani, Malte Vogl, Hassan El-Hajj, and Kim Pham, The Network of Early Modern Printers and Its Impact on the Evolution of Scientific Knowledge: Automatic Detection of Awareness Relationships, *Histories* 2, no. 4 (2022): 466–503, <https://doi.org/10.3390/histories2040033>, accessed February 22, 2025; James Misson and Devani Mandira Singh, Computing Book Parts with EEBO-TCP, *Book History* 25, no. 2 (2022): 503–529, <https://doi.org/10.1353/bh.2022.0018>, accessed February 22, 2025; Andrea Sangiacomo, Raluca Tanasescu, Hugo Hogenbirk, and Silvia Donker, Recreating the Network of Early Modern Natural Philosophy: A Mono- and Multilingual Text Data Vectorization Method, *Journal of Historical Network Research* 7, no. 1 (2022): 33–85, <https://doi.org/10.25517/jhnr.v7i1.129>, accessed February 22, 2025.

[12] Mayank Kejriwal, Craig A. Knoblock, and Pedro Szekely (eds.), *Knowledge Graphs: Fundamentals, Techniques, and Applications*, Cambridge (MA): The MIT Press, 2021; Arianna Ciula, Øyvind Eide, Cristina Marras, and Patrick Sahle, *Modelling between Digital and Humanities: Thinking in Practice*, Cambridge (UK): Open Book Publishers, 2023, <https://books.openbookpublishers.com/10.11647/obp.0369.pdf>, accessed February 22, 2025.

[13] Mark Bevir, *The Logic of the History of Ideas*, Cambridge (UK): Cambridge University Press, 1999, <https://doi.org/10.1017/CBO9780511490446>, accessed February 22, 2025; Andrew Piper, *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*, Cambridge (UK): Cambridge University

attempts to outline this framework by way of a concrete example from the Magnetic Margins Database.[14] This research does thus not aim at presenting a specific research case in its own right but uses a specific case to address more general preliminaries and implications. In doing so, its contribution lies in a critical and reflexive take on the advancement of computational methods in the historical humanities, narrowed down to the intricate subject of readers' marks. Concrete technical details of the data model and computational algorithms – which are evolving rapidly – are beyond the scope of this chapter and will be provided in a dedicated repository and future publications.[15]

First, I will present the methodological, epistemological, and ontological presuppositions that form the framework for the subsequent analysis. Then, I will provide a brief overview of the dataset and its historical significance. Finally, I will examine the potential of computational analysis through examples from the sample dataset.

1 Presuppositions

The presuppositions outlined here, which form the conceptual framework of the proposed approach, are not intended as exhaustive ontological claims. Rather, they serve as pragmatic assumptions underpinning the methodology and epistemology at hand. Some of them draw on established formalized ontologies from the Semantic Web domain.

(1) Concrete, physical annotations and marks – referred to hereafter as readers' marks (RMs) – are understood as psychological indicators or markers of attention. These RM are interpreted as traces that provide insight into historical interactions between a reader and a text. Interest in a particular passage or

Press, 2020, <https://doi.org/10.1017/9781108922036>, accessed February 22, 2025; David John Hand, *Dark Data: Why What You Don't Know Matters*, Princeton: Princeton University Press, 2020.

[14] Christoph Sander, *Magnetic Margins: A Census and Annotations Database*, October 25, 2023, <https://doi.org/10.48431/res/qk19-bj96/magmar>, accessed February 22, 2025; Christoph Sander, Hassan El-Hajj, and Alessandro Adamou, Magnetic Margins: A Census and Reader Annotations Database, in: *Digital Humanities 2023: Collaboration as Opportunity (DH2023)*, Graz: Zenodo, 2023, <https://doi.org/10.5281/zenodo.8107608>, accessed February 22, 2025.

[15] See <https://doi.org/10.5281/zenodo.14851170> and <https://github.com/ch-sander/raramagnetica/tree/main/analysis/magnetic_margins>, accessed February 22, 2025. A comprehensive account of the census as well as the census itself will be published in Christoph Sander, Magnetic Margins: Insights from the Digital Descriptive Census of William Gilbert's "De magnete", *Annals of Science*, forthcoming.

content may prompt the creation of a RMs. In this sense, they are viewed as traces of reading, sense-making, and evaluative engagement.[16]

(2) Text (as printed in books, in this context) is regarded as the meaningful expression of content – hereafter referred to as semantic expressions (SEs) – which can be characterized by a set of semantic features (SFs). These features are recurrent across different texts, whether as explicit topics or as latent embeddings in high-dimensional vector spaces produced by modern language models. Thus, a single text (or textual unit, such as a sentence) may refer to multiple topics, and conversely, a single topic may be addressed by multiple texts. The scope and nature of these features will be elaborated below.

(3) Any RM that resembles another RM (e.g., by sharing a particular property), or any SE that shares a relevant SF with another SE, is interpreted as expressing semantic similarity. This similarity gives rise to a semantic topology among RMs and SEs. Two RMs or SEs that share any SF are considered semantically related – or similar – with respect to that feature.

(4) The identification and description of any RM, SE, or SF constitutes an abstraction from a unique material or conceptual context. Connections among these elements, based on abstracted commonalities, are necessarily tentative, subject to uncertainty, vagueness, and contestability. Often, little is known about the provenance of an RM, and whether two SEs genuinely share an SF may depend on interpretive judgment or probabilistic modeling. Despite these inherent challenges of underdetermination, the overall quantitative analysis proceeds on the assumption that such biases in the ground truth or algorithm are statistically negligible or can be addressed through critical discussion of results and the presentation of comparative analytical scenarios.

While philosophical in nature, these aforementioned presuppositions have partly a technical correspondence in the domain of Semantic Web technologies.[17] Digital Semantic Web ontologies (such as CIDOC-CRM, FRBR(oo), FaBiO, DoCO, and CiTO) structure semantic data by categorizing entities and properties into relationships.[18]

[16] This justifies, therefore, the exclusion of annotations related to ownership marks, librarians' cataloging notes, repagination, and similar non-reader-oriented markings.

[17] Richard Gartner, *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*, Cham: Springer, 2016; Elena Spadini, Francesca Tomasi, and Georg Vogeler (eds.), *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, vol. 15, Norderstedt: BoD, 2021, <https://kups.ub.uni-koeln.de/54577>, accessed February 22, 2025.

[18] See Silvio Peroni and David Shotton, The SPAR Ontologies, in: *The Semantic Web – ISWC 2018*, Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, and Elena Simperl (eds.), 119–136, Cham: Springer, 2018,

The following discussion – subscribing to FRBR-ontologies –, particularly leverages the distinction between concrete physical objects that exist(ed) at some period of time at some place on the one hand, and immaterial conceptual objects that carry meaning on the other hand.[19] The latter classes of immaterial entities include intellectual works, e.g. some treatise as a generic entity, and their expressions (e.g. its original linguistic form or a translation of that work) and manifestations (e.g. a published edition of that work). Instances (or, parts of them) of these classes are SE and have SF.[20] In contrast, any concrete material representation of such an immaterial conceptual object is seen as an item that has material properties and a spatiotemporal existence. By itself, no item is considered a SE, but it can be a (or, more precisely, its manifestation's) physical medium or carrier of some SE. As we shall see, RMs are strictly speaking items, i.e. physical unique items, but their mere existence can also be interpreted on a semantic level as a SE. While any full-fledged Semantic Web ontology is much more complex, this bifold distinction material-immaterial is the most crucial one for the approach pursued here.[21]

<https://doi.org/10.1007/978-3-030-00668-6_8>, accessed February 22, 2025; and <https://www.spar-ontologies.net>, accessed February 22, 2025.

[19] For general methodological reflections and a rationale of bibliography, see Bevir, *The Logic of the History of Ideas*; George Thomas Tanselle, *A Rationale of Textual Criticism*, Philadelphia: University of Pennsylvania Press, 1992; George Thomas Tanselle, *Descriptive Bibliography*, Charlottesville (VA): The Bibliographical Society of the University of Virginia, 2020, 122–160.

[20] <http://purl.org/spar/fabio> and <https://cidoc-crm.org/frbroo> (renamed to LRMoo), accessed April 8, 2024; Silvio Peroni and David Shotton, FaBiO and CiTO: Ontologies for Describing Bibliographic Resources and Citations, *Journal of Web Semantics* 17 (2012): 33–43, <https://doi.org/10.1016/j.websem.2012.08.001>, accessed February 22, 2025; Karen Coyle, *FRBR, Before and After: A Look at Our Bibliographic Models*, Chicago: ALA Editions, 2016. See also Spadini, Tomasi, and Vogeler (eds.), *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*; Katrina Fenlon, Modeling Digital Humanities Collections as Research Objects, in: *Proceedings of the 18th Joint Conference on Digital Libraries (JCDL 2019)*, 138–147, Champaign (IL): IEEE Press, 2020, <https://doi.org/10.1109/JCDL.2019.00029>, accessed February 22, 2025; Koraljka Golub and Ying-Hsang Liu (eds.), *Information and Knowledge Organisation in Digital Humanities: Global Perspectives*, London: Routledge, 2021, <https://doi.org/10.4324/9781003131816>, accessed February 22, 2025.

[21] The hierarchical structure of these classes also defines the propagation of their properties. A SF of a work also applies to its expressions and manifestations, while specific properties of an expression (e.g. the language of a translation) does not apply to the work etc. Any item that is an exemplar of a manifestation or represents an expression also propagates their properties: Any copy of Plato's *Parmenides* in some German translation T in an edition E has printed text that deals with the same philosophical puzzles of that dialogue. Moreover, in a concrete encoding for the Semantic Web, any entity or property is ideally linked to a resource that provides information and context. Entities of classes that include persons, institutions (like libraries), places, editions, and specific copies are linked to authority files such as Wikidata, GND, and USTC to provide standardized references that ensure con-

What then are the most relevant properties of these classes? The scope of semantic data on works, expressions, and manifestations includes the substructure of works such as its chapters, its text and images (expression), and bibliographic information on editions (i.e. manifestations) such as place, year, publisher – but also the work's genre and creator, a (primary) language, topics, latent semantic embeddings, intertextual references, as well as tonality of the text.[22] Information about analog items like individual copies of editions centers on provenance information, material conditions (binding etc.), and conditions of the printed pages such as marks or damages.[23]

2 Magnetic Margins

Before exploring how these ontological presuppositions and commitments drive and affect modeling a concrete dataset, our example corpus deserves a historiographical introduction. The project Magnetic Margins investigates how and by whom the most important early-modern book publications on magnetism were read and annotated.[24] Magnetism emerged as one of the most important topics in natural philosophy and experimental research in the early modern period. This interest led to the publication of several extensive volumes in the seventeenth century, most famously William Gilbert's *De Magnete* (1600), Niccolò Cabeo's *Philosophia Magnetica* (1629), and Athanasius Kircher's *Magnes* (1641).[25] This project aims at investigat-

sistency and accuracy in data linkage across different systems and datasets. Moreover, controlled semantic vocabularies, managed under frameworks like SKOS, are used to define descriptive concepts.

[22] Typically, some of this information is metadata retrieved from external resources or defined manually, other can be retrieved through computation, including lemma frequency analysis, Named Entity Recognition (NER), sentiment analysis, analysis of chapter headings, Latent Dirichlet Allocation (LDA) for topic modeling, Vision Language Modeling (VLM). Large Language Models representtextual meaning through embeddings, i.e., numerical vector representations in which the dimensions are latent (not human-interpretable) and encode compressed semantic information. See note 9.

[23] Retrieval of this material information utilizes metadata from library cataloging resources but potentially also employs Computer Vision technologies used for scans of individual objects.

[24] A few preliminaries are developed in Christoph Sander, *Magnes: Der Magnetstein und der Magnetismus in den Wissenschaften der Frühen Neuzeit*, Leiden: Brill, 2020, 833–838, <https://doi.org/10.1163/9789004419414>, accessed February 22, 2025; Christoph Sander, Magnetism in an Aristotelian World (1550–1700), in: *Alte und neue Philosophie: Aristotelismus und protestantische Gelehrsamkeit in Helmstedt und Europa (1600–1700)*, Bernd Roling, Sinem Kılıç, Benjamin Wallura, and Hartmut Beyer (eds.), 69–105, Wiesbaden: Harrassowitz, 2023, 90–98.

[25] See esp. Sander, *Magnes*, 792–874.

ing these publications through the eyes of their contemporary readers, traceable through their handwritten RMs in the editions' individual copies.

The dataset used in the following focuses on the work by William Gilbert, extant in three editions (and one title issue) and 481 known copies (although some are spurious or privately owned). These copies contain ca. 10,000 reader annotations.[26] Gilbert's work is known for its aspiration to establish a new natural philosophy centered on magnetism, for employing a rigorous experimental approach, harsh criticisms of previous bookish views on the phenomena, and a defense and physical explanation of Copernican cosmology. *De Magnete* was highly influential and impacted on better known scholars of the era, such as Johannes Kepler, Galileo Galilei, Francis Bacon, and René Descartes.

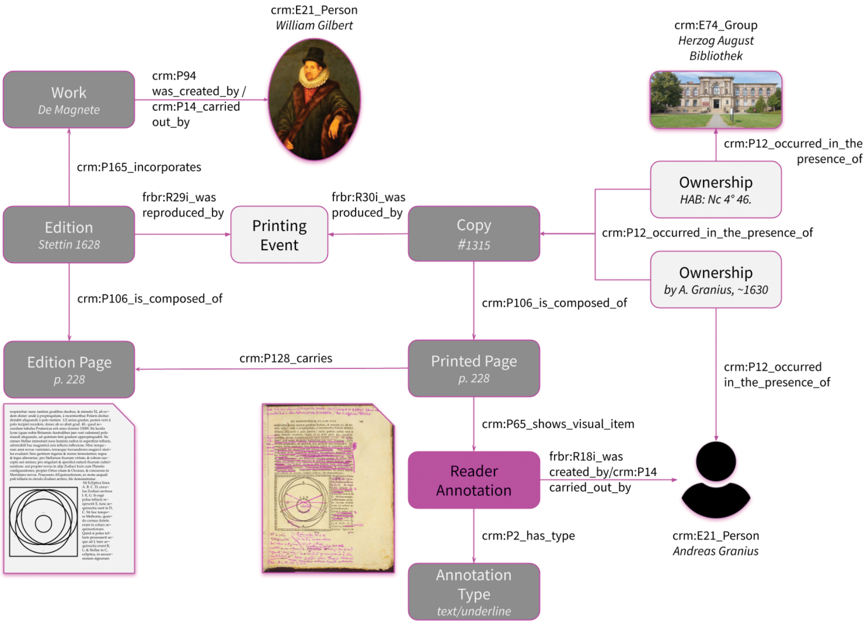


Figure 2: Simplified ontology diagram of the Magnetic Margins knowledge graph (see <https://magnetic-margins.com>, accessed February 22, 2025).

[26] On how to count these, see below.

The Magnetic Margins ontology (Figure 2) follows the CIDOC-CRM, FRBRoo, and FaBiO models.[27] The conceptual objects within this database include works (and expressions), which represent structured scientific knowledge foundational to the thematic and intellectual content. The physical objects encompass book copies (items), which are individual instances of editions (manifestations) that have been printed, bound, and distributed. Each copy carries its own history of ownership, study, and use, evidenced by physical marks and wear. Annotations or RMs, as physical objects, are indirect indications of reader engagement and include notes, underlinings, marginalia, and other markings made by readers as they interacted with the texts.

Works are represented as textual expressions by way of their manifestations and have been structured into chapters. Editions are dissected into edition pages (as prepared by the typesetter) that contain the text of the work's expression in one manifestation.[28] Each printed page (a physical item), that is bound in one copy of an edition, is a carrier of its respective edition page. The texts of chapters and edition pages are described through SFs, by (my own) manual annotation employing a controlled vocabulary, retrieved from LDA topic modeling, and most frequent lemmata, both in English and Latin.[29] Printed images are described by a hierarchical controlled vocabulary.[30] Metadata on each edition includes typical biographical information.[31]

Key agents in the creation and dissemination of these works include authors, who created the works, and publishers or printers, who were responsible for the financing, creation, and distribution of editions and copies. Owners, which can be individuals or organizations such as libraries or institutions, held and hold possession of the book copies. The spatial and geographical data covered in the database include places of publication, which are specific cities where publishers operated and where the books were printed and/or initially distributed. Locations of copies

[27] Some classes were added to better capture the phenomena under scrutiny. In the diagram (Figure 2), prefixes denote namespaces (i.e. ontologies and vocabularies). Classes are linked to each other by properties, directed by the arrows.

[28] In the case of *De Magnete*, variants between the editions/manifestations are minimal, i.e. all manifest the same work in almost the same way.

[29] Topic modeling refers to a computational technique to extract major topics from text. See note 9.

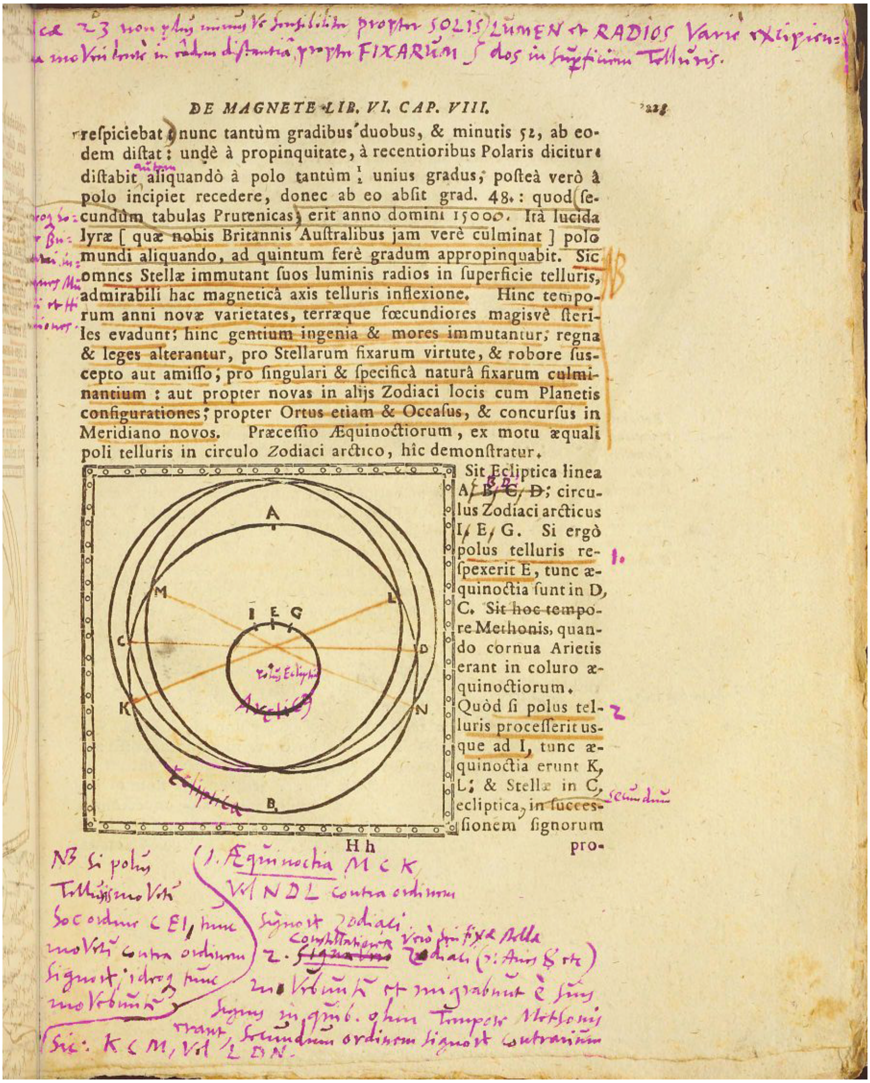
[30] See Christoph Sander, *Early Modern Magnetism Image Database (1500–1650)*, 2022, <https://doi.org/10.48431/res/qk19-bj96/vikus/vismag>, accessed February 22, 2025.

[31] For copies individual library catalogs were used for metadata, while most metadata has been added manually when reviewing the copies directly. By way of the study of all copies, different issues and even different states of printed sheets have been identified as well. See Tanselle, *Descriptive Bibliography*, 122–160.

reflect the current or historical locations of individual book copies, showing the geographical spread and the movement of items over time. The temporal data in the database is also crucial, with composition dates suggesting when works were conceptually completed, providing insights into the intellectual timeline of scientific ideas. Publication dates indicate when editions were physically printed, providing a timeline for the material production of scientific knowledge. Ownership dates track when specific copies were acquired or cataloged by owners, offering a chronological footprint of the book's life and usage.

In the database, annotations, or RM, are defined broadly to include any intentional and detectable engagement with a printed text. Each RM is exclusively categorized under one formal type, which is determined by its visible characteristics rather than the annotator's intentions. This method ensures that data is structured consistently across various instances. The database identifies several types of RM, such as text or numbers added by a reader in their own handwriting (Figure 3), markings that underline or highlight text (Figure 4), the use of symbols like "NB" for *nota bene* or manicules to denote importance, and instances where text has been obscured or blackened out by the reader. Other types include drawings added by the reader (Figure 5), the folding down of a page corner to create a dog-ear, scribbles that test a writing tool, stamps indicating ownership, traces of wax used to adhere materials to the page, and labels or bookplates pasted onto the page.

Each type of annotation, for example, "underlining," encompasses all occurrences of that type on a single page. The database also evaluates the density or weight of RMs on a scale from 1 to 3, though finding a practical and conceptual way to measure this quantitatively has proven challenging. RMs are further detailed through four categories of tags. Position tags indicate the specific locations of RMs on the page. Descriptive tags aim to capture the semantic meaning or intended purpose of the RMs. Content tags describe the conceptual content related to the RM, such as scientific concepts, while material tags provide information about the materials used in the annotation, including the type of writing tool. Whenever possible, individual RMs are transcribed and stored, and any damage or modification to the page associated with an annotation itself is recorded. If the creator of an annotation is known, this information is linked to the ownership history of the book copy, providing a pragmatic method for tracing the provenance of RMs.



ca 23 non plus minus ve sensibiliter propter SOLIS LUMEN et RADIOS Variè excipien= a moveri lente in eadem distantia, propter FIXARUM [illegible] dos in superficiem Telluris.

DE MAGNETE LIB. VI. CAP. VIII. 228

reſpiciebat ; nunc tantùm gradibus duobus, & minutis 52, ab eodem diſtat : undè à propinquitate, à recentioribus Polaris dicitur. diſtabit autem aliquandò à polo tantùm ½ unius gradus; poſteà verò à polo incipiet recedere, donec ab eo abſit grad. 48. : quod (ſecundùm tabulas Prutenicas) erit anno domini 15000. Ità lucida lyræ [quæ nobis Britannis Auſtralibus jam verè culminat] polo mundi aliquando, ad quintum ferè gradum appropinquabit. Sic omnes Stellæ immutant ſuos luminis radios in ſuperficie telluris, admirabili hac magneticâ axis telluris inflexione. Hinc temporum anni novæ varietates, terræque fœcundiores magisvè ſteriles evadunt; hinc gentium ingenia & mores immutantur; regna & leges alterantur, pro Stellarum fixarum virtute, & robore ſuſcepto aut amiſſo; pro ſingulari & ſpecificâ naturâ fixarum culminantium : aut propter novas in alijs Zodiaci locis cum Planetis configurationes; propter Ortus etiam & Occaſus, & concurſus in Meridiano novos. Præceſſio Æquinoctiorum, ex motu æquali poli telluris in circulo Zodiaci arctico, hîc demonſtratur.

NB

Sit Ecliptica linea A, ~~B, C, D~~ B, D; circulus Zodiaci arcticus I, E, G. Si ergò polus telluris reſpexerit E, tunc æquinoctia ſunt in D, C. ~~Sit hoc tempore Methonis~~, quando cornua Arietis erant in coluro æquinoctiorum. 1.

Quòd ſi polus telluris proceſſerit uſque ad I, tunc æquinoctia erunt K, L; & Stellæ in C, ecliptica, ~~in ſucceſſionem~~ ſecundum ſignorum pro- 2.

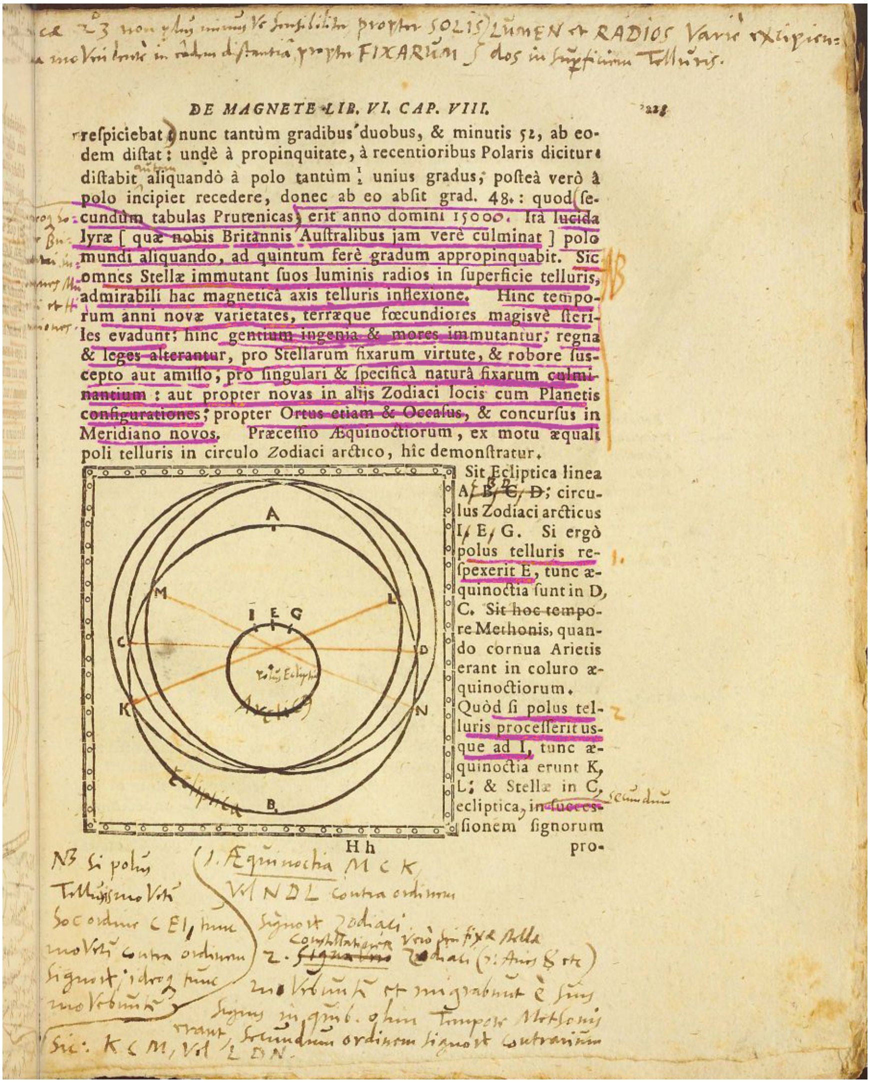
H h

NB Si polus Telluris [illegible] moveri secundum CEI, tunc moveri contra ordinem signorum, idque tunc movebuntur

(1. Æquinoctia MCK vel NDL contra ordinem signorum Zodiaci.
2. ~~Signatura~~ constellationes verò seu fixæ stellæ Zodiaci (v: Aries & etc) movebuntur et migrabunt è suis signis in quib. olim tempore Methonis erant, secundum ordinem signorum contrarium

Sic: KCM, vel LDN.

Figure 3: A page from William Gilbert, *De Magnete* (1628), annotated by Andreas Granius. Pink highlights (added digitally) extend to the annotation type for additions of handwritten text (Source: HAB Wolfenbüttel: M: Nc 4° 46 <https://diglib.hab.de/drucke/nc-4f-46/start.htm>, accessed February 22, 2025).



ca 23 non plus minus ve sensibiliter propter SOLIS LUMEN et RADIOS Varie excipien-tes Veri tracti in eadem distantia, propter FIXARUM [illegible] dos in superficiem Telluris.

DE MAGNETE LIB. VI. CAP. VIII. 228

reſpiciebat) nunc tantùm gradibus duobus, & minutis 52, ab eodem diſtat: undè à propinquitate, à recentioribus Polaris dicitur: diſtabit autem aliquandò à polo tantùm ½ unius gradus; poſteà verò à polo incipiet recedere, donec ab eo abſit grad. 48.: quod (ſecundùm tabulas Prutenicas) erit anno domini 15000. Ità lucida lyræ [quæ nobis Britannis Auſtralibus jam verè culminat] polo mundi aliquando, ad quintum ferè gradum appropinquabit. Sic omnes Stellæ immutant ſuos luminis radios in ſuperficie telluris, admirabili hac magneticâ axis telluris inflexione. Hinc temporum anni novæ varietates, terræque fœcundiores magisvè ſteriles evadunt; hinc gentium ingenia & mores immutantur; regna & leges alterantur, pro Stellarum fixarum virtute, & robore ſuſcepto aut amiſſo; pro ſingulari & ſpecificâ naturâ fixarum culminantium: aut propter novas in aliis Zodiaci locis cum Planetis configurationes; propter Ortus etiam & Occaſus, & concurſus in Meridiano novos. Præceſſio Æquinoctiorum, ex motu æquali poli telluris in circulo Zodiaci arctico, hîc demonſtratur.

Sit Ecliptica linea A, ~~B, C, D~~; circulus Zodiaci arcticus I, E, G. Si ergò polus telluris reſpexerit E, tunc æquinoctia ſunt in D, C. ~~Sit hoc tempore Methonis~~, quando cornua Arietis erant in coluro æquinoctiorum. Quòd ſi polus telluris proceſſerit uſque ad I, tunc æquinoctia erunt K, L; & Stellæ in C, ecliptica, ~~in ſucceſ~~ſionem ſecundum ſignorum pro-

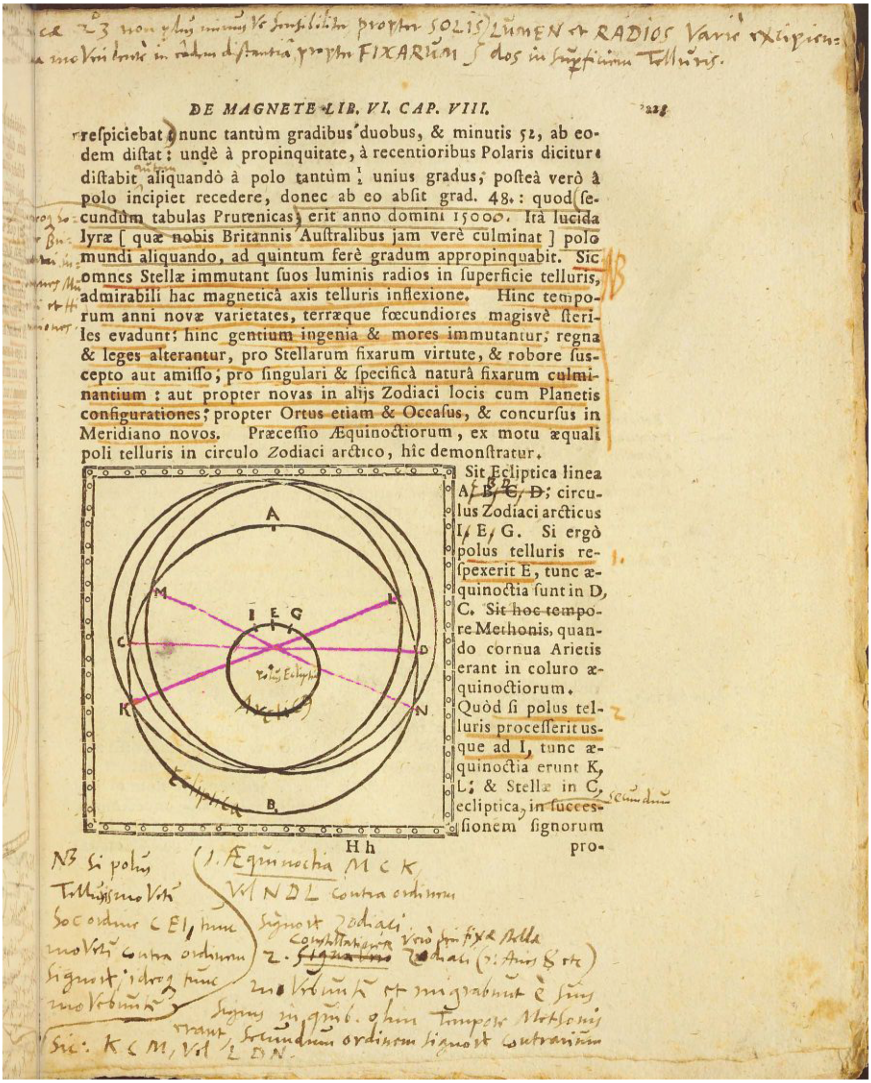
H h

NB Si polus Telluris moveti Sec ordinem C E I, tunc moveti contra ordinem signorum, idque tunc movebuntur

(1. Æquinoctia M C K vel N D L contra ordinem signorum Zodiaci
constellationes vero seu fixæ stellæ
2. ~~signa~~ Zodiaci (v: Aries & etc) movebuntur et migrabunt e suis signis in quib. olim tempore Methonis erant, secundum ordinem signorum contrarium

Sic: K C M, vel L D N.

Figure 4: A page from William Gilbert, *De Magnete* (1628), annotated by Andreas Granius. Pink highlights (added digitally) extend to the annotation type for underlinings of printed text (Source: HAB Wolfenbüttel: M: Nc 4° 46 <https://diglib.hab.de/drucke/nc-4f-46/start.htm>, accessed February 22, 2025).



...ca 23 non plus minus ve sensibiliter propter SOLIS LUMEN et RADIOS varie excipien-
...a mediocri in eadem distantia propter FIXARUM radios in superficiem Telluris.

DE MAGNETE LIB. VI. CAP. VIII. 228

reſpiciebat; nunc tantùm gradibus duobus, & minutis 52, ab eodem diſtat: undè à propinquitate, à recentioribus Polaris dicitur. diſtabit autem aliquandò à polo tantùm ½ unius gradus; poſteà verò à polo incipiet recedere, donec ab eo abſit grad. 48.: quod (ſecundùm tabulas Prutenicas) erit anno domini 15000. Ità lucida lyræ [quæ nobis Britannis Auſtralibus jam verè culminat] polo mundi aliquando, ad quintum ferè gradum appropinquabit. Sic omnes Stellæ immutant ſuos luminis radios in ſuperficie telluris, admirabili hac magneticâ axis telluris inflexione. Hinc temporum anni novæ varietates, terræque fœcundiores magisvè ſteriles evadunt; hinc gentium ingenia & mores immutantur; regna & leges alterantur, pro Stellarum fixarum virtute, & robore ſuſcepto aut amiſſo; pro ſingulari & ſpecificâ naturâ fixarum culminantium: aut propter novas in alijs Zodiaci locis cum Planetis configurationes; propter Ortus etiam & Occaſus, & concurſus in Meridiano novos. Præceſſio Æquinoctiorum, ex motu æquali poli telluris in circulo Zodiaci arctico, hîc demonſtratur.

Sit Ecliptica linea A, ~~B, C, D~~ B D; circulus Zodiaci arcticus I, E, G. Si ergò polus telluris reſpexerit E, tunc æquinoctia ſunt in D, C. ~~Sit hoc tempore Methonis~~, quando cornua Arietis erant in coluro æquinoctiorum. Quòd ſi polus telluris proceſſerit usque ad I, tunc æquinoctia erunt K, L; & Stellæ in C, ecliptica, ~~in ſucceſſionem~~ secundum ſignorum pro-

H h

NB Si polus Telluris moveretur hoc ordine C E I, tunc moveretur contra ordinem signorum, idemque tunc movebuntur

1. Æquinoctia M C K vel N D L contra ordinem signorum Zodiaci.
2. ~~signatum~~ constellationes vero seu fixæ stellæ Zodiaci (i.e. Aries & etc) movebuntur et migrabunt e suis signis in quibus olim tempore Methonis erant, secundum ordinem signorum contrarium

Sic: K C M, vel L D N.

Figure 5: A page from William Gilbert, *De Magnete* (1628), annotated by Andreas Granius. Pink highlights (added digitally) extend to the annotation type “drawing” (Source: HAB Wolfenbüttel: M: Nc 4° 46 <https://diglib.hab.de/drucke/nc-4f-46/start.htm>, accessed February 22, 2025).

3 Analysis

This section will focus on three computational techniques employed to interpret the complex relationships between textual content and reader interactions: frequency analysis, co-occurrence analysis, and the exploration of multidimensional feature spaces. These analytical methods allow for a discovery or confirmation of (perhaps presumed but to be confirmed) patterns that characterize historical reader engagement in their thematic emphasis.

Before presenting the methods and illustrating their potential using the Magnetic Margins dataset, it is important to note that this dataset is based on a census, that is, a comprehensive catalog of all copies associated with the same work. This enables us to focus on specific chapters that appears in all editions and copies. We safely assume that the content of each chapter is identical across these copies and can therefore investigate how frequently it was annotated. If the thematic content of the chapter is known, this already provides insight into the interests of historical readers. In this way, a SF (i.e., some conceptual content) can be proxied directly by a SE, in this case, the chapter itself. This approach does not apply when dealing with multiple distinct works within a single dataset, as each work will have its own chapter structure and content. In such cases, observing that some "Chapter 2" is the most frequently annotated across different works merely reflects a shared topological position, not identical content or comparable reader interests. The following analyzes are therefore based on the *De Magnete* census, which ensures content comparability, but the methods are adaptable to other – non-census included – datasets by identifying more appropriate SF, such as recurrent topics instead of chapters.

The dataset is queried for RMs in chapters (SE) of the work *De Magnete*, including all its editions and all their copies.[32] RMs were filtered to exclude editorial marks by librarians or book sellers and marks of ownership as they are not seen as an intellectual interaction with a text's content, occur in great numbers, and are mostly easy to identify. At the same time, RMs are weighted by a value indicating "how much annotation" one annotation record comprises. Chapters were filtered to only include content of the main text, excluding paratexts such as title pages or tables of contents for which any intellectual interaction is difficult to relate to the

[32] Any analysis requires a preparation of analytical data that is generated from queries against the database. Put simply, querying the data means that a complex structured database is "asked" to return certain values, typically as a table. Tables are flat as they only contain rows and columns while the data itself is "deep." Querying all RMs in one work means to "flatten" the structure in between, including editions and copies.

printed text. As a result, each copy is represented as a vector of SFs (e.g., a topic in a chapter or the chapter itself, depending on what has been defined as the type of SF taken into consideration). If, for example, $copy_1$ had SF_1 with a weighted sum of 3 and SF_2 of 1, the vector for this copy would be [3,1]. Vectors like this are the basis for the subsequent analyzes.[33] An abstract outline of the three analytical methods – frequency analysis, co-occurrence analysis, and complexity analysis – will be followed by a more concrete example and a discussion of the insights they yield.

Frequency analysis aims to examine the distribution of data points across a single variable. The data consists of a one-dimensional set of values for the variable X. By determining the frequency of values within X these frequencies can be represented, e.g., as bars in a histogram. Here, the focus is on the sum of RMs per SF in every copy (F(X)). This analysis helps in understanding which SFs were most frequently highlighted by readers.

In the example dataset, it is less evident to search for the number of RMs per topic as SF, given that all copies ultimately contain the same text, which exhibits notable overlaps in its topics. Put simply, Several chapters deal with very similar content. To achieve a greater spread of topics, it is better to analyze the RMs per chapter across all copies of different editions. Counting the number of RMs per chapter diversifies the features – each chapter is a SF then. Since the chapters themselves undoubtedly have disparate thematic foci the result still allows for semantic interpretation.[34]

In order to further enhance the significance, two different ways of data normalization and scaling can be applied. A simple aggregation of weighted RMs per copy per SF, however, skews the data: Copies with a very high number of RMs will overstress the distribution of annotation per SF. A few heavily annotated copies would determine which SFs dominate the entire analysis across all copies, which is not desirable. Instead, in a normalization scenario we might call “frequency” the number of weighted RMs per copy per SF is multiplied by the number of how many copies overall have annotated this SF. This still reflects greater or less interest into some SF per copy but takes into account if the SF is also annotated in other copies:

[33] Vectors can be understood as coordinates in a space whose number of dimensions corresponds to the number of features or positions in the vector. For example, the vector [3, 1] defines a point in two-dimensional space, i.e., a coordinate on a plane. When vectors span hundreds of dimensions, they exceed the limits of human phenomenological intuition, yet the underlying principle remains the same.

[34] A manual annotation per chapter is an alternative approach and will be described below.

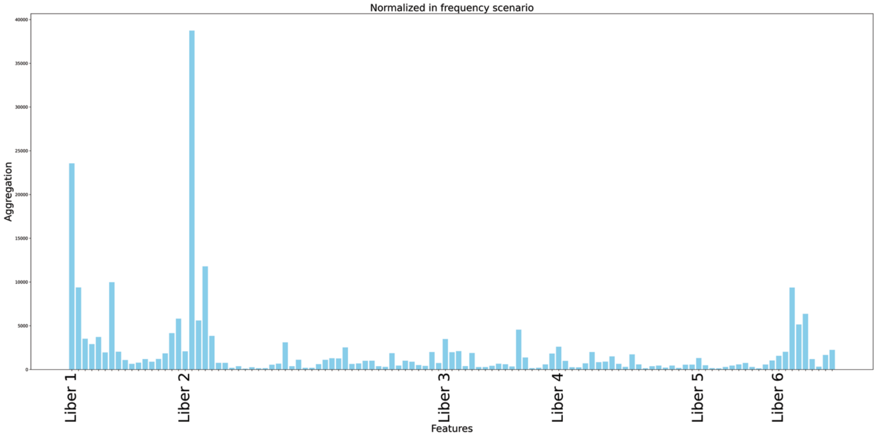


Figure 6

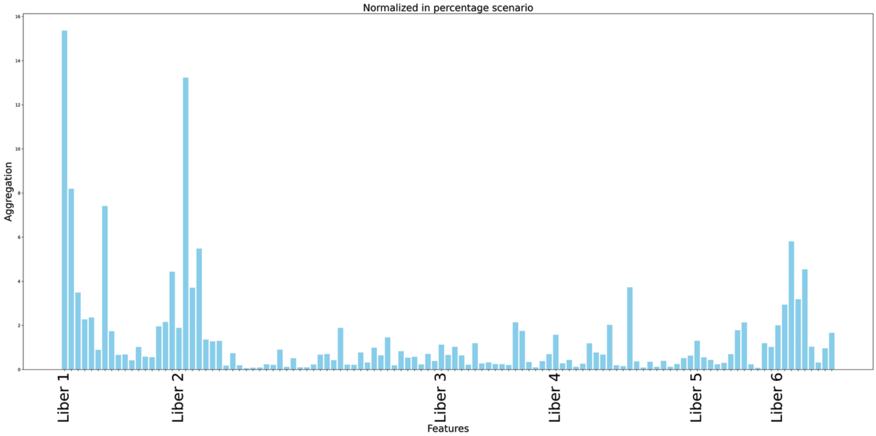


Figure 7

Figures 6 and 7: Bar charts of aggregated and weighted annotations per chapter in Gilbert's *De Magnete* across all analyzed copies, ordered by the chapters' original order in the work. Figures 6 ("frequency") and 7 ("percentage") have different normalizations for counting annotations per copy, resulting in different profiles. Source: Plots created by Christoph Sander, using Matplotlib in Python.

SF (e.g., chapters) that are annotated more frequently are weighted more strongly in the final dataset.

For a normalization scenario called “percentage” the aggregated RMs per copy are divided by the total RMs of the respective copy as values normalized, resulting in the percentage of a feature per copy.[35] This ensures that a few copies with a disproportionate absolute number of RMs for specific features do not influence the overall impression. Instead, this normalization focuses on the distributions of the interest into some SF within one single copy. So, while “frequency” normalization still accounts if a copy has 10 or 100 RMs, “percentage” disregards absolute counts and instead focuses on the relative distribution of RMs across features within each copy. In the former scenario, a two-feature vector would distinguish between [1,2] and [10,20] while in the latter scenario both would be interpreted as [0.25,0.75]. “Percentage” normalization thus highlights *what* a reader was most interested in, without indicating *how much* it interested them. In contrast, “frequency” normalization captures both the direction and intensity of attention – adjusted by how many other copies contain RMs for the respective semantic feature.

Upon examination of the results (Figures 6 and 7), two chapters, Chapter 1 in Book 1 and Chapter 2 in Book 2, in particular stand out, regardless of the normalization scenario. These chapters can be described as doxographic chapters, i.e., chapters with a high density of references to authors. Concurrently, in a topological perspective, the analysis demonstrates that the middle section of the work in particular shows significantly fewer RMs. This is the part of the work with the most experiments and the practical application of magnetism. A highly abbreviated interpretation suggests that the RMs are concentrated above all where natural philosophical and natural history debates and discussions with Gilbert’s predecessors take place rather than in the chapters in which Gilbert describes his experiments in detail.[36] The relatively minor peak at the end of the book also indicates that the cosmological topics are of greater interest than the more experimental and practical sections on navigation and instrumentation. An alternative and supplementary interpretation – that however needs corroboration from other data – would suggest that readers focus(sed) on the initial and the final part of a book they read, regardless of content or interest.

[35] Other scales and normalizations of the numerical data were applied as well, with quite different outcomes in subsequent analysis. This study lacks the space to discuss each scenario.

[36] Margócsy, Somos, and Joffe, *The Fabrica of Andreas Vesalius*, 95: “While it is true that Vesalius exhorted his readers to adopt a hands-on approach to anatomy, his readers did not necessarily pay attention. Instead, they turned to books. Less than 10% of the annotators mention their own experiences with studying the human body, while more than half include a reference to another book or author.”

However, if all RMs are aggregated not per chapter but per book, encompassing several chapters, and these books are manually tagged with twenty thematic features, a confirming but less differentiated picture emerges (Figures 8 and 9). Thematic differentiation within a given book, comprising many chapters with different thematic foci, is given up in favor of a more general semantic description of its contents. For Gilbert's work, this is a sound approach as each book has its designated overarching topics.

Co-occurrence analysis aims to identify relationships and patterns between two variables (X and Y) by examining how often specific combinations of their values appear together. The result is a two-dimensional dataset, F(X, Y), that records the frequency of each X–Y pairing. In this context, it involves counting how often one semantic feature (SF_1) co-occurs with another (SF_2) within the same copy, e.g., how frequently two topics are annotated together by multiple readers. This yields a co-occurrence matrix, which can be visualized as a heatmap: SF_1 and SF_2 define the rows and columns, while each cell indicates the frequency with which the corresponding pair appears together. A different type of normalization of the data here is useful. PMI (Pointwise Mutual Information) quantifies how much more often two values co-occur than would be expected by chance, thereby highlighting meaningful associations.[37] If, e.g. most copies annotate some SF_x, it is likely for this SF_x to co-occur with any other SF_y, hence their correlation is less specific and insightful and would likely happen "by chance." PMI corrects for this by emphasizing co-occurrences that are unexpectedly frequent, helping to surface more expressive and significant patterns.

When a chapter covers multiple topics and topic co-occurrence is based on aggregated annotations per chapter, the resulting correlations primarily reflect relationships between topics within the chapter itself, rather than meaningful associations between topics and RMs. Given the thematic narrowness of the single work represented in this dataset, such an approach does not yield analytically useful results and is therefore not a suitable methodology in this context. Nevertheless, correlations between individual chapters are analytically significant, as they reflect which chapters attract the interest of the same readers (Figure 10), with chapters serving as proxies for thematic focal points. Particularly noteworthy are the strong correlations between topologically adjacent chapters – an unsurprising result, given their often close conceptual relationship and the tendency of readers to engage with them jointly. Additionally, a striking pattern emerges: readers who

[37] See Kenneth Ward Church and Patrick Hanks, Word Association Norms, Mutual Information, and Lexicography, *Computational Linguistics* 16, no. 1 (1990): 22–29.

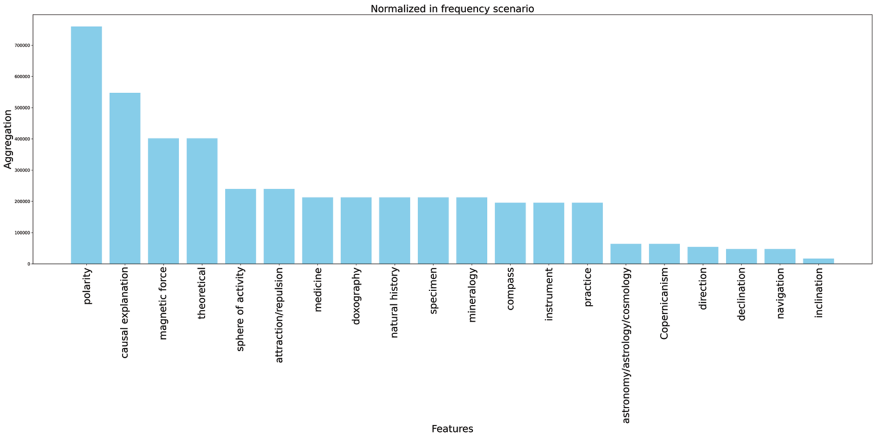


Figure 8

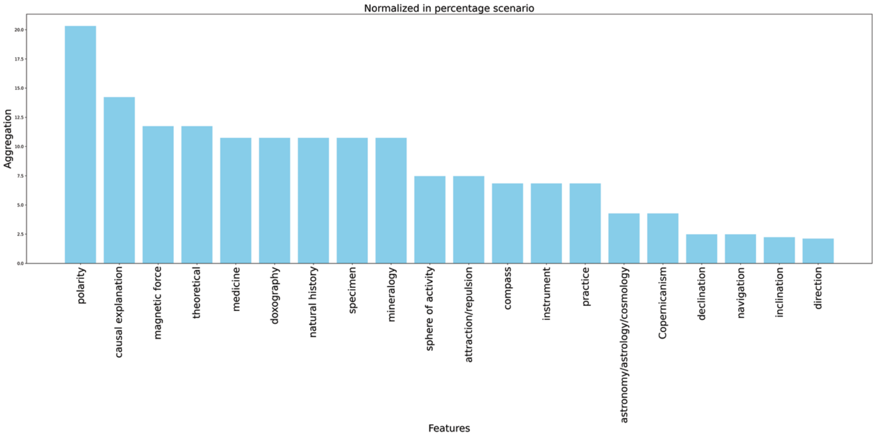


Figure 9

Figures 8 and 9: Bar charts of aggregated and weighted annotations per topic in Gilbert's *De Magnete* across all analyzed copies, ordered by accumulated sum per topic. Figures 8 ("frequency") and 9 ("percentage") display different normalizations of the data for counting annotations per copy, resulting in different profiles and in a different descending ordering of topics. Source: Plots created by Christoph Sander, using Matplotlib in Python.

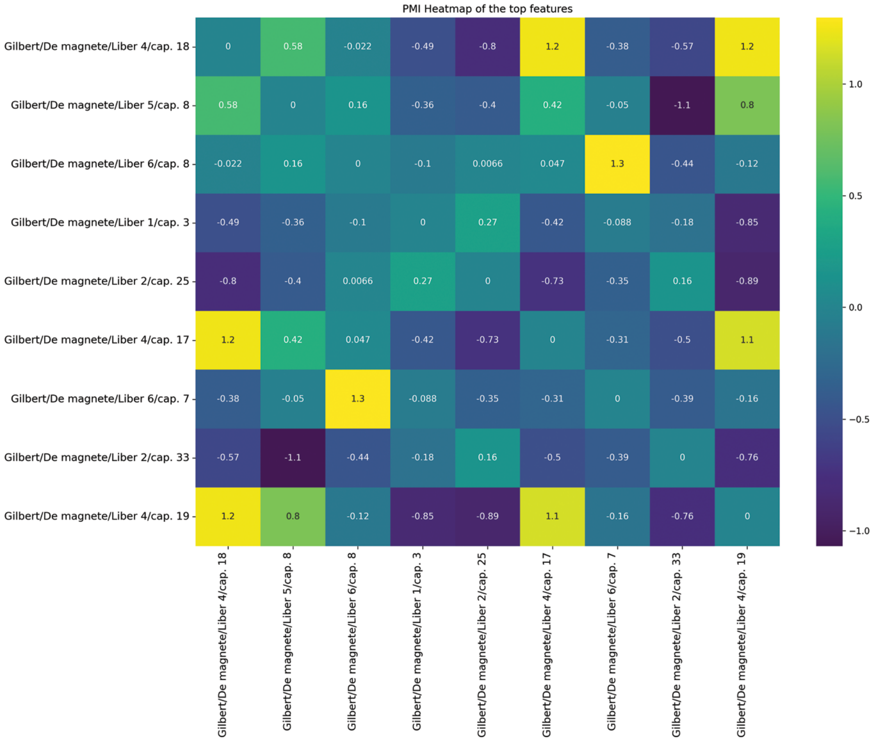


Figure 10

showed marked interest in the first two books of Gilbert's oeuvre frequently displayed little interest in his concluding cosmological work, and vice versa.

These correlations can be analyzed not only for semantic associations but also with regard to formal characteristics. Different types of annotation are significantly more often found in certain chapters (Figure 11) sometimes due to contingent factors. For example, drawings added by readers often feature on chapters that give a woodcut diagram themselves, and crossings out are particularly often found in chapters in book six on the Copernican hypothesis: a clear indication of the censorship this part underwent. For example, analyzing the types of RMs in relation to their positions reveals clear and often trivial patterns (Figure 12). When relating annotation types to the positions they occur in, pasted labels are typically found on blank pages, while underlining is almost never found in the margins. This is to be expected. Moreover, it confirms the previous observation on readers' draw-

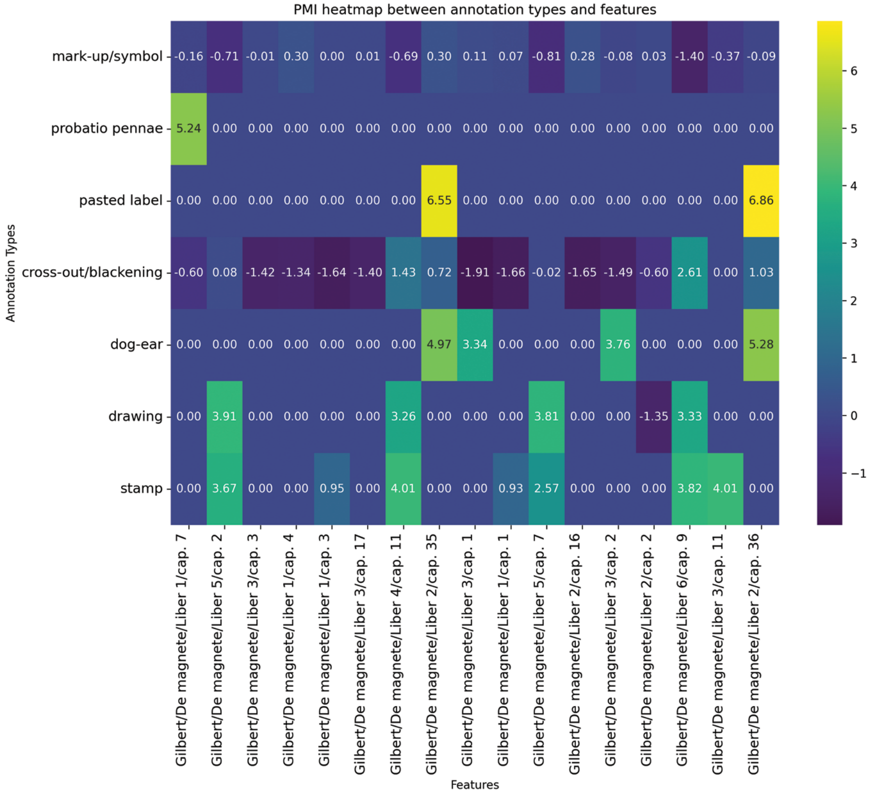


Figure 11

ings that the addition of drawings was not significantly more frequent in or next to the text, but rather within printed illustrations in the book. This indicates a dialog between two graphical representations rather than a multimodal visualization of the textual content.

Cluster analysis and complexity reduction aims to identify groups or clusters of similar data points within a high-dimensional dataset. The data consists of multiple vectors (X_1, X_2, ..., X_n), each having the same number of dimensions or features described as a numerical value.[38] The process involves reducing the dimension-

[38] These values can be a one-hot encoding, if only 0 and 1 are realized, e.g. if some feature is present or not. This is usually done for categorical data, e.g. if some feature is a category such as

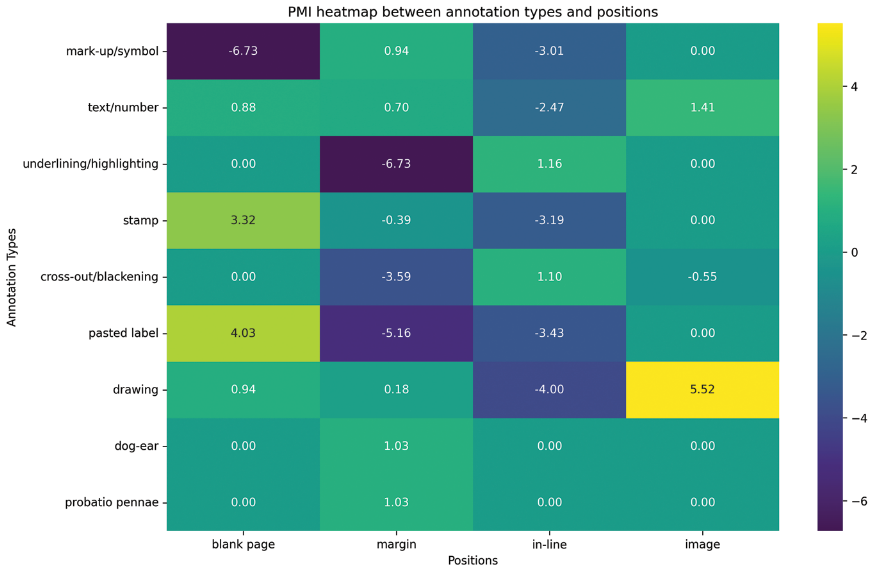


Figure 12

Figures 10–12: Heatmaps showing top and bottom co-occurring features computed as Pointwise Mutual Information. High positive values indicate frequent significant associations, low negative values indicate significant and exceptional associations. Figure 10 associates chapters with chapters, revealing how significantly annotations in chapter A and in chapter B co-occur in the same copy. Figure 11 associates chapters with annotation types, revealing how significantly annotations of some type are found in some chapter across copies. Figure 12 associates annotation types with annotation positions, revealing how significantly annotations of some type are found in some position on the page across copies. Source: Plots created by Christoph Sander, using Seaborn in Python.

ality of the data using dedicated algorithms to project the high-dimensional data onto a 2D plane.[39] This allows for a visual identification of clusters. Then, clustering algorithms mathematically group similar data points based on vector similar-

"drawing" or "underlining." While each category can be represented by a number, e.g. 1 for "drawing" and 2 for "underlining," this is not reasonable for a vector unless the order of these numbers reflects a semantic order or ranking as well. A higher number would mean a "more" of something which is for strict categories not the case. However, the (normalized) sum of weighted RMs per SF is a suitable numerical value to use, while it neglects the RMs' categorical information and only scales their weighted occurence.

[39] Algorithms include PCA, t-SNE, and UMAP. All of them are included in the above-mentioned scikit-learn Python package (see note 11).

ity.[40] These clusters often match visual clusters in the projection but are based on different algorithms and may rely on the high-dimensional feature space instead of the projection. A third step in this pipeline is the identification of features that drive the assignment to a cluster, which is referred to as feature importance.[41] Modern machine learning techniques readily support all three tasks: dimensionality reduction, cluster detection, and interpretation through feature importance.

This analysis's main advantage is that it helps in understanding which copies are "similarly annotated" and may share underlying connections or patterns. We thereby can identify which copies of an edition resemble others based on whether the same chapters or SF were annotated. The data thus consists of multiple dimensions or components, with each dimension representing a SF, e.g., a chapter in the work. Each copy's vector indicates with each its dimension how much (numerically) a particular SF was annotated. This dimension value depends on the above-described normalization scenarios. Applying the "frequency" scenario will highlight mostly similarities between densely annotated copies, as they have high scores related to the same SF. The "percentage" scenario aims at the focal SF per copy and not so much on the mere number of RMs per copy. Whether the SF is represented by the chapter, the book, or a (manually assigned) tag of a book allows for another variance of the results.

An examination of calculated projections reveals varying clusters depending on the normalization scenario. Inspecting local neighborhoods allows for the verification of marked similarities on a case-by-case basis. In the "frequency" scenario, for some of the heavily annotated copies, which are plotted with correspondingly scaled dots in the plot (Figure 13), the discrimination into different clusters is particularly noteworthy. A comprehensive analysis is beyond the scope of this study. However, an examination of one cluster, for instance, reveals that a copy from the Herzog August Library in Wolfenbüttel (copy with ID 1357) stands out from the corpus due to its extensive RMs, with these RMs concentrated primarily on Books 4 and 5, which are comparatively less annotated in other copies. Other copies have high number RMs as well – but not in these chapters. The identification of such an outlier can thus be significantly facilitated by the projection. This oddball would be described as a "heavy-annotator" with an interest in topics that are not what other "heavy-annotators" were concerned with.

[40] Algorithms include k-Means and (H)DBSCAN. All of them are included in the above-mentioned scikit-learn Python package (see note 11).

[41] Algorithms include RandomForestClassifier and SHAP. The former is included in the above-mentioned scikit-learn Python package (see note 11), the latter in its own.

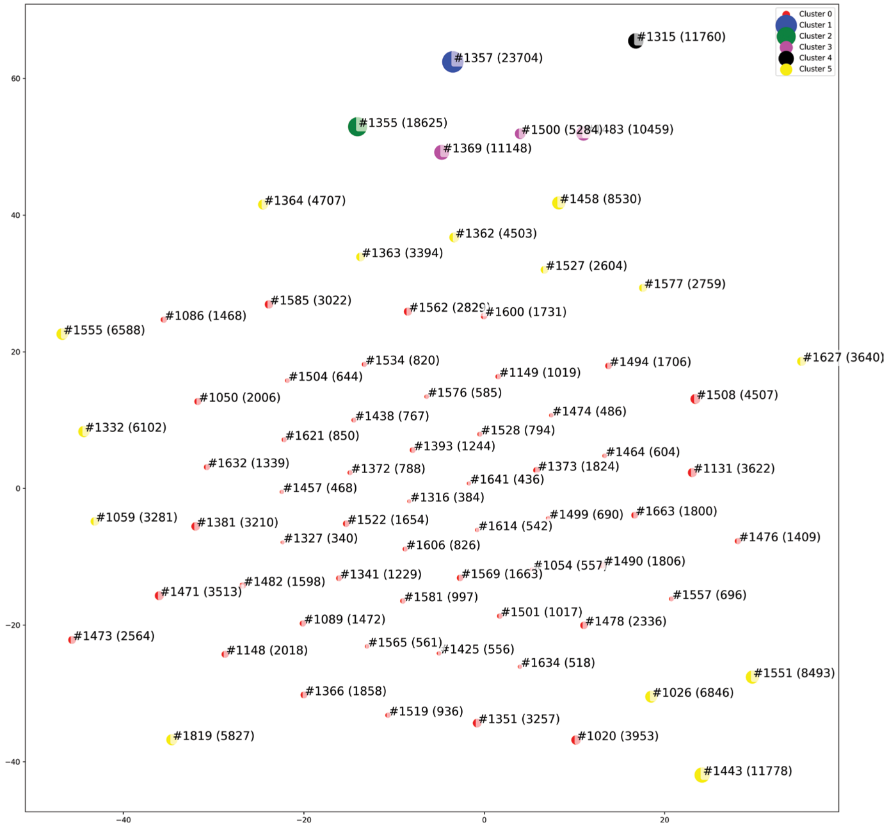


Figure 13

In the "percentage" scenario, when RMs are aggregated per book, grouping multiple chapters into a single unit, we find different clusters.[42] While a six-dimensional space, equivalent to the number of books in Gilbert's work, is rather dense, the six books can neatly be associated with a range of distinct topics (as SF), making this feature space semantically richer. The assignment of distinct topics to each book serves both to enhance the interpretability of the results as well as for accounting for potential thematic overlap across individual books. When this multidimensional feature space is projected and grouped into distinct clusters, clear

[42] If the feature space is based on chapters annotated, it becomes excessively broad due to the high numbers and a large variance of values. A significant proportion of chapters are rarely or never annotated, which results in high perplexity and indistinct clusters.

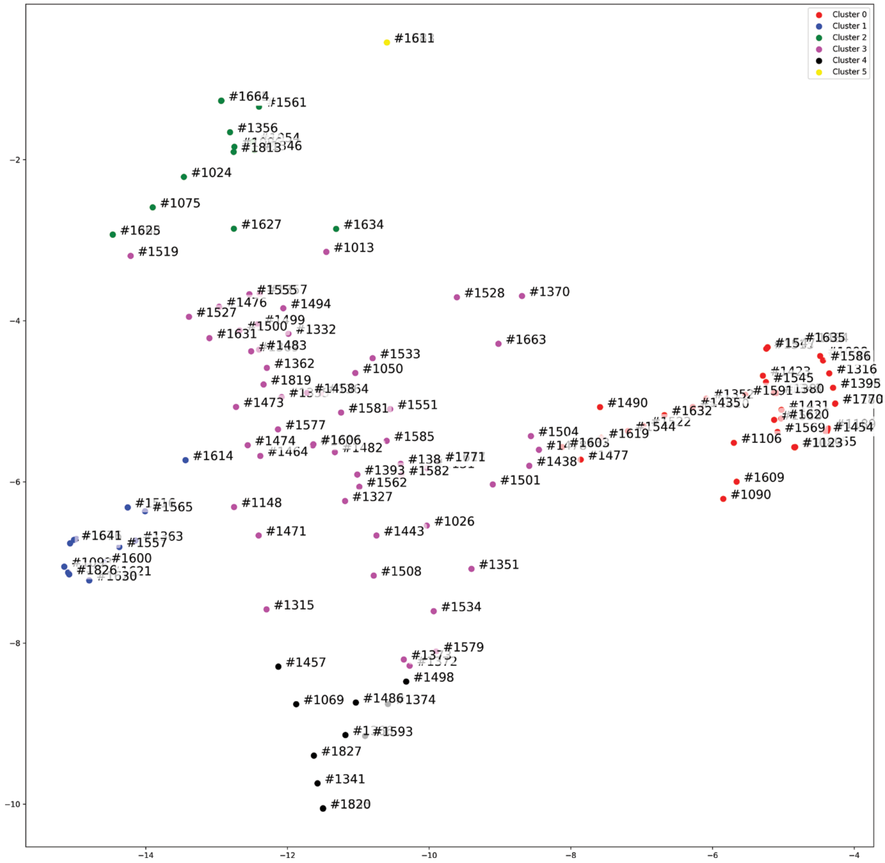


Figure 14

Figures 13 and 14: The t-SNE (t-Distributed Stochastic Neighbor Embedding) algorithm projecting multidimensional data (copies represented as n-dimensional vectors) onto a 2D plane in a plot. Colors map to k-Means clusters, based on high-dimensional feature space. Included copies were thresholded with regard to the amount of annotations to avoid sparse vectors and noisy results. Figure 13 is based on the "frequency" normalization scenario for chapters as features, showing smaller clusters (less members, clusters of outliers) of semantic resemblance. Figure 14 is based on the "percentage" normalization scenario for semantic tags as features, showing larger clusters and a more pronounced topology of semantic neighborhoods. Source: Plots created by Christoph Sander, using Matplotlib in Python.

profiles (Figure 14) emerge. Indexing RMs by book-level topics and applying "percentage"-based normalization proves especially insightful: each copy receives a fingerprint of "focal interest" based on the relative attention given to each topic. This approach highlights clusters of copies that share similar thematic emphases, rather than merely distinguishing between heavily and lightly annotated texts.

To better understand the profiles of the identified clusters, it is crucial to examine the feature importance associated with each cluster. This allows for a more nuanced understanding of the motivations behind a cluster assignment. Since clustering algorithms do not assign clusters based on feature importance directly, uncovering which features drive these assignments requires a post-hoc explanatory step: training a prediction model (a subsymbolic, weak and supervised artificial intelligence model) on the given cluster assignment. The model is then observed during its prediction process, allowing us to infer which features it relies on. This explainable AI approach helps to open the "black box" of unsupervised clustering and reveals the thematic dimensions that differentiate the clusters. The result, as one might expect, is not grounded in any feature's frequency in a cluster, but rather on a normalized weighting that also considers which themes are addressed together and which are not addressed at all. A bar chart (Figure 15) makes it evident that the absolute frequency of a feature (feature sum, blue axis) is not a reliable indicator of its importance in determining cluster assignment.[43] Some features are frequently assigned to multiple clusters and hence do not distinguish them, e.g. an interest in "causal explanation" – a feature assigned to various chapters.

Looking at each cluster reveals that some clusters represent a rather diffuse annotation pattern across the majority of books and themes while others exhibit a more specific focus. From a historiographical perspective we can identify one outstanding cluster of eighteen copies with a primary and specific interest in the Copernican aspects of Gilbert's work (Figure 16). This cluster comprises copies that exhibit a distinct and overrepresented interest in this topic relative to other topics. Consequently, it excludes copies with a substantial interest in the Copernican issues but also a relatively high interest in other topics concurrently.[44] This cluster of readers displays a pronounced focus on book six, with comparatively limited interest in other content.

This "Copernican" cluster overrepresents the features "astronomy/astrology/cosmology" and "Copernicanism." These two features make up more than 50% of what distinguishes this cluster from others, while other features are frequently present as well but do not distinguish this cluster from others. Validating this

[43] A box plot can also be employed to graphically display variance, thereby indicating whether the presence of a feature is relatively uniform across all instances of the cluster or if there are significant outliers.

[44] In this case, we might justify their exclusion as these readers didn't show a distinct interest into the Copernican parts, which might be, but not necessarily is, a different "type of reader."

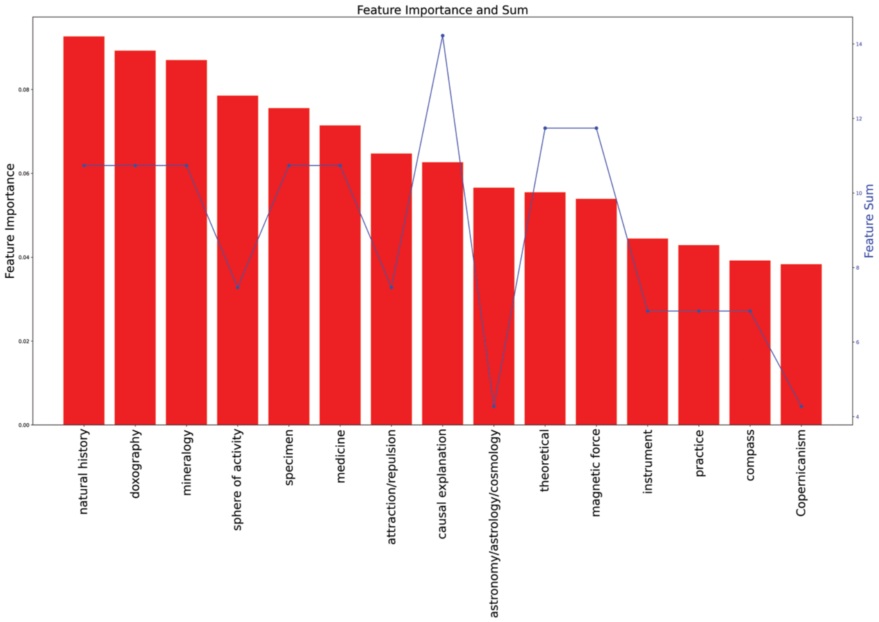


Figure 15

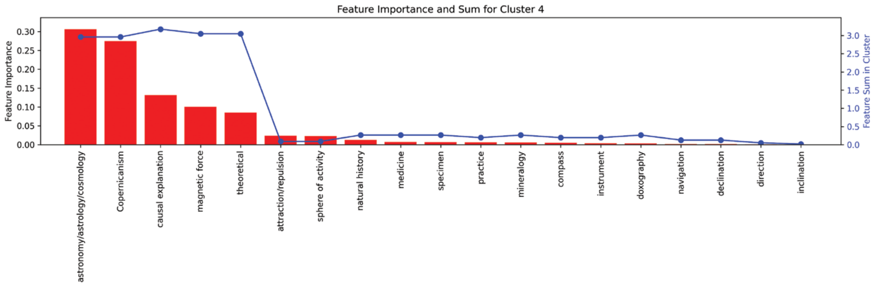


Figure 16

Figures 15 and 16: Charts visualizing feature frequency (blue line) and importance (red bars), based on “percentage” data normalization. Frequencies are aggregated sums of all annotations per feature of all copies (Figure 15) or those in Cluster 4 (Figure 16). Deviations between both metrics indicate that high frequency does not indicate high importance for cluster assignment. High importances may also indicate the low presence of a feature in copies in a cluster. Importances were calculated through RandomForestClassifier and Mean Decrease in Impurity in the feature_importances_ attribute in Python’s scikit-learn package. Source: Plots created by Christoph Sander, using Matplotlib in Python.

cluster integrity by studying its members, i.e., the copies, is out of scope here, but as a proof of concept a few preliminary observations can be made.

1. Bologna, Università di Bologna, Dipartimento di Fisica ed Astronomia DIFA. Biblioteca di Astronomia, DC-4 0038
2. Florence, Biblioteca Nazionale Centrale di Firenze, RARI B.R.121
3. Milan, Archivio Storico Civico e Biblioteca Trivulziana, Mor.F.20
4. Rome, Biblioteca dell'Accademia Nazionale dei Lincei e Corsiniana, ACCAD 349. H. 25
5. Rome, Biblioteca Nazionale Centrale di Roma, 55. 10.D.9
6. Offenbach a. M., Deutsche Meteorologische Bibliothek, 219060
7. Vienna, Österreichische Nationalbibliothek, Rara *69.B.74
8. Wrocław, Biblioteka Uniwersytecka we Wrocławiu, BU-Stare Druki 401922
9. Gent, Universiteitsbibliotheek, BIB.PHYSCHIM.000067
10. London, British Library, 33.b.20.
11. Jerusalem, The National Library of Israel, 4= 62 D 1035
12. Pittsburgh, Carnegie Mellon University Libraries, Posner Memorial Collection, QC751 .G44, copy 2
13. Berkeley, The Bancroft Library, University of California, Bancroft Vault QC751 .G44 1600
14. Boston, Boston Athenæum, HCM .G37
15. Cambridge, University of Cambridge, Pembroke College, 10.9. 9
16. Glasgow, University of Glasgow Library, Sp Coll Ferguson Ag-a.27
17. London, Royal College of Physicians Library, D2/55-h-4
18. New York, The New York Academy of Medicine Library, RB

Its 18 copies include, e.g., a copy owned by Galileo Galilei (no. 2).[45] In fact, Galileo made few annotations, but only related to the work's astronomical sections. A similar focus can be corroborated for other instances, too. This clustering and feature importance detection thus reveals a group of clusters that share characteristics that may lead to historical insights, such as: These copies were read by readers, who predominantly annotated astronomical content.

[45] See Mario Loria, William Gilbert e Galileo Galilei: La terrella e le calamite del granduca, in: *Saggi su Galileo Galilei*, vol. 2, Carlo Maccagni (ed.), 208–247, Florence: Giunti Barbèra, 1972; Sander, *Magnes*, 833.

4 Discussion

The hermeneutic principle underlying the previous analysis is based on the assumption that the intentional material engagement with material carriers of semantic content can reveal insights about the reception and assessment of that content. This approach aims to integrate material history with intellectual history. To employ this approach effectively, it relies on an interoperable modeling of relevant phenomena as structured data, in addition to a transparent analytical pipeline that forms the basis for discovery and explanation of pattern.[46]

On an interpretive level, it is important to recognize that thematic foci for a reader to annotate may not necessarily or exclusively reflect substantive individual interests, but also testify to wider practices of annotating and reading itself. The significant focus on doxographical knowledge in readers' marks may relate to the humanistic practice of commonplacing, rather than suggesting that doxography was the most important thematic aspect for the readers of Gilbert's work.[47] Consequently, it necessitates a critical examination of the practice of annotation with regard to specific content, and compared to other datasets studied by a similar methodology.

The computational methods described enable researchers to identify relevant relationships and patterns within the data. These connections and patterns are often not unexpected to the researchers, who compiled the dataset, as they are usually already part of an "informed intuition." An important function of quantitative methods is to substantiate this intuition through scientific and verifiable means. This approach effectively counters potential risks of cherry-picking and ensures the reproducibility of the research, provided that the data processing and collection are documented by the researchers. While computed results must not be taken uncritically at face value they suggest a course for spot tests and case studies. Moreover, the efficacy of computational analysis is particularly useful when applied to previously unknown datasets for offering an initial semantic and meaningful overview. The method's flexibility in defining relevant features and its capacity to apply various combinations of queries against the dataset render it sufficiently versatile to enable comparable analyzes in diverse datasets.

[46] For example, interventions by librarians, such as the retrospective pagination of an entire work, must be identified and excluded from the analytical data, as they do not constitute meaningful traces of intellectual engagement.

[47] See also Ann M. Blair, Humanist Methods in Natural Philosophy: The Commonplace Book, *Journal of the History of Ideas* 53, no. 4 (1992): 541–551; Ann M. Blair, The Rise of Note-Taking in Early Modern Europe, *Intellectual History Review* 20, no. 3 (2011): 303–316.

While the creation of suitable corpora through manual research is a time-consuming process that primarily enables researchers to develop qualified intuitions that can later be statistically validated, the use of artificial intelligence and machine learning offers a novel and promising perspective. The application of computer vision in detecting readers' marks, the semantic interpretation and text mining of full texts on printed pages, and the representation of visual elements like illustrations as semantic embeddings through Vision Language Models has the potential to significantly expand the scope of data collection. It is acknowledged that a dataset created in this manner cannot possibly meet the standards of historical research that are justifiably high. It is of the utmost importance to meticulously examine and contextualize the inherent biases of the machine learning models used. Nevertheless, the tasks performed by artificial intelligence are relatively straightforward for qualified researchers to verify.

It is crucial to highlight that the analytical or heuristical methodologies presented in this paper do not use pre-trained models. This does not render them entirely neutral research tools; however, it mitigates the risk of the analytical judgments themselves being truly hallucinated or biased from training data – much discussed issues of large foundation AI models. The mindful combination of stochastic, statistical, and rule-based approaches with the generation of semantic research data through more complex machine learning methods represents a promising avenue for further research. The genesis of hermeneutic inference and historical interpretation is not a black box in this case; rather, it is modular, individually adaptable, and verifiable. Concurrently, the analytical work conducted through statistical methods is considerably less time-consuming than the manual collection of semantic data. The resources required for the manual creation of suitable datasets should be reallocated to the critical assessment, reflection, and evaluation of analytical methods and the oversight of automatically generated datasets. Research, ideally, is about interpreting, understanding, and explaining phenomena – not (primarily) about recording them.

Bibliography

Acheson, Katherine O. (ed.). *Early Modern English Marginalia*. New York: Routledge, 2019.

Bevir, Mark. *The Logic of the History of Ideas*. Cambridge (UK): Cambridge University Press, 1999. <https://doi.org/10.1017/CBO9780511490446>, accessed February 22, 2025.

Brett, Megan R. Topic Modeling: A Basic Introduction, *Journal of Digital Humanities* 2, no. 1 (2012). <https://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/>, accessed February 22, 2025.

Blair, Ann M. Humanist Methods in Natural Philosophy: The Commonplace Book. *Journal of the History of Ideas* 53, no. 4 (1992): 541–551.

Blair, Ann M. The Rise of Note-Taking in Early Modern Europe. *Intellectual History Review* 20, no. 3 (2011): 303–316.

Brockstieger, Sylvia and Rebecca Hirt (eds.). *Handschrift im Druck (ca. 1500–1800): Annotieren, Korrigieren, Weiterschreiben*. Berlin: De Gruyter, 2023. <https://doi.org/10.1515/9783111191560>, accessed February 22, 2025.

Büttner, Jochen, Julius Martinetz, Hassan El-Hajj, and Matteo Valleriani. CorDeep and the Sacrobosco Dataset: Detection of Visual Elements in Historical Documents. *Journal of Imaging* 8, no. 10 (2022): 285. <https://doi.org/10.3390/jimaging8100285>, accessed February 22, 2025.

Chang, Ku-ming (Kevin), Anthony Grafton, and Glenn Warren Most (eds.). *Impagination: Layout and Materiality of Writing and Publication: Interdisciplinary Approaches from East and West*. Berlin: De Gruyter, 2021. <https://doi.org/10.1515/9783110698756>, accessed February 22, 2025.

Church, Kenneth Ward and Patrick Hanks. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics* 16, no. 1 (1990): 22–29.

Ciula, Arianna, Øyvind Eide, Cristina Marras, and Patrick Sahle. *Modelling between Digital and Humanities: Thinking in Practice*. Cambridge (UK): Open Book Publishers, 2023. <https://books.openbookpublishers.com/10.11647/obp.0369.pdf>, accessed February 22, 2025.

Coyle, Karen. *FRBR, Before and After: A Look at Our Bibliographic Models*. Chicago: ALA Editions, 2016.

El-Hajj, Hassan, Oliver Eberle, Anika Merklein, Anna Siebold, Noga Shlomi, Jochen Büttner, Julius Martinetz, Klaus-Robert Müller, Grégoire Montavon, and Matteo Valleriani. Explainability and Transparency in the Realm of Digital Humanities: Toward a Historian XAI. *International Journal of Digital Humanities* 5, no. 2 (2023): 299–331. <https://doi.org/10.1007/s42803-023-00070-1>, accessed February 22, 2025.

Feingold, Mordechai and Andrej Svorenčík. A Preliminary Census of Copies of the First Edition of Newton's "Principia" (1687). *Annals of Science* 77, no. 3 (2020): 253–348. <https://doi.org/10.1080/00033790.2020.1808700>, accessed February 22, 2025.

Fenlon, Katrina. Modeling Digital Humanities Collections as Research Objects. In *Proceedings of the 18th Joint Conference on Digital Libraries (JCDL 2019*), 138–147. Champaign (IL): IEEE Press. <https://doi.org/10.1109/JCDL.2019.00029>, accessed February 22, 2025.

Gartner, Richard. *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*. Cham: Springer, 2016.

Gingerich, Owen. *An Annotated Census of Copernicus' "De Revolutionibus" (Nuremberg, 1543 and Basel, 1566)*. Leiden: Brill, 2002.

Golub, Koraljka and Ying-Hsang Liu (eds.). *Information and Knowledge Organisation in Digital Humanities: Global Perspectives*. London: Routledge, 2021. <https://doi.org/10.4324/9781003131816>, accessed February 22, 2025.

Grafton, Anthony. *Inky Fingers: The Making of Books in Early Modern Europe*. Cambridge (MA): Harvard University Press, 2020.

Greg, Walter Wilson. Bibliography: An Apologia. *The Library*, 4th series, 13, no. 2 (1932): 113–143. <https://doi.org/10.1093/library/s4-XIII.2.113>, accessed February 22, 2025.

Hand, David John. *Dark Data: Why What You Don't Know Matters*. Princeton: Princeton University Press, 2020.

Impett, Leo and Fabian Offert, There Is a Digital Art History. *Visual Resources* 38, no. 2: 186–209, <https://doi.org/10.1080/01973762.2024.2362466>, accessed February 22, 2025.

Jackson, Heather Joanna. *Marginalia: Readers Writing in Books*. New Haven (CT): Yale University Press, 2001.

Jacquart, Danielle and Charles S. F. Burnett (eds.). *Scientia in margine: Études sur les marginalia dans les manuscrits scientifiques du moyen âge à la renaissance*. Geneva: Droz, 2005.

Jardine, Lisa and Anthony Grafton. "Studied for Action": How Gabriel Harvey Read His Livy. *Past & Present*, no. 129 (1990): 30–78.

Jünger, Jakob and Chantal Gärtner. *Computational Methods für die Sozial- und Geisteswissenschaften*. Wiesbaden: Springer, 2023.

Kejriwal, Mayank, Craig A. Knoblock, and Pedro Szekely (eds.). *Knowledge Graphs: Fundamentals, Techniques, and Applications*. Cambridge (MA): The MIT Press, 2021.

LLaVa. *Hugging Face*. May 7, 2024. <https://huggingface.co/docs/transformers/model_doc/llava>, accessed May 7, 2024.

Loria, Mario. William Gilbert e Galileo Galilei: La terrella e le calamite del granduca. In *Saggi su Galileo Galilei*, vol. 2, Carlo Maccagni (ed.), 208–247. Florence: Giunti Barbèra, 1972.

Margócsy, Dániel, Mark Somos, and Stephen N. Joffe. *The Fabrica of Andreas Vesalius: A Worldwide Descriptive Census, Ownership, and Annotations of the 1543 and 1555 Editions*. Leiden: Brill, 2018.

Mayer, Thomas F. An Interim Report on a Census of Galileo's Sunspot Letters. *History of Science* 50, no. 2 (2012): 155–196. <https://doi.org/10.1177/007327531205000202>, accessed February 22, 2025.

Misson, James and Devani Mandira Singh. Computing Book Parts with EEBO-TCP. *Book History* 25, no. 2 (2022): 503–529. <https://doi.org/10.1353/bh.2022.0018>, accessed February 22, 2025.

Mu, Yida, Chun Dong, Kalina Bontcheva, and Xingyi Song. Large Language Models Offer an Alternative to the Traditional Approach of Topic Modelling, *arXiv:2403.16248. Preprint, arXiv*, March 26, 2024. <https://doi.org/10.48550/arXiv.2403.16248>, accessed February 22, 2025

Orgel, Stephen. *The Reader in the Book: A Study of Spaces and Traces*. Oxford: Oxford University Press, 2015.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, no. 85 (2011): 2825–2830.

Peroni, Silvio and David Shotton. FaBiO and CiTO: Ontologies for Describing Bibliographic Resources and Citations. *Journal of Web Semantics* 17 (2012): 33–43. <https://doi.org/10.1016/j.websem.2012.08.001>, accessed February 22, 2025.

Peroni, Silvio and David Shotton. The SPAR Ontologies. In *The Semantic Web – ISWC 2018*, Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, and Elena Simperl (eds.), 119–136. Cham: Springer, 2018. <https://doi.org/10.1007/978-3-030-00668-6_8>, February 22, 2025.

Piper, Andrew. *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*. Cambridge (UK): Cambridge University Press, 2020. <https://doi.org/10.1017/9781108922036>, accessed February 22, 2025.

Sander, Christoph. *Early Modern Magnetism Image Database (1500–1650)*, 2022. <https://doi.org/10.48431/res/qk19-bj96/vikus/vismag>, accessed February 22, 2025.

Sander, Christoph. *Magnes: Der Magnetstein und der Magnetismus in den Wissenschaften der Frühen Neuzeit*. Leiden: Brill, 2020. <https://doi.org/10.1163/9789004419414>, accessed February 22, 2025.

Sander, Christoph. *Magnetic Margins: A Census and Annotations Database*, October 25, 2023. <https://doi.org/10.48431/res/qk19-bj96/magmar>, accessed February 22, 2025.

Sander, Christoph. Magnetic Margins: Insights from the Digital Descriptive Census of William Gilbert's "De magnete". *Annals of Science*, forthcoming.

Sander, Christoph. Magnetism in an Aristotelian World (1550–1700). In *Alte und neue Philosophie: Aristotelismus und protestantische Gelehrsamkeit in Helmstedt und Europa (1600–1700)*, Bernd Roling, Sinem Kılıç, Benjamin Wallura, and Hartmut Beyer (eds.), 69–105. Wolfenbütteler Forschungen 175. Wiesbaden: Harrassowitz, 2023.

Sander, Christoph, Hassan El-Hajj, and Alessandro Adamou. Magnetic Margins: A Census and Reader Annotations Database. In *Digital Humanities 2023: Collaboration as Opportunity (DH2023)*. Graz: Zenodo, 2023. <https://doi.org/10.5281/zenodo.8107608>, accessed February 22, 2025.

Sangiacomo, Andrea, Raluca Tanasescu, Hugo Hogenbirk, and Silvia Donker. Recreating the Network of Early Modern Natural Philosophy: A Mono- and Multilingual Text Data Vectorization Method. *Journal of Historical Network Research* 7, no. 1 (2022): 33–85. <https://doi.org/10.25517/jhnr.v7i1.129>, accessed February 22, 2025.

Sherman, William H. *Used Books: Marking Readers in Renaissance England*. Philadelphia: University of Pennsylvania Press, 2010.

Spadini, Elena, Francesca Tomasi, and Georg Vogeler (eds.). *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, vol. 15. Norderstedt: BoD, 2021. <https://kups.ub.uni-koeln.de/54577>, accessed February 22, 2025.

Tanselle, George Thomas. *A Rationale of Textual Criticism*. Philadelphia: University of Pennsylvania Press, 1992.

Tanselle, George Thomas. *Descriptive Bibliography*. Charlottesville (VA): The Bibliographical Society of the University of Virginia, 2020.

Valleriani, Matteo, Florian Kräutli, Maryam Zamani, Alejandro Tejedor, Christoph Sander, Malte Vogl, Sabine Bertram, Gesa Funke, and Holger Kantz. The Emergence of Epistemic Communities in the Sphaera Corpus. *Journal of Historical Network Research* 3, no. 1 (2019): 50–91. <https://doi.org/10.25517/jhnr.v3i1.63>, accessed February 22, 2025.

Valleriani, Matteo, Malte Vogl, Hassan El-Hajj, and Kim Pham. The Network of Early Modern Printers and Its Impact on the Evolution of Scientific Knowledge: Automatic Detection of Awareness Relationships. *Histories* 2, no. 4 (2022): 466–503. <https://doi.org/10.3390/histories2040033>, accessed February 22, 2025.

Viola, Lorella. *The Humanities in the Digital: Beyond Critical Digital Humanities*. Heidelberg: Springer, 2023.

Westmann, Robert S. The Reception of Galileo's "Dialogue": A Partial World Census of Extant Copies. In *Novità celesti e crisi del sapere: atti del convegno internazionale di studi galileiani*, Paolo Galluzzi (ed.), 329–371. Florence: Giunti Barbèra, 1984.