Digital Studies / Le champ numérique

Estill, Laura. 2025. "Digital Text Analysis and Early Shakespeare Bibliography: Using Voyant Tools with Bad OCR." *Digital Studies/Le champ numérique* 15(1): 1–38. https://doi.org/10.16995/dscn.18897.

Open Library of Humanities

Digital Text Analysis and Early Shakespeare Bibliography: Using Voyant Tools with Bad OCR

Laura Estill, English, St Francis Xavier University, lestill@stfx.ca

Enumerative bibliographies are lists of scholarship that capture the state of a field. This article first evaluates digital texts of one such bibliography, Franz Thimm's *Shakspeariana from 1564–1864* (second edition, 1872), before applying textual analysis using Voyant Tools. The takeaways are both methodological and interpretive: we can use inaccurate online texts ("dirty OCR," that is, optical character recognition); we can fruitfully apply text analysis to printed bibliographies; and we can learn about bibliographies with Voyant Tools even if they are multilingual. This research shows how Thimm's bibliography emphasizes Shakespeare publication from major urban centres and surfaces the importance of nineteenth-century German translation and scholarship on Shakespeare, while inviting us to reconsider how we credit translators (or not) as we name them in our lists. Ultimately, experimenting with digital tools to analyze early bibliographies can help us better understand the history of our scholarship.

Les bibliographies énumératives sont des listes de travaux académiques qui reflètent l'état d'un domaine. Cet article commence par évaluer les textes numériques d'une telle bibliographie, *Shakspeariana from 1564–1864 de Franz Thimm* (deuxième édition, 1872), avant d'appliquer une analyse textuelle à l'aide de Voyant Tools. Les conclusions sont à la fois méthodologiques et interprétatives : nous pouvons utiliser des textes en ligne inexacts (« ROC sale », c'est-à-dire la reconnaissance optique de caractères défectueuse) ; nous pouvons appliquer de manière productive l'analyse textuelle aux bibliographies imprimées ; et nous pouvons explorer les bibliographies avec Voyant Tools même lorsqu'elles sont multilingues. Cette recherche montre comment la bibliographie de Thimm met en avant les publications sur Shakespeare provenant des grands centres urbains et révèle l'importance de la traduction et des études allemandes du XIXe siècle sur Shakespeare, tout en nous invitant à reconsidérer la manière dont nous reconnaissons (ou non) les traducteurs en les nommant dans nos listes. En fin de compte, expérimenter avec des outils numériques pour analyser les premières bibliographies peut nous aider à mieux comprendre l'histoire de notre production académique.

This article started with a simple question: what can we learn about a nineteenth-century bibliography that lists publications about Shakespeare using an "out-of-the-box" text analysis tool, that is, Voyant Tools? As I started undertaking this research, additional questions emerged: can this be used with pre-existing digital texts, or would I have to transcribe the volume by hand or run my own optical character recognition model? Can we fruitfully apply text analysis to a book that contains multiple languages? Is there a value in applying text analysis to a bibliography, that is, a book that was not meant to be read from front to back?

As this article shows, you can use pre-existing OCR (optical character recognition) from large-scale digitization projects (though you have to evaluate your OCR and nuance your claims accordingly). Multilingual text analysis is also possible if you know your text. And yes, applying text analysis to enumerative bibliographies is a useful way to understand not only the contents of a bibliography but also, by proxy, to get a sense of the field it covers. Indeed, using text analysis to consider the contents of a bibliography is a particularly apt method of engagement because we skim lists rather than "read" them (Smith 1991). Geoffrey Rockwell and Stéfan Sinclair designed Voyant to support consultative reading (Rockwell and Sinclair 2016, 48), that is, moving beyond linear reading and instead seeking information. Both print bibliographies and Voyant are tools to facilitate research, but, as with all tools, researchers need to understand their uses to make effective claims.

This article opens by introducing the value of studying enumerative bibliographies as a snapshot of a scholarly field, then introduces the bibliography explored in this article, Franz Thimm's *Shakspeariana from 1564–1864* (London, 1865; second edition, 1872). I compare the different OCR'd versions of Thimm's *Shakspeariana* on the Internet Archive in order to choose an existing digital text to analyze. I then model an iterative mode of inquiry with Voyant Tools following the precepts outlined in Rockwell and Sinclair's *Hermeneutica*. As Rockwell and Sinclair emphasize, it is important to show the steps of digital textual analysis: "If you hide the technique, you lose the logical force of an argument, in addition to losing any reader who might be interested in the technique itself" (Rockwell and Sinclair 2016, 35). As the final section of this article tracts, even with imperfect digital texts, when it comes to Thimm's *Shakspeariana*, we can see the outlines of a scholarly discipline appear using text analysis on an early bibliography.

Why study bibliographies?

By their nature, bibliographies offer a snapshot of studies about their time. My focus here is enumerative bibliographies (comprehensive lists) and not, say, descriptive or analytic bibliography (the study of how old books were created and assembled).

Enumerative bibliographies list and often categorize materials; before digital projects, these printed books were an important tool for finding scholarship on a particular subject. Unlike a works cited list, which could be considered a small enumerative bibliography, the bibliographies that I analyze here aim for comprehension, that is, they attempt to list every relevant item in their declared scope. Although no bibliography is complete, the materials they capture offer a glimpse into a moment of publication or scholarly history. By text mining a bibliography about Shakespeare publications, we learn *where* Shakespeare scholarship was being published, *who* was involved in these publications, while also reflecting on *what* claims we can make. As Thimm's *Shakspeariana* reveals, Shakespeare studies in the nineteenth century was increasingly global.

This project came to be as I was working with Heidi Craig, Kris L. May, and Dorothy Todd to create a history of Shakespearean bibliography (Craig et al. 2026). With an emphasis on how we list and find scholarship about Shakespeare, in *Collaboration, Technologies, and the History of Shakespearean Bibliography*, we argue that understanding bibliographies is foundational to how we research. The history of Shakespeare bibliography both supports and reflects the history of Shakespeare scholarship. Centuries before the digital turn, scholars lamented the glut of Shakespeare scholarship and the difficulty of gathering and assessing it (Craig et al. 2026). Digital bibliography applied computational tools to the gathering and indexing of scholarship and related materials, although it had to contend with an explosion in research, partly due to the same digital affordances. Rather than considering digital bibliographies, in this article, I apply digital methodologies to learn about a nineteenth-century bibliography about Shakespeare.

Nineteenth-century Shakespeare scholarship was published in many countries across Europe and North America. The bibliographies that attempted to catalogue this scholarship, as well as translations and editions of Shakespeare's plays, were similarly published in multiple countries. In England, for instance, John Wilson's *Shaksperiana* (London, 1827), James Orchard Halliwell-Phillipps's *Shakesperiana* (London, 1841), and Henry J. Bohn's *The Biography and Bibliography of Shakespeare* (London, 1863) are but a few examples of Shakespeare bibliographies (see Craig et al. 2026, for more examples; for additional examples of nineteenth-century bibliographies, with links to those that are available open access, see Estill, forthcoming). Beyond England, Jurriaan Moulin published *Omtrekken eener algemeene litteratuur over William Shakespeare en deszelfs werken* (Outlines of a General Literature on William Shakespeare and His Works) ([Kampen, the Netherlands], 1845); Max Koch included over sixty pages of bibliography in his tome *ШЕКСПИРЪ* (Shakespeare) (Moscow, 1888); and Morgan

Appleton's *Digest Shakespeareanæ* (New York, 1886–1887) was published for the New York Shakespeare Society. Some bibliographers emphasized the contributions to Shakespeare scholarship from their region, such as Karl Knortz's *An American Shakespeare Bibliography* (Boston, 1876); Thomas James I. Arnold's *Shakespeare-Bibliography in the Netherlands | Shakespeare in de Nederlandsche letterkunde en op het Nederlandsch tooneel. Bibliographisch overzicht* (The Hague, 1879); and Ludwig Unflad's *Die Shakespeare-literatur in Deutschland* (Shakespeare Literature in Germany) (Munich, 1880). Shakespeare bibliography was undertaken in multiple countries and multiple languages; many of these bibliographies tried to capture the multilingual scholarship of the day. Thanks to digital archives such as HathiTrust, the Internet Archive, and Google Books, we now have access to facsimiles and full texts of many of these enumerative bibliographies that document publications about Shakespeare.

In this article, I focus on Thimm's *Shakspeariana from 1564–1864* because it exemplifies nineteenth-century Shakespeare bibliography, with its emphasis on English and German contributions and its attempts to include more global content. Thimm aimed to provide a comprehensive bibliography: in his introduction to the first edition, he boasted that his volume contained more materials ("over 600 more [entries]") than previous bibliographies (Thimm 1865, v). In his second edition, he included the perennial plea of bibliographers: "Bibliographers are aware that it is almost impossible to collect every known book on Shakspeare, I therefore appeal to the kindness of those who may use my book and find any thing missing, to inform me of any full titles, omissions or errors, which information will be received with thanks, and duly incorporated with future editions" (Thimm 1872, [iii]). Despite his call for more information, Thimm did not complete a third edition or further supplements. I focus on the second edition of Thimm's bibliography because of its expanded global content (adding eight pages of international content and ten pages of publications from 1865–1872).

Thimm's work announces its multilingual audience and contents by providing two title pages with slightly varying content (see **Figure 1**). The German title page for the second edition announces that it covers material from 1564–1871; the English title page emphasizes the original purview, 1564–1864, while adding in smaller font below that the second edition "contain[s] the literature from 1864–1871" (Thimm 1872). The German title page announces that this bibliography covers "aller Länder der Welt [all the countries of the world]" (Thimm 1872), whereas the English title page highlights that it covers literature from "England, Germany, and France" in a large typeface, followed by a much smaller "and other European countries" (Thimm 1872). The first edition of Thimm's bibliography covered only England, Germany, and

France. In the second edition, while the emphasis remained on England (47 pages not including general introduction), Germany (49 pages), and to a lesser degree, France (13 pages), Thimm added coverage of a number of additional languages: Italian, Spanish, Portuguese, Danish, Swedish, Dutch, Friesic (Frisian), Bohemian (Czech), Hungarian, Walachian (Romanian), Modern Greek, Polish, Russian, and Bengalee (Bengali/Bangla) (in Thimm's order).

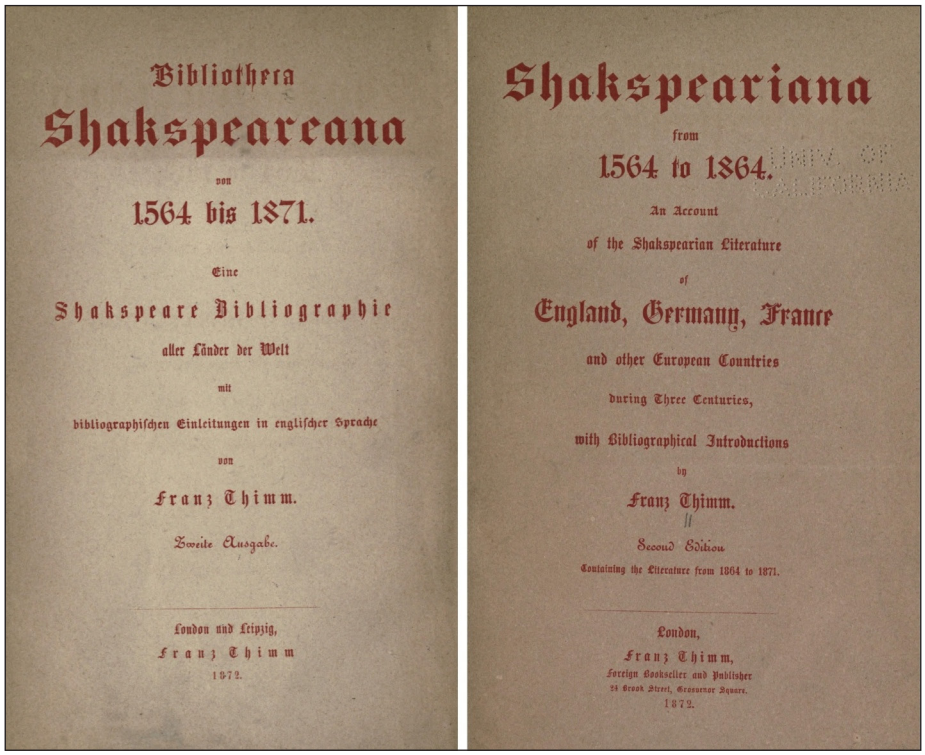
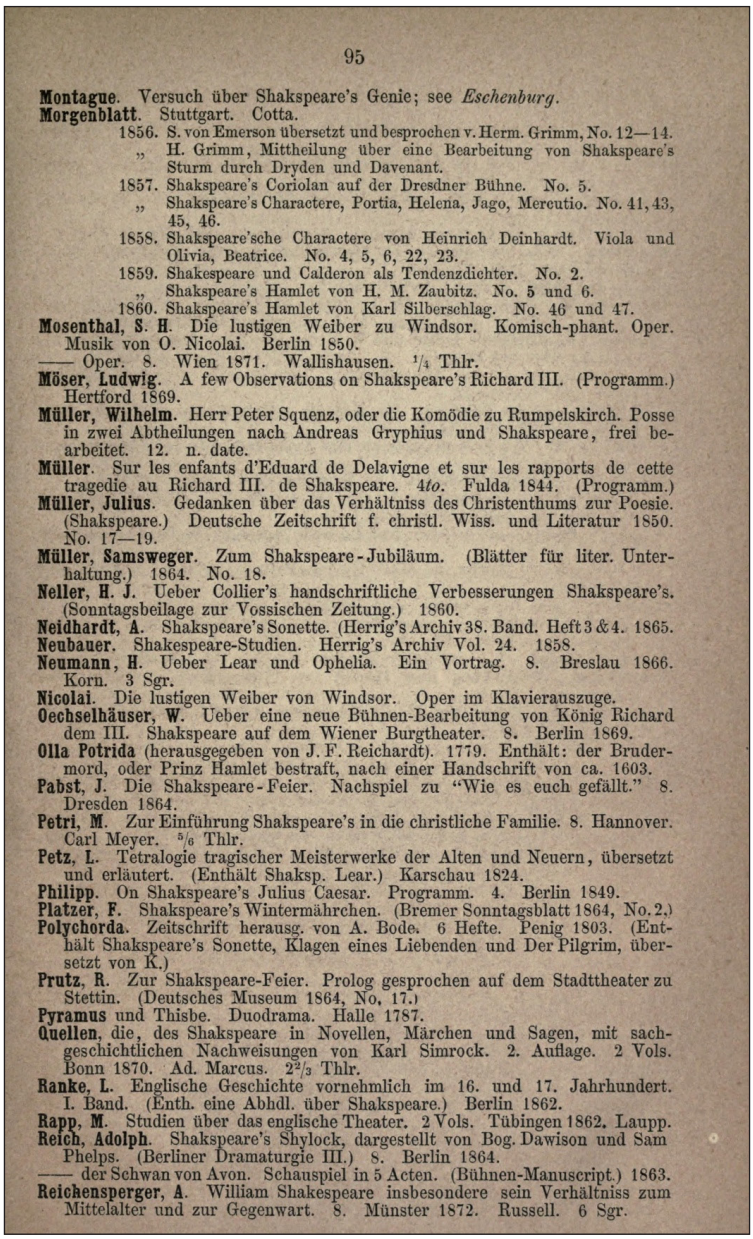


Figure 1: Facing-page title pages from the second edition of Thimm's *Shakspeariana from 1564–1864* (1872), digitized by the Internet Archive, https://archive.org/details/shakspearianafro00thimrich.

Most nineteenth-century bibliographies offer a brief introduction and then provide a list of scholarship, often subdivided by category (sometimes by their place of origin, other times by the topic, for instance, the play)—Thimm's bibliography is no exception. (In Craig et al. 2026, we touch briefly on the challenges of taxonomizing Shakespeare scholarship, editions, and translations, though there is much more work to be done in this area. In short: the way we categorize research to make it findable offers a snapshot of the value we place on it and offers one way to understand a field of study.) **Figure 2** and **Figure 3** offer an example of the kinds of material contained in these editions and their multilingual contents. The bulk of Thimm's volume is

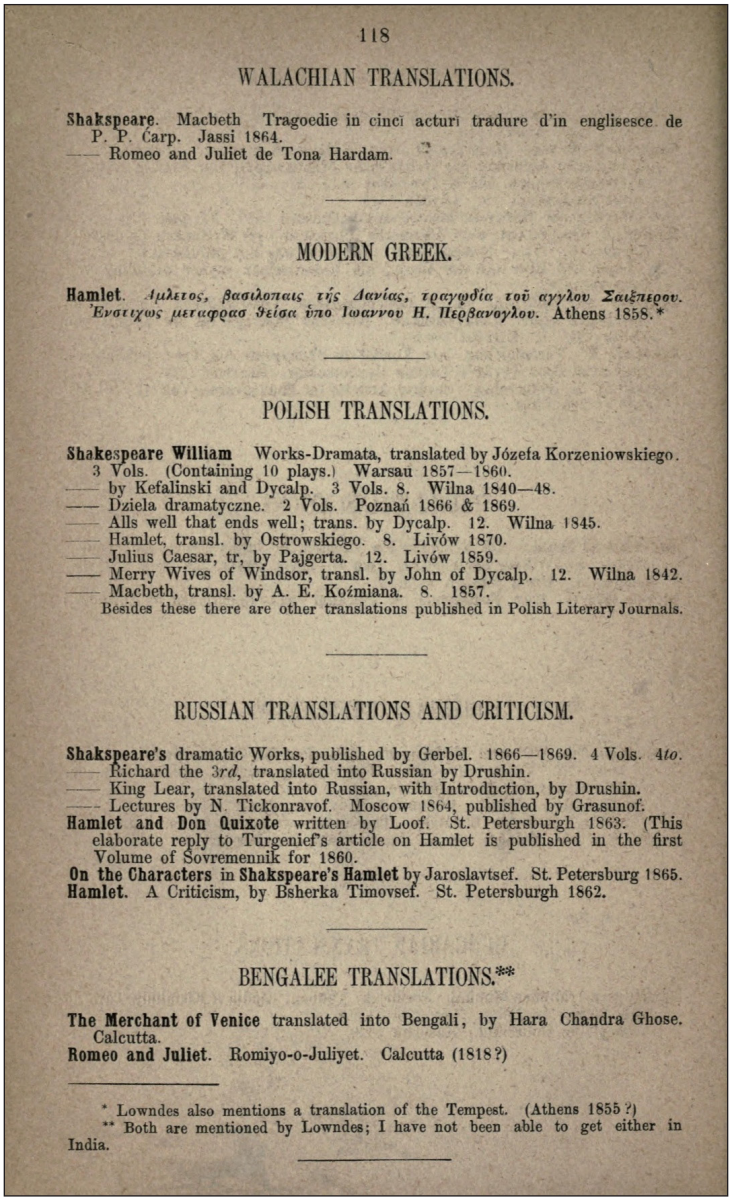
typical of enumerative Shakespeare bibliographies. Thimm's 1872 volume is divided geographically, offering three brief introductions to the three main sections about Shakespeare in England, Germany, and France. The bulk of the volume is comprised of lists of editions Shakespeare's work (in English and in translation) and lists of publications about Shakespeare's life or work ("Commentaries, Essays, and Plates").



95

Montague. Versuch über Shakspeare's Genie; see *Eschenburg.*
Morgenblatt. Stuttgart. Cotta.
1856. S. von Emerson übersetzt und besprochen v. Herm. Grimm, No. 12—14.
„ H. Grimm, Mittheilung über eine Bearbeitung von Shakspeare's Sturm durch Dryden und Davenant.
1857. Shakspeare's Coriolan auf der Dresdner Bühne. No. 5.
„ Shakspeare's Charactere, Portia, Helena, Jago, Mercutio. No. 41, 43, 45, 46.
1858. Shakspeare'sche Charactere von Heinrich Deinhardt. Viola und Olivia, Beatrice. No. 4, 5, 6, 22, 23.
1859. Shakespeare und Calderon als Tendenzdichter. No. 2.
„ Shakspeare's Hamlet von H. M. Zaubitz. No. 5 und 6.
1860. Shakspeare's Hamlet von Karl Silberschlag. No. 46 und 47.
Mosenthal, S. H. Die lustigen Weiber zu Windsor. Komisch-phant. Oper. Musik von O. Nicolai. Berlin 1850.
—— Oper. 8. Wien 1871. Wallishausen. 1/4 Thlr.
Möser, Ludwig. A few Observations on Shakspeare's Richard III. (Programm.) Hertford 1869.
Müller, Wilhelm. Herr Peter Squenz, oder die Komödie zu Rumpelskirch. Posse in zwei Abtheilungen nach Andreas Gryphius und Shakspeare, frei bearbeitet. 12. n. date.
Müller. Sur les enfants d'Eduard de Delavigne et sur les rapports de cette tragedie au Richard III. de Shakspeare. *4to.* Fulda 1844. (Programm.)
Müller, Julius. Gedanken über das Verhältniss des Christenthums zur Poesie. (Shakspeare.) Deutsche Zeitschrift f. christl. Wiss. und Literatur 1850. No. 17—19.
Müller, Samsweger. Zum Shakspeare-Jubiläum. (Blätter für liter. Unterhaltung.) 1864. No. 18.
Neller, H. J. Ueber Collier's handschriftliche Verbesserungen Shakspeare's. (Sonntagsbeilage zur Vossischen Zeitung.) 1860.
Neidhardt, A. Shakspeare's Sonette. (Herrig's Archiv 38. Band. Heft 3 & 4. 1865.
Neubauer. Shakespeare-Studien. Herrig's Archiv Vol. 24. 1858.
Neumann, H. Ueber Lear und Ophelia. Ein Vortrag. 8. Breslau 1866. Korn. 3 Sgr.
Nicolai. Die lustigen Weiber von Windsor. Oper im Klavierauszuge.
Oechselhäuser, W. Ueber eine neue Bühnen-Bearbeitung von König Richard dem III. Shakspeare auf dem Wiener Burgtheater. 8. Berlin 1869.
Olla Potrida (herausgegeben von J. F. Reichardt). 1779. Enthält: der Brudermord, oder Prinz Hamlet bestraft, nach einer Handschrift von ca. 1603.
Pabst, J. Die Shakspeare-Feier. Nachspiel zu "Wie es euch gefällt." 8. Dresden 1864.
Petri, M. Zur Einführung Shakspeare's in die christliche Familie. 8. Hannover. Carl Meyer. 5/6 Thlr.
Petz, L. Tetralogie tragischer Meisterwerke der Alten und Neuern, übersetzt und erläutert. (Enthält Shaksp. Lear.) Karschau 1824.
Philipp. On Shakspeare's Julius Caesar. Programm. 4. Berlin 1849.
Platzer, F. Shakspeare's Wintermährchen. (Bremer Sonntagsblatt 1864, No. 2.)
Polychorda. Zeitschrift herausg. von A. Bode. 6 Hefte. Penig 1803. (Enthält Shakspeare's Sonette, Klagen eines Liebenden und Der Pilgrim, übersetzt von K.)
Prutz, R. Zur Shakspeare-Feier. Prolog gesprochen auf dem Stadttheater zu Stettin. (Deutsches Museum 1864, No. 17.)
Pyramus und Thisbe. Duodrama. Halle 1787.
Quellen, die, des Shakspeare in Novellen, Märchen und Sagen, mit sachgeschichtlichen Nachweisungen von Karl Simrock. 2. Auflage. 2 Vols. Bonn 1870. Ad. Marcus. 2 2/3 Thlr.
Ranke, L. Englische Geschichte vornehmlich im 16. und 17. Jahrhundert. I. Band. (Enth. eine Abhdl. über Shakspeare.) Berlin 1862.
Rapp, M. Studien über das englische Theater. 2 Vols. Tübingen 1862. Laupp.
Reich, Adolph. Shakspeare's Shylock, dargestellt von Bog. Dawison und Sam Phelps. (Berliner Dramaturgie III.) 8. Berlin 1864.
—— der Schwan von Avon. Schauspiel in 5 Acten. (Bühnen-Manuscript.) 1863.
Reichensperger, A. William Shakespeare insbesondere sein Verhältniss zum Mittelalter und zur Gegenwart. 8. Münster 1872. Russell. 6 Sgr.

Figure 2: Thimm's *Shakspeariana from 1564–1864* (1872), p. 95, from the section on "German Shakspeariana." Digitized by the Internet Archive, https://archive.org/details/shakspearianafro00thimrich.



118

WALACHIAN TRANSLATIONS.

Shakspeare. Macbeth Tragoedie in cinci acturi tradure d'in englisesce de P. P. Carp. Jassi 1864.
—— Romeo and Juliet de Tona Hardam.

MODERN GREEK.

Hamlet. *Ἁμλετος, βασιλοπαις τής Δανίας, τραγῳδία τοῦ ἀγγλου Σαιξπειρου. Ἐνατιχως μεταφρασ θεῖσα ὑπο Ιωαννου Η. Περβανογλου.* Athens 1858.*

POLISH TRANSLATIONS.

Shakespeare William Works-Dramata, translated by Józefa Korzeniowskiego. 3 Vols. (Containing 10 plays.) Warsau 1857—1860.
—— by Kefalinski and Dycalp. 3 Vols. 8. Wilna 1840—48.
—— Dziela dramatyczne. 2 Vols. Poznań 1866 & 1869.
—— Alls well that ends well; trans. by Dycalp. 12. Wilna 1845.
—— Hamlet, transl. by Ostrowskiego. 8. Livów 1870.
—— Julius Caesar, tr, by Pajgerta. 12. Livów 1859.
—— Merry Wives of Windsor, transl. by John of Dycalp. 12. Wilna 1842.
—— Macbeth, transl. by A. E. Koźmiana. 8. 1857.
Besides these there are other translations published in Polish Literary Journals.

RUSSIAN TRANSLATIONS AND CRITICISM.

Shakspeare's dramatic Works, published by Gerbel. 1866—1869. 4 Vols. 4*to*.
—— Richard the 3*rd*, translated into Russian by Drushin.
—— King Lear, translated into Russian, with Introduction, by Drushin.
—— Lectures by N. Tickonravof. Moscow 1864, published by Grasunof.
Hamlet and Don Quixote written by Loof. St. Petersburgh 1863. (This elaborate reply to Turgenief's article on Hamlet is published in the first Volume of Sovremennik for 1860.
On the Characters in **Shakspeare's Hamlet** by Jaroslavtsef. St. Petersburg 1865.
Hamlet. A Criticism, by Bsherka Timovsef. St. Petersburgh 1862.

BENGALEE TRANSLATIONS.**

The Merchant of Venice translated into Bengali, by Hara Chandra Ghose. Calcutta.
Romeo and Juliet. Romiyo-o-Juliyet. Calcutta (1818?)

* Lowndes also mentions a translation of the Tempest. (Athens 1855?)
** Both are mentioned by Lowndes; I have not been able to get either in India.

Figure 3: Thimm's *Shakspeariana from 1564–1864* (1872), p. 118, offering an example of Thimm's international coverage. Digitized by the Internet Archive, https://archive.org/details/shakspearianafro00thimrich.

Today, of course, researchers rely on digital bibliographies, such as the *World Shakespeare Bibliography* and the *MLA International Bibliography* to find scholarship about Shakespeare. With today's digital bibliographies, we can more easily undertake quantitative analysis (see, for example, Estill, Klyve, and Bridal 2015) and we have different expectations for ease of access to this information, including search, browse, sort, and export (Craig and Estill 2022). Early bibliographies warrant our

attention because they offer an overview of a moment of scholarship; they were intended to help researchers and collectors know about the state of the scholarly conversation to help them find the materials therein. Thimm, for instance, adds a note "To Shakspearian Collectors"/"An Shakespeare-Forscher" in his second edition, in facing page English and German (Thimm 1872, 119–120). Understanding past scholarship is crucial to our scholarship today, just as understanding past research practices—which, of course, relied on bibliographies—reveals how that scholarship was created and places it in contexts so that we can effectively build on and interrogate these longstanding traditions.

Online texts of Thimm's *Shakspeariana* and optical character recognition

Gathering and analyzing digital texts of Thimm's bibliography are important because they set the stage for the interpretive caveats to come, while also offering a snapshot of the kinds of digital texts we regularly use in our research and their all-too-frequent shortcomings.

Before attempting text analysis using Voyant Tools, we need to consider the digital text we are analyzing. Though a PDF downloaded from the Internet Archive or Google Books can be uploaded into Voyant and will produce results, we need to determine what text Voyant will "see." That is to say, while the page images will be, more often than not, perfectly readable to humans, they might not have been processed adequately for a machine. This process, called optical character recognition (OCR) or automated text recognition (ATR), is rapidly improving and will continue to improve. However, for decades, the OCR that has been mass-produced has not been perfect (Gupta et al. 2015; Christy et al. 2017; Hill and Hengchen 2019); mass digitization efforts have resulted in reams (or, should we say, gigabytes) of digital texts that appear online with pre-existing yet flawed optical character recognition. Historical documents offer particular challenges for optical character recognition.

Humanists receive mixed messages about what we can do with these already-digitized texts. One the one hand, Jørgen Burchardt asks, "Can we trust searches performed in archives generated by optical character recognition (OCR)?" (Burchardt 2023, 31), to which he offers the answer, "This article presents enough examples of flawed results produced by the technology to suggest that the short answer to the question is a resounding no" (Burchardt 2023, 31–32). On the other hand, Mark J. Hill and Simon Hengchen show that for some natural language processing, "the impact of OCR … is perhaps less problematic than one may initially guess" (Hill and Hengchen 2019, 840).

For Ryan Cordell, OCR, and particularly "dirty OCR," creates "a new edition of the text" (Cordell 2017, 196) and offers the chance to undertake "speculative bibliography" (Cordell 2020). Cordell asserts that we can apply digital (analytic) bibliography to understand the history, provenance, and textuality of digital texts (Cordell 2017). Cordell points out that "Errorful OCR influences our research in ways by now well expounded by scholars, inhibiting, for instance, comprehensive search" (Cordell 2017, 195), but goes on to assert that "critiques that both begin and end with the imperfections of OCR foreshorten the bibliographic imagination" (Cordell 2017, 196). Cordell calls for scholars to better understand digital archives; here, I compare different OCR-generated texts across multiple digital archives.

When we turn to Thimm's *Shakspeariana from 1564–1864*, there are multiple texts online, and each has different optical character recognition. Here I use "text" and "document" following G. Thomas Tanselle's usage (Tanselle 1989). The Internet Archive offers five different texts of Thimm's bibliography: two of the first edition and three of the second edition (see **Table 1**). These represent four different physical documents: the final two versions (D and E) are different digitizations of the same volume held at the University of California Berkeley's Shields Library. Although D and E represent the same physical object (document), here, I treat them separately because they are two different digital objects, offering (following Cordell 2017) two different digital texts. **Figure 9** and **Figure 10** emphasize how different these digital texts are: for instance, where the facsimile clearly reads "Shakspearian Literature," the OCR of Text C reads "Sljakspearian Ctterature" and the OCR of Text D reads "5l)ttk0pearion f itctttturc."

Table 2 traces seven additional full-text downloads of Thimm's *Shakspeariana*, many of which were undertaken by Google Books (for more on Google Books, the Internet Archive, and mass digitization projects by universities and cultural heritage entities, see Barnett 2020). There are, of course, multiple additional copies listed on Google Books, including many that are not available to download and print-on-demand reprints (see Trettien 2013 for more on print-on-demand editions of out-of-copyright texts and how they clog our search results and affect digital textuality). Google Books lists a phantom third edition with a date of 1890 (Google Books 2025), drawing on a WorldCat entry based on a library holding cataloguing notes in a copy of the second edition by his son, Carl Albert, for an unrealized third edition (WorldCat 2025). WorldCat shows many other libraries with a copy of Thimm's *Shakspeariana* in both the first and second edition, as well as later print-on-demand and ebook versions (see Trettien 2013); when many of those libraries link to digital versions of the text, they link to one of the available resources on HathiTrust. The Stanford University Online

	Year of Publication	Internet Archive URL (the Internet Archive identifier appears after the last slash)	Digitized by	Notes	Repository, call number, and link to catalogue
A	1865	https://archive.org/details/shakespeariana-fr00thimuoft	Robarts Library – University of Toronto	There is a typo in the volume's title in the metadata in the library catalogue and on the Internet Archive: "Shakespeariana" for "Shakspeariana," which means this won't appear in a title search.	University of Toronto, Robarts Library Stacks Call number: Z8811.T5 https://librarysearch.library.utoronto.ca/permalink/01UTORONTO_INST/blpd0s/alma991106763672806196
B	1865	https://archive.org/details/cu31924029648601	Cornell University Library	Also appears in HathiTrust: https://hdl.handle.net/2027/coo1.ark:/13960/t5v70124k Cornell has two copies of Thimm's *Shakspeariana*, only one (the Goldwin Smith copy) is digitized.	Cornell University Library Call number: Z8811.T44 https://catalog.library.cornell.edu/catalog/4687748
C	1872	https://archive.org/details/shakspeariana-fr00thimgoog	Google Books (Google ID: 3YYLAAAAIAAJ)	This volume's relation to Stanford is not listed in the Internet Archive metadata but is clear from the bookstamps that appear in the digitized volume.	Stanford University Library Call number: 822.33.AT44 https://searchworks.stanford.edu/view/2253668
D	1872	https://archive.org/details/shakspeariana-fro00thimrich	The Internet Archive	This version also appears in HathiTrust: https://hdl.handle.net/2027/uc2.ark:/13960/t7fq9st8j	University of California Davis Library, Shields Library General Collection Call number: Z8811.T442 1872 https://search.library.ucdavis.edu/permalink/01UCD_INST/1birqoj/alma990017017470403126
E	1872	https://archive.org/details/shakspeariana-fr01thimgoog	Google Books (Google ID: axU_AAAAIAAJ)	This version also appears in HathiTrust: https://hdl.handle.net/2027/uc1.$b152230	University of California Davis Library, Shields Library General Collection Call number: Z8811.T442 1872 https://search.library.ucdavis.edu/permalink/01UCD_INST/1birqoj/alma990017017470403126

Table 1: Digital versions of Thimm's *Shakspeariana* on the Internet Archive.

	Year of Publication	URL	Digitized by	Notes	Repository, call number, and link to catalogue
F	1865	Google Books, https://books.google.ca/books?id=m4ZTAAAAcAAJ	Google Books	Metadata error: listed as 1863 publication because of unclear facsimile	Dutch Koninklijke Bibliotheek (National Library of the Netherlands) Call number: 9107 G 14 Link to physical text in catalogue: https://webggc.oclc.org/cbs/DB=2.37/XMLPRS=Y/PPN?PPN=197320783 Link to digital text in catalogue: https://webggc.oclc.org/cbs/DB=2.37/XMLPRS=Y/PPN?PPN=354475541
G	1865	Hathi Trust, https://hdl.handle.net/2027/mdp.39015082234496	University of Michigan	For more on HathiTrust's digitization partners, see https://www.hathitrust.org/member-libraries/contribute-content/.	University of Michigan Library, Special Collections Shakespeare Collection Call number: PR 2885.A1 T44 1865 https://search.lib.umich.edu/catalog/record/990013673240106381
H	1865	Hathi Trust, https://hdl.handle.net/2027/nyp.33433069267460	Google Books (Google vID: NYPL:33433069267460)		New York Public Library Call number: *NCI (Thimm, F. Shakspeariana from 1564 to 1864. 1865) https://www.nypl.org/research/research-catalog/bib/b13490090
I	1865	Google Books, https://books.google.ca/books?id=ixO7CFRkU7MC	Google Books	Due to the cyberattacks on the British Library in October 2023, this item was not available to request (as of June 2025).	British Library Call number: Digital Store 2785.cm.3 https://bll01.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma-990197558640109251&context=L&v-id=44BL_INST:BLL01&lang=en

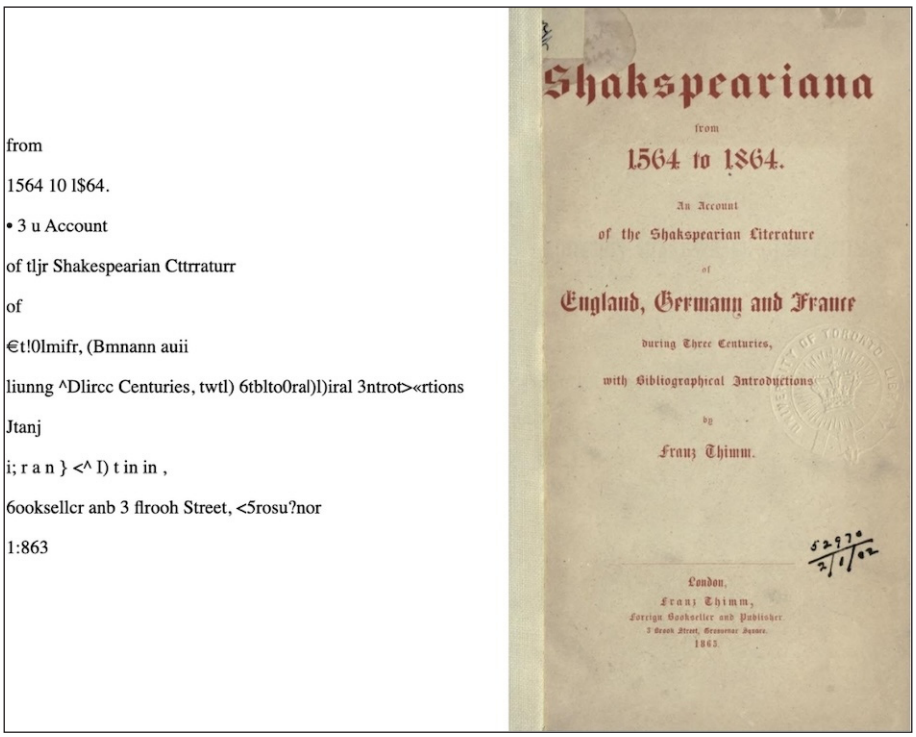
(Contd.)

	Year of Publication	URL	Digitized by	Notes	Repository, call number, and link to catalogue
J	1865	Austrian National Library (ONB), http://data.onb.ac.at/rep/10452D9B	Google Books (Google ID: id=eH_5wS-iJCUC)	For more on the partnership between Austrian Books Online (the Austrian National Library) and Google Books, see https://www.onb.ac.at/en/digital-offers/austrian-books-online. Google Books undertakes the digitization and OCR for this partnership.	Österreichische Nationalbibliothek (Austrian National Library) Call number: 161985-B https://search.onb.ac.at/permalink/f/sb7jht/ONB_alma21325497820003338
K	1872	Austrian National Library (ONB), http://data.onb.ac.at/rep/1062C8A5	Google Books (Google ID: kmDNZFMpY-FwC)		Österreichische Nationalbibliothek (Austrian National Library) Call number: 99654-B https://search.onb.ac.at/permalink/f/sb7jht/ONB_alma21325497660003338
L	1872	Google Books, https://books.google.ca/books?id=P91yuT4cKTEC	Google Books		British Library Call number W84/1428 https://bll01.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma-990115179220109251&context=L&v-id=44BL_INST:BLL01&lang=en

Table 2: Additional digital versions of Thimm's *Shakspeariana*.

catalogue, for instance, links to the digitized copy from the University of California Libraries on HathiTrust (D) as well as to the digitized copy from their own library on Google Books (C). These digital texts are created from physical copies yet have their own digital textuality.

Turning to the digitized texts from the Internet Archive (**Table 1**), we can see that across the board, the OCR is unusable (see **Figure 4, Figure 5, Figure 6, Figure 7**, and **Figure 8**; these OCR shown here are taken from the "text" option of the Internet Archive). Thimm's title page was printed using a typeface from the Fraktur family, also known as blackletter or gothic. As Jens Bjerring-Hansen and colleagues note, these typefaces can "pose technical and methodological challenges in terms of processing the text from printed page to digital corpus" (Bjerring-Hansen et al. 2022, 177). Even the page images provided in this article are not too clear; the quality stems from the fact that these are the images that appear online, not new images taken for publication. Many of our digital texts are the result of scanned microfilms (Cordell 2017), as early modernists who turn to *Early English Books Online* or *Eighteenth-Century Collections Online* are all too aware. (For more on digitizing from microfilm, see Martin 2007 and Verheusen, van Dormolen, and Wilms 2011, among others; and on the labour of microfilming and digitizing, see Quiring 2024).



from

1564 10 l$64.

• 3 u Account

of tljr Shakespearian Cttrraturr

of

€t!0lmifr, (Bmnann auii

liunng ^Dlircc Centuries, twtl) 6tblto0ral)l)iral 3ntrot><rtions

Jtanj

i; r a n } <^ I) t in in ,

6ooksellcr anb 3 flrooh Street, <5rosu?nor

1:863

Shakspeariana

from

1564 to 1864.

An Account

of the Shakspearian Literature

of

England, Germany and France

during Three Centuries,

with Bibliographical Introductions

by

Franz Thimm.

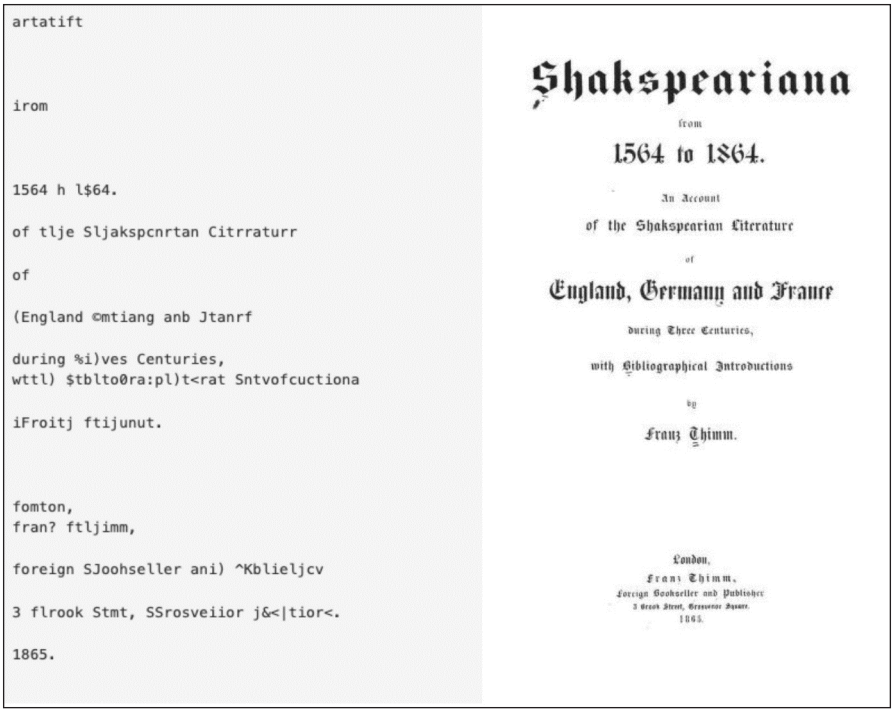
London.

Franz Thimm,

Foreign Bookseller and Publisher.

1865.

Figure 4: OCR and image of English title page from A.



```
artatift


irom


1564 h l$64.

of tlje Sljakspcnrtan Citrraturr

of

(England ©mtiang anb Jtanrf

during %i)ves Centuries,
wttl) $tblto0ra:pl)t<rat Sntvofcuctiona

iFroitj ftijunut.


fomton,
fran? ftljimm,

foreign SJoohseller ani) ^Kblieljcv

3 flrook Stmt, SSrosveiior j&<|tior<.

1865.
```

Shakspeariana

from

1564 to 1864.

An Account

of the Shakspearian Literature

of

England, Germany and France

during Three Centuries,

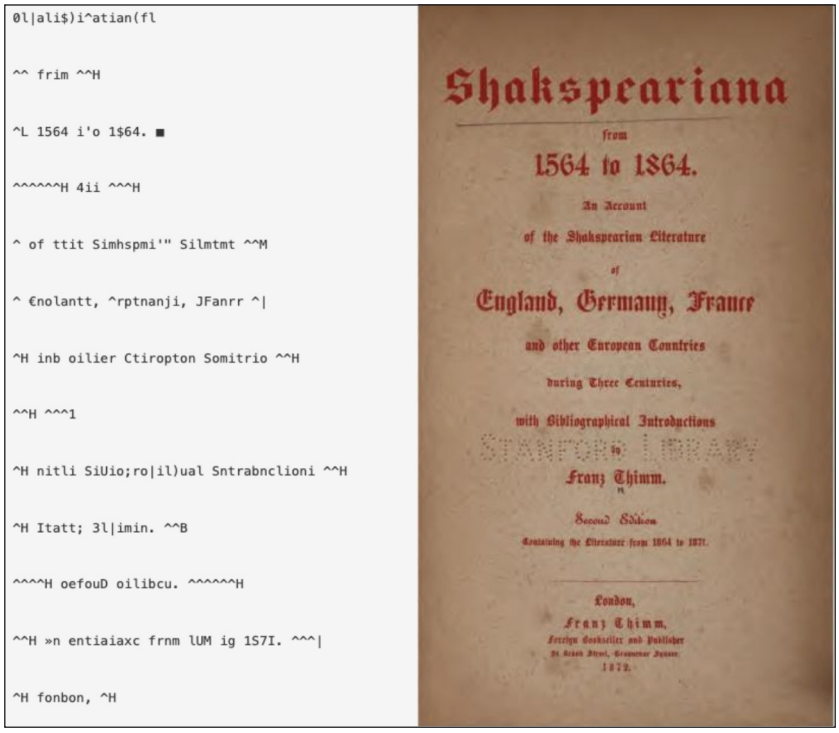
with Bibliographical Introductions

by

Franz Thimm.

London,
Franz Thimm,
Foreign Bookseller and Publisher
3 Brook Street, Grosvenor Square.
1865.

Figure 5: OCR and image of English title page from B.



```
0l|ali$)i^atian(fl


^^ frim ^^H


^L 1564 i'o 1$64. ■


^^^^^^H 4ii ^^^H


^ of ttit Simhspmi'" Silmtmt ^^M


^ €nolantt, ^rptnanji, JFanrr ^|


^H inb oilier Ctiropton Somitrio ^^H


^^H ^^^1


^H nitli SiUio;ro|il)ual Sntrabnclioni ^^H


^H Itatt; 3l|imin. ^^B


^^^^H oefouD oilibcu. ^^^^^^H


^^H »n entiaiaxc frnm lUM ig 1S7I. ^^^|


^H fonbon, ^H
```

Shakspeariana

from

1564 to 1864.

An Account

of the Shakspearian Literature

of

England, Germany, France

and other European Countries

during Three Centuries,

with Bibliographical Introductions

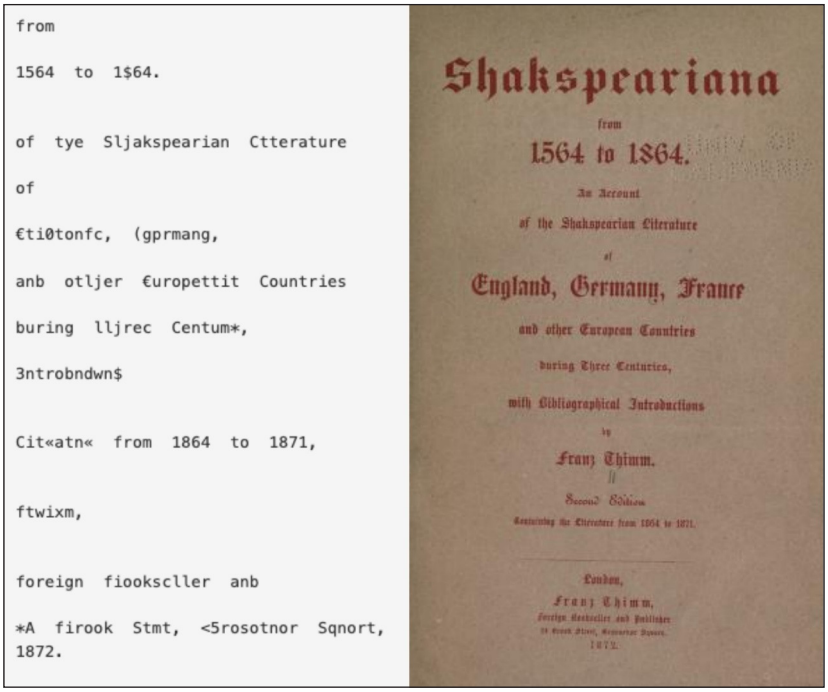
by

Franz Thimm.

Second Edition

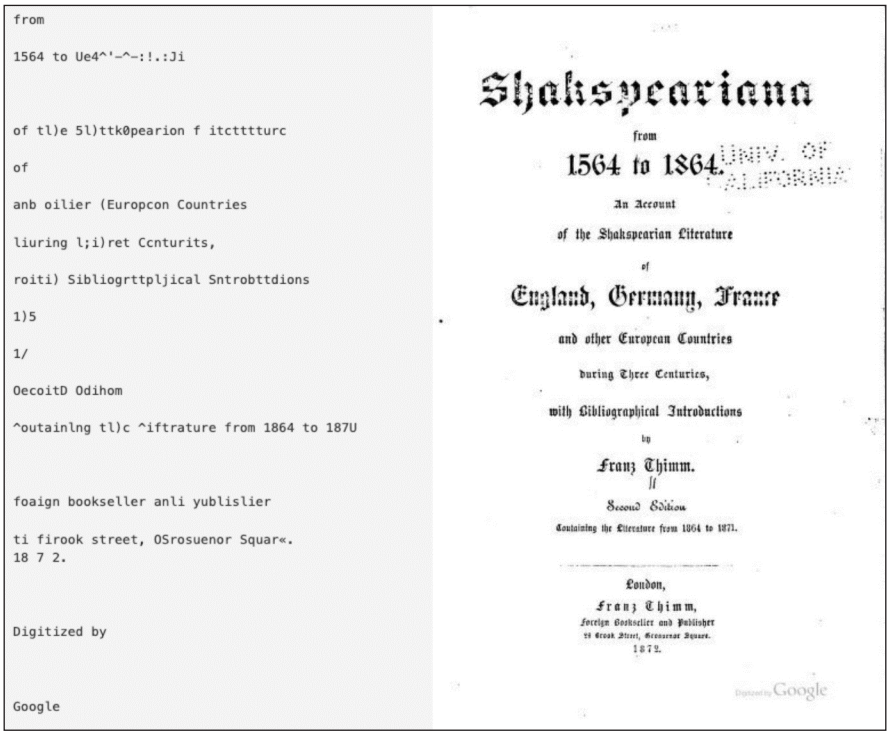
London,
Franz Thimm,
Foreign Bookseller and Publisher
3a Brook Street, Grosvenor Square.
1872.

Figure 6: OCR and image of English title page from C.



```
from

1564  to  1$64.


of  tye  Sljakspearian  Ctterature

of

€ti0tonfc,  (gprmang,

anb  otljer  €uropettit  Countries

buring  lljrec  Centum*,

3ntrobndwn$


Cit«atn«  from  1864  to  1871,


ftwixm,


foreign  fiookscller  anb

*A  firook  Stmt,  <5rosotnor  Sqnort,
1872.
```


Figure 7: OCR and image of English title page from D.



```
from

1564 to Ue4^'-^~:!.:Ji


of tl)e 5l)ttk0pearion f itcttttturc

of

anb oilier (Europcon Countries

liuring l;i)ret Ccnturits,

roiti) Sibliogrttpljical Sntrobttdions

1)5

1/

OecoitD Odihom

^outainlng tl)c ^iftrature from 1864 to 187U


foaign bookseller anli yublislier

ti firook street, OSrosuenor Squar«.
18 7 2.


Digitized by


Google
```


Figure 8: OCR and image of English title page from E.

When we turn to Thimm's introduction, however, we see markedly fewer errors in the OCR'd transcription (see **Figure 9**, **Figure 10**, and **Figure 11**, all digitized versions of the 1872 edition). When it comes to the introduction, Text C is the least faithful to the page. (For more on how OCR is undertaken at the Internet Archive, see Wajer [2020] 2023.) Both the C and E texts include an added prefatory page from Google in the file and in the transcription, which will be familiar to scholars (see **Figure 12**; the OCR of this boilerplate also differs between both texts, with more errors in C). Text E also includes a digital watermark on each page that reads "digitized by Google." The OCR transcription includes this watermark; this means that the word "Google" appears 134 times in this digital text, and not at all in the physical one. The added material from Google (the prefatory page and digital watermark) does not exclude this digital text from possible text analysis, but rather reminds us of the importance of understanding, and perhaps cleaning, our text before we undertake text analysis. (These could, for instance, be easily removed from the digital text before text analysis, though it would be important to document these changes).

```
SKETCn OF THE PP.OGRESS OF SIIAKSPEARIAN GKITICISM,

AND or THE GRAnCAI, AI'PREUUTIOX OF SBAKSPEABB'^ -.



ENGLAND.



Tbe history of Sbakspearian criticism ib one which goes hand in
hand with that of the general literary and critical art of England:
uay, Shakapeare'a works would seem lo have been particularly deaigned
to test the march of English intellect. It will therefore be uecessai'y
to glance nt the successive publications of his works, in order to show
the effect Ihoy produced on English writers.

The separate- plays of the great dramatist were -issued during his
life-time; in what consecutive order it is now impossible to say; though
certain it is that Shakspeare himself conld never Ijave seen them, even
separately, through the press. They appeared in a corrupt state from
the begiuning; for, being printed and puhlished as acting plays, they
were --altered, corrected and "improved" by both actors and man-
agers.
```

Figure 9: OCR from the Internet Archive for Thimm's introduction from C.

Dirty OCR (as we see in **Figure 4, Figure 5, Figure 6, Figure 7**, and **Figure 8**; and **Figure 9, Figure 10**, and **Figure 11**) has downstream effects on scholarly research. David A. Smith and Ryan Cordell point out that "researchers and students alike rely on search in OCR collections, often unwittingly" (Smith and Cordell 2018, 8). When we search large corpora such as the Internet Archive, or, indeed, the results that we find from any internet search, we find texts that have been automatically recognized. Cordell summed it up: "most humanities scholars … rely on the output of OCR algorithms. In other words, OCR is a fundamental element of our digital research infrastructure

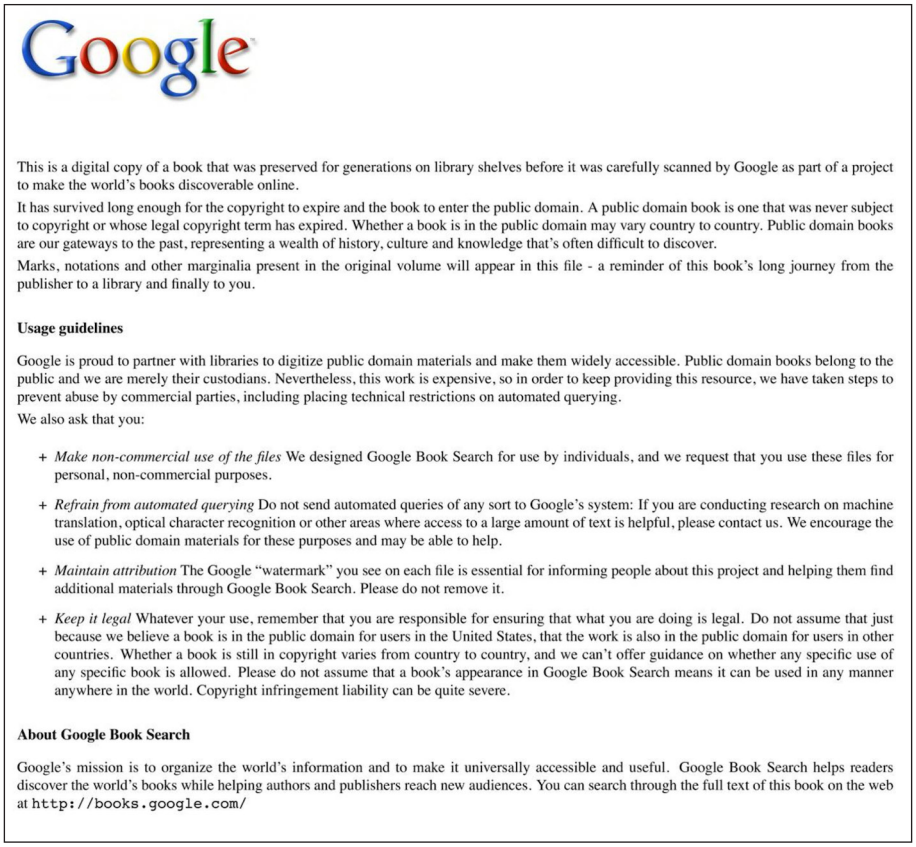
```
SKETCH  OF  THE  PROGRESS  OF  SHAKSPEARIAK  CRITICISM,

AND  OF  THE  GRADUAL  APPRECIATION   OP  SHAKSPEAEE

IN

ENGLAND.


The  history  of  Shakspearian  criticism  is  one  which  goes  hand  in
hand  with  that  of .  the  general  literary  and  critical  art  of  England :
nay,  Shakspeare's  works  would  seem  to  have  been  particularly  designed
to  test  the  march  of  English  intellect.  It  will  therefore  be  necessary
to  glance  at  the  successive  publications  of  his  works,  in  order  to  show
the  eifect  they  produced  on  English  writers.

The  separate  plays  of  the  great  dramatist  were  issued  during  his
life-time ;  in  what  consecutive  order  it  is  now  impossible  to  say ;  though
certain  it  is  that  Shakspeare  himself  could  never  have  seen  them,  even
separately,  through  the  press.  They  appeared  in  a  corrupt  state  from
the  beginning;  for,  being  printed  and  published  as  acting  plays,  they
were  altered,  corrected  and  "improved"  by  both  actors  and  man-
agers.
```

Figure 10: OCR from the Internet Archive for Thimm's introduction from D.

```
SKETCH OF THE PROGRESS OF SHAKSPEARIAN CRITICISM,

AND OF THE GRADUAL APPRECUTION OF 8HAKSPEARE

IN

ENGLAND.



The history of Shakspearian criticism is one which goes hand in
hand with that of. the general literary and critical art of England:
nay, Shakspeare's works wonld seem to have heen particularly designed
to test the march of English intellect. It will therefore be necessary
to glance at the successive publications of his works, in order to show
the effect they produced on English writers.

The separate plays of the great dramatist were issued during his
life-time ; in what consecutive order it is now impossible to say ; though
certain it is' that Shakspeare himself could never have seen them, even
separately, through the press. They appeared in a corrupt state from
the beginning; for, being printed and published as acting plays, they
were altered , corrected and " improved " by both actors and man-
agers.
```

Figure 11: OCR from the Internet Archive for Thimm's introduction from E.

that's also easily overlooked because we tend to focus on the images of historical pages rather than the underlying text data that helped us find them" (Cordell 2019). Because researchers are often unaware of the quality of the OCR they are searching or using to draw quantitative information from, Ian Milligan suggests that "there is a good chance that we are now ascribing to ourselves greater authority than is warranted" (Milligan 2013, 564). Milligan calls for "a critical methodology" when it comes to using scholarly databases and corpora of historical texts. In short: we must be critical about the texts we use and the claims we make as we undertake our scholarly searches.



Google

This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book's long journey from the publisher to a library and finally to you.

Usage guidelines

Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

+ *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.
+ *Refrain from automated querying* Do not send automated queries of any sort to Google's system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.
+ *Maintain attribution* The Google "watermark" you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.
+ *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can't offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book's appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

About Google Book Search

Google's mission is to organize the world's information and to make it universally accessible and useful. Google Book Search helps readers discover the world's books while helping authors and publishers reach new audiences. You can search through the full text of this book on the web at `http://books.google.com/`

Figure 12: The added front matter from Google Books in Text E.

Although our searches are proscribed by the often-flawed digital texts in historic databases, when it comes to text analysis, we are not, of course, limited to the OCR provided. The best practice to obtain clean OCR of large swathes of texts available to us online would begin with re-imaging the texts to get clearer scans, which would result in better OCR (Smith and Cordell 2018). We could also feed the existing scans through improved OCR pipelines, by for instance, using existing solutions for OCR (see Appendix A for examples). We could train our own OCR models (for a tutorial, see Pinche and Spychala 2024 and the other modules on "Automatic Text Recognition: Harmonising ATR Workflows," available in English, German, and French). Recent scholarship suggests the possibility of using large language models to correct OCR'd texts (Thomas, Gaizauskas, and Lu 2024; Veninga 2024; Do et al. 2025), though there is still much work to be done, particularly for non-English languages (Kanerva et al. 2025; Sohail, Masood, and Iqbal 2024). As Laura Mandell and Elizabeth Grumbach put it bluntly, however: "The cost of microfilming, digitizing, cameras, scanners, servers, programmers, associating metadata with files, and OCR'ing texts is very high" (Mandell and Grumbach 2015, 2).

And yet, as we know from our own research practices, we already have extensive texts that are digitized and OCR'd. What kind of text analysis can we do without unlimited time, training, and funding? The rest of this article offers an attempt at answering Cordell's question: "What tasks might be possible with existing OCR?" (Cordell 2019). In order to answer Cordell's question, I turned to the OCR from the Internet Archive to choose which text to analyze. Comparing the existing digital transcriptions of Thimm's bibliography discussed in the first section of this article, I selected digital Text D as best suited for analysis, because it was the expanded second edition with the cleanest OCR. In this case, I looked over the existing transcriptions and judged which was the most accurate. For larger texts or corpora, you can evaluate the accuracy of automatic text recognition algorithmically (see, for instance, Gupta et al. 2015; Hartel and Dunst 2018). I offer some conclusions that can be drawn from this digital text analysis. At the article's conclusion, I compare the results to the clean OCR from the Austrian National Library.

Text analysis with Voyant Tools

Voyant Tools is a longstanding, open-access, web-based digital text analysis toolkit that is often used in humanities, social sciences, libraries, and adjacent fields of scholarship. As of this writing, there are over five thousand results when you search "Voyant Tools" in Google Scholar. Many of these publications described the results from using Voyant Tools to analyze literature or primary sources such as social media posts; some suggest using Voyant Tools to aid in creating secondary information such as metadata or search keywords (such as McGowan 2021 and Gregory, Geiger, and Salisbury 2022). Recently, Janelle Bitter applied Voyant Tools to bibliographic metadata, considering the titles, summaries, and subjects of materials held in an academic library to consider the loaded language in catalogue subject headings (Bitter 2024). While this article explores the merits of undertaking text analysis on dirty OCR, it also shows the value of using Voyant to analyze enumerative bibliographies.

And so, I loaded the Internet Archive Text D into Voyant Tools, using the predefined English-language stopwords. Stopwords are the words that the computer ignores to undertake text analysis, often including numbers, articles, and prepositions, among other things. For more on the importance of stopwords, see Rockwell and Sinclair (Rockwell and Sinclair 2016, esp. 35–40) and Miller (Miller 2018, esp. 190–194). Note that stylometric analyses approach stopwords differently and might choose to emphasize how an author uses articles and prepositions, many of which are excluded in the default Voyant stopword list (Sinclair and Rockwell 2025a). Loading Text D into Voyant was easy to undertake, because I could simply download the PDF from the Internet Archive

and then upload it. Unsurprisingly, for a volume dedicated to capturing information about publications about Shakespeare, "Shakspeare" and variations on Shakespeare's name featured prominently (**Figure 13**).

Figure 13: Visualization from unchanged text. Stéfan Sinclair and Geoffrey Rockwell, "Cirrus," Voyant Tools.

In order to better, as the Voyant slogan says, "see through your text," I added "Shakespeare," "Shakspeare," "Shakespeare's," and "Shakspear's" to the pre-existing stopword list. Voyant documentation explains how to change the stopword list (Sinclair and Rockwell 2025a). As Voyant's document summary told me, these words appeared many times throughout the corpus (538, 599, 224, and 174 times, respectively), so removing them changed the text analysis results greatly (see **Figure 14**).

Figure 14: Visualization after removing Shakespeare's name and variants. Stéfan Sinclair and Geoffrey Rockwell, "Cirrus," Voyant Tools.

With Shakespeare out of the picture, so to speak, the Cirrus visualization revealed … that we needed to remove German stopwords as well. The most frequent words being counted (that is, excluding stopwords) were now "von" (881), "und" (386), "edition" (333), "lond" (289), and "der" (276), as the summary panel helpfully explained. The choice to remove the stopwords from multiple languages should not be undertaken lightly in a book that includes multilingual content. For instance, "die" means two wildly different things in German and in English: a version of the definite article "the" (German); or a conjugation of the verb relating to death (English). Looking at the contexts panel in Voyant revealed, however, that in Thimm's bibliography, "die" was only used in German and so could safely be excluded from our analysis; in a book or corpus where "die" was being used in both German and English, this would not be the appropriate course of action. For more on the contexts panel and its value, sometimes referred to as KWIC or keyword in context, see Sinclair and Rockwell (Sinclair and Rockwell 2016, esp. ch. 3).

"Voyant Tools supports multilingualism in various ways," the help documentation announces (Sinclair and Rockwell 2025b), before specifying that this entails offering multiple language options for the interface and having pre-loaded stopwords available in multiple languages. As Masoud Ghorbaninejad, Nathan P. Gibson, and David Joseph Wrisley explain, as they detail the creation of an Arabic Voyant interface, different languages will have different interface and textual analysis needs (Ghorbaninejad, Gibson, and Wrisley 2023). As the Voyant Tools help documentation notes, however, "in some ways Voyant Tools supports analysis in any language since it mostly operates on character sequences" (Sinclair and Rockwell 2025b). While there are lots of examples of Voyant Tools being applied to non-English texts and corpora, as well as its use as a tool to support translators (Horenberg 2023), so far as I know, this essay is the first attempt to use it to analyze a multilingual text, that is, a text comprising multiple languages. This is perhaps foolhardy, because, as Julie McDonough Dolmaya points out, "for bilingual and multilingual corpora, tools such as Sketch Engine and AntPConc are frequently used" (Dolmaya 2023, 110). As of 2024, Voyant provided pre-loaded lists of stopwords for thirty-eight languages; these lists of stopwords for each language can be accessed on GitHub (Voyant Tools 2025). Although you cannot select more than one language of stopwords at once in the online interface, you can go to the GitHub page to copy and paste multiple lists together to upload your own list. Rockwell and Sinclair warn that "Interpretive tools extend interpretation in ways that require caution" (Rockwell and Sinclair 2016, 19); and so, let us proceed cautiously.

With German stopwords removed, the new word cloud suggested this analysis could also be facilitated by removing the French stopwords, which were starting to appear in the Cirrus wordcloud ("et," "en," "par," etc). The resulting visualization

and data (**Figure 15**, with stopwords from English, German, and French removed, as well as variations on Shakespeare's name) was more useful but obviously could still be improved: for instance, the OCR clearly read "ü" as "ii" in "über" (over) and "übersetzt" (translated). The word "af" appeared, which, the Voyant "Contexts" panel showed came entirely in one section of the book: the Swedish and Danish materials (where "af" means "of").

Figure 15: Visualization after removing multilingual stopwords. Stéfan Sinclair and Geoffrey Rockwell, "Cirrus," Voyant Tools.

After removing stopwords from English, German, French, and a handful of other languages, and with the most minimal intervention in the text provided from the Internet Archive, we now have a visualization that offers information about Thimm's bibliography, and, by proxy, the state of Shakespeare studies in the mid nineteenth-century. (For a complete list of the stopwords used and for the final text used for analysis, see Estill 2025.) These shared datasets are one step towards answering calls for increased replicability in digital humanities and OCR (Verhaar 2022; Cooke and Litvack-Katzman 2024). As far as cleaning the text, I removed the first page of unintelligible OCR and removed the final text that was captured from the University of California library borrowing cards. I used find and replace to replace "iiber" with "über" and "fiir" with "für"; "frangais" with "français"; and "^" with "8" when it is next to another number (in this volume, "8" is often unevenly inked). A quick search revealed that "ii" could not be universally replaced with "ü" because it appeared in other places, like "Haliiwell," a poor-OCR reading of "Halliwell." While much of the OCR is adequate, Page 36 of the scan is blurry, resulting in OCR like: "Ihe Bafcop of Gloueester4* ooarrel wWi die kmet, abotK fab eiUkw of | Sfaakeapeaie's pbj^ to wtidi is added an impartial aemat of die | extraonfinary means used to snncw fie remarkable letter, *. aod foL." Despite an imperfectly OCR'd text, we can start to see information emerge including places, types of publication, and names of people and works (**Figure 16**).

Figure 16: Final visualization. Stéfan Sinclair and Geoffrey Rockwell, "Cirrus," Voyant Tools.

Thimm's bibliography emphasizes publications from major Western European cities, especially London (and "Lond"; "Londres" also appears a handful of times in the French section), Paris, Leipzig, Vienna (Wien), and Berlin. We are beginning to see where scholarship about Shakespeare was being published. We cannot take this information as a reliable quantitative truth and make claims that, for instance, 208 works were published in Leipzig and 146 in Berlin, both because Thimm's bibliography, like all bibliographies, is incomplete, and, as Milligan warns, flawed OCR will lead to flawed qualitative analysis. However, this visualization tells us that English publishing of and about Shakespeare was more concentrated in London (with Stratford also appearing multiple times), whereas German publishing of and about Shakespeare was spread more evenly across cities. We see, too, that New York appears less frequently than many other places of publication by quite a margin: it is listed by an order of magnitude less than Paris, Berlin, and London. From Voyant's "Document Terms" tool, which lists all word occurrences in a document and can be sorted by the number of a times a term appears in a corpus, we can see that Gotha, a smaller German city, appears as a publishing hub, with its outputs outstripping New York's.

Finding Gotha represented in the imperfect OCR 49 times raised the question of what Shakespeare-related works were being published there. Turning to Voyant's "Contexts" panel reveals that the works Thimm indexed that were published in Gotha were mainly translations, primarily by Heinrich Döring and Carl Joseph Meyer. Christa Jansohn (Jansohn 1995) argues for the importance of considering publication history to the reception of history, with a focus on nineteenth-century Germany, relating how Meyer started publishing Shakespeare translations in Gotha, translating some himself and securing Döring for others. Döring's/Meyer's translations were published in over 21,000 copies (Jansohn 1995, 551) and sold by traveling salesmen ("kolporteure") with subscription offers, though buyers could also choose to buy only a single edition

(Jahnson 1995, 547–548; McCarthy 2018, esp. 6, 9). Jahnson demonstrates that by offering cheap individual translations, Meyer participated in "the establishment of Shakespeare as a national poet in Germany" (Jansohn 1995, 555). Meyer's *Shakespeare's Sämmtliche Schauspiele* (published in Gotha from 1824–1834) is often noted in lists of translations or bibliographies as a single work comprising fifty-two volumes, but Thimm's bibliography reminds us of the importance of considering these fifty-two volumes as separate books. Voyant, here, validates Jahnson's research and suggests the importance of further studying the Meyer's publications with an attention to Döring's contributions.

Considering *Shakespeare's Sämmtliche Schauspiele* as separate books also emphasizes the importance of Döring's contributions to the volumes (he completed over half of the translations in this series, as Voyant reveals) and encourages us to re-evaluate Döring's contributions to the history of German Shakespeare. For comparison, in this uncorrected text, August Wilhelm von Schlegel appears fifty-three times and Döring appears thirty-three times. On the one hand, Schlegel is well known and has been studied for years as an influential German translator, appearing in, for instance, the Arden "Great Shakespeareans" series (Roger and Paulin 2010). Döring, on the other hand, has been generally overlooked to date.

Using Voyant to analyze Thimm's bibliography emphasizes the importance of translations and translators: both "übersetzt" and "ubersetzt" (German for "translated") appear prominently in the Cirrus visualization. Many of the most frequent names that appear in Cirrus and in the document terms are those of German translators: Voss, Schlegel, Tieck, and Ortlepp appear more than forty times each, despite the imperfect OCR. This is equivalent to the number of times major English scholars and editors are mentioned, such as Capell and Collier, adding to the evidence that Thimm positioned German Shakespeare scholarship on a par with its English counterparts.

When undertaking text analysis with Voyant, disambiguation of names is important to take into account, as are conventions of credit and citation. There were two German Shakespeareans named Tieck active in the nineteenth century, Ludwig and his daughter, Dorothea, though only Ludwig was credited on the original publications. The complete works known as the "Schlegel-Tieck Shakespeare" (published 1825–1833) was incredibly influential (Smith 2021); it is "*the* German Shakespeare" (Newman 2011, 116, emphasis in the original). Despite advertising itself as a collaboration between Schlegel and Ludwig Tieck, the translations were actually completed by Schlegel, Dorothea Tieck, and Wolf Graf Baudissin (Larson 1987), with contributions from Caroline Schelling (who was for a time married to August Schlegel) and Friedrich Schlegel (Stott 2009; Smith 2021) and Sophie Tieck (Doleschal 2017). The title page of the second edition of the Schlegel Tieck Shakespeare, which was instrumental in this

translation's popularity (Newman 2011, 116), credits Ludwig Tieck with translating plays he did not translate. Christian Smith writes: "after the first edition of the translation, von Baudissin's name was added to the title page; Dorothea Tieck's and Caroline (Schelling) Schlegel's were not" (Smith 2023, 21). As Smith notes, "it is well known that Schlegel-Tieck is its brand name, not the complete roster of who performed the translations" (Smith 2021, 234). In Voyant's text analysis, "Tieck" is not a single person with that name, or even, two: rather, it is symbolic of collective authorship and branding that both built and built on Ludwig Tieck's cultural capital.

Although Thimm's bibliography replicates the publication of Ludwig Tieck's name on this edition, Thimm also added additional information about play translators. Thimm writes, "A great many of the plays were translated by Count Wolff von Baudissin, a very elegant translator; and six were the work of Tieck's daughter, Dorothea" (Thimm 1872, 54). Thimm credited Dorothea and Baudissin for the plays they translated, despite the fact that Dorothea's name does not appear anywhere in the Schlegel-Tieck edition until twentieth century editions (Smith 2023, 5). Dorothea completed six translations (alphabetically: *Coriolanus*, *Cymbeline*, *Macbeth*, *Timon of Athens*, *Two Gentlemen of Verona*, *Winter's Tale*) (Smith 2023); Thimm credits her for each of these translations in his list of German translations. Thimm similarly credits Baudissin for each play he translated. The bulk of the references to Tieck, however, as Voyant reveals, are to Ludwig. Thimm justified continuing to give him credit for the work: "but he was the editor and critic of the whole work, and went over all the translations with great care. His corrections indeed were so numerous, that it would be difficult to deny him the credit of having taken a share in the work" (Thimm 1872, 54). When it comes to "Tieck" in the text analysis of Thimm's bibliography, Ludwig is represented more than twice as much as Dorothea, even though Thimm acknowledged that "Ludwig Tieck himself did not even translate a whole play" (Thimm 1872, 54).

Thimm's bibliography begins to point to German women translating Shakespeare, but it is still not representative of their actual contributions. Schelling played a "considerable" role in Schlegel's translations, but Schlegel himself downplayed her contributions (Paulin 2016, 92–93). While now sometimes credited as co-translator, Thimm credits Caroline's contributions to only a single play: *Hamlet*. With under-represented women translators, Thimm's bibliography reflects the practices of his time that viewed women's contributions to literary studies as "ancillary" (Smith 2021).

Dorothea Tieck's and Caroline Schelling's contributions offer two examples of the disdain with which women's contributions were seen. Dorothea's translations were derisively described in the 1919 *Encyclopedia Americana*: "Only 17 plays of the so-called Schlegel-Tieck Shakespeare were translated by Schlegel (1797–1810); the remainder were added, from 1825 to 1833, by Graf Wolf Baudissin and (very inadequately) by

Dorothea Tieck, while Tieck himself acted only as a reviser and annotator" (quoted in Stott 2009). In his 1913 edition of Caroline's letters, Ernst Schmidt wrote, "a more thorough study still remains to be done concerning Caroline's contribution to Wilhelm Schlegel's translation of Shakespeare to determine the extent to which, during the process of copying, Caroline also arbitrarily made things worse through otherwise well-intentioned corrections or choices" (quoted in Stott 2009). Using Voyant will only capture people named in the texts analyzed, and as such, it is important to understand the contents and historical contexts of the works being interpreted.

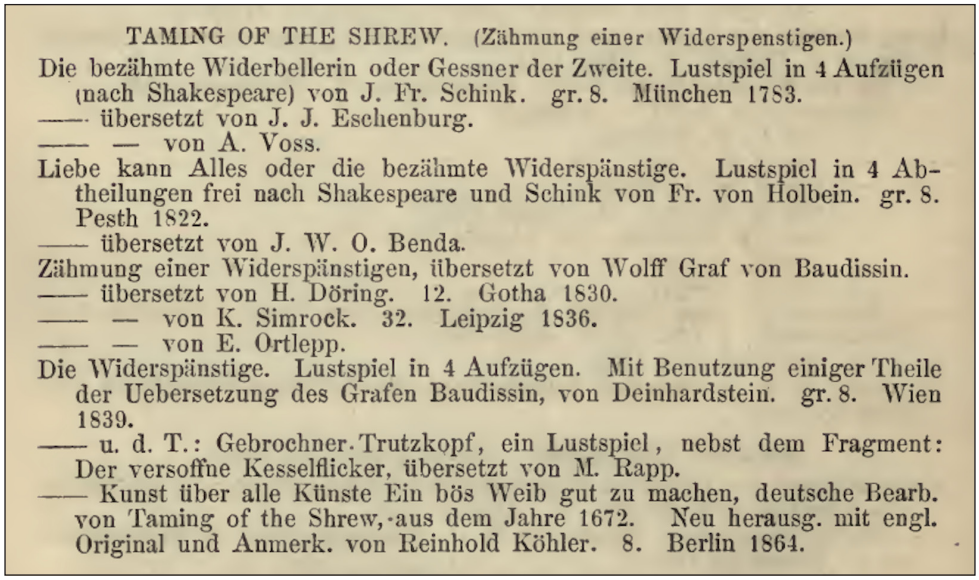
If even the surnames present in this text can cause disambiguation problems (with, for instance, the two Tiecks), the most captured given names in Thimm's bibliography seem certain to likewise stymy quantitative analysis. In this OCR, "John" is the most prominent given name (appearing 82 times), followed by "Richard" (62), and "William" (56). Even "Henry," which is in the title of six separate Shakespeare plays (*1 Henry IV, 2 Henry IV, 1 Henry VI, 2 Henry VI, 3 Henry VI,* and *Henry VIII*) is counted only 48 times. ("Hamlet," "Macbeth," and "Romeo" also feature prominently but are used to primarily refer to play titles.) While Shakespeare does have a play titled *King John*, the Voyant collocates panel reveals that only a handful of the uses of John refer to that history play. The collocates panel "is a table view of which terms appear more frequently in proximity to keywords across the entire corpus" (Sinclair and Rockwell 2025c). And, according to Voyant, the word most frequently found with John is Falstaff.

John Falstaff is the famous drunkard of *1* and *2 Henry IV*—who was allegedly so popular that Queen Elizabeth herself requested that Shakespeare write another play featuring Falstaff, to which, as the apocryphal story goes, Shakespeare responded by making Falstaff the main character of the *Merry Wives of Windsor* (Hepokoski 1983, 161). Applying Voyant Tools to even the uncorrected OCR of Thimm's bibliography bears out what we have long known: Falstaff was one of the most popular characters of the nineteenth century as Rosemary Gaby (Gaby 2019) and others have traced. For instance, "in 1817 William Hazlitt claimed that Falstaff was one of the greatest comic characters ever invented" (Gaby 2019). Voyant lets us quickly navigate Thimm's text to see that Falstaff appears in multiple places: in the titles of works adapted from Shakespeare's plays, and also in the title of scholarly articles about Shakespeare.

Beyond places of publication and names, Voyant also surfaces many play titles. Shakespearean bibliography allows us to see what plays were being published, translated, and written about. By manually enumerating centuries of Spanish Shakespeare, for instance, Ángel-Luis Pujante and Juan F. Cerdá can claim that, in Spain, the major tragedies have been of longstanding interest, whereas some plays like *Coriolanus* are only now achieving popularity (Pujante and Cerdá 2015, XL). With Dominic Klyve and Kate Bridal, in a previous publication, I quantified the scholarly

research about Shakespeare in the late twentieth century from the *World Shakespeare Bibliography* data set, and we established that *Hamlet* was more studied than any other work, and by such a large margin that it was statistically important (Estill, Klyve, and Bridal 2015; and the dataset, Estill and Klyve 2016). We were able to undertake this analysis because the *World Shakespeare Bibliography* categorizes the materials it lists.

One limitation of applying Voyant Tools to Thimm's bibliography appears with play titles. Without understanding Thimm's text, a researcher might attempt to quantify which plays were most published or written about by finding the number of times play titles appear in the text. In **Figure 16**, "Hamlet" features prominently (the Voyant "Contexts" panel tells us this refers primarily to the play title, not the character); less prominently, "Romeo" and "Macbeth" can be seen. Thimm's bibliography, however, does not lend itself to the analysis of play titles because (1) Thimm used the common bibliography practice of not repeating titles in lists, and (2) many of the play titles will appear in different languages. **Figure 17** shows Thimm's list of the German translations of *Taming of the Shrew*, which Thimm notes is translated as *Zähmung einer Widerspentigen.* Of course, not every translator uses the same title: the first translation listed, by Johann Friedrich Schink, offers a different title *Die bezähmte Widerbellerin.* The changes in titles across languages also accounts for why "Romeo" appears with much more frequency than "Juliet," whose name is sometimes given as "Julie," "Julia," and "Giulietta." And while most German-language translations use the classic title for *Taming of the Shrew* (*Zähmung einer Widerspentigen*), Thimm does not repeat titles in lists, instead offering the familiar line or "do" (ditto) for repetition (see **Figure 17**):



TAMING OF THE SHREW. (Zähmung einer Widerspenstigen.)

Die bezähmte Widerbellerin oder Gessner der Zweite. Lustspiel in 4 Aufzügen (nach Shakespeare) von J. Fr. Schink. gr. 8. München 1783.

—— übersetzt von J. J. Eschenburg.

—— — von A. Voss.

Liebe kann Alles oder die bezähmte Widerspänstige. Lustspiel in 4 Abtheilungen frei nach Shakespeare und Schink von Fr. von Holbein. gr. 8. Pesth 1822.

—— übersetzt von J. W. O. Benda.

Zähmung einer Widerspänstigen, übersetzt von Wolff Graf von Baudissin.

—— übersetzt von H. Döring. 12. Gotha 1830.

—— — von K. Simrock. 32. Leipzig 1836.

—— — von E. Ortlepp.

Die Widerspänstige. Lustspiel in 4 Aufzügen. Mit Benutzung einiger Theile der Uebersetzung des Grafen Baudissin, von Deinhardstein. gr. 8. Wien 1839.

—— u. d. T.: Gebrochner-Trutzkopf, ein Lustspiel, nebst dem Fragment: Der versoffne Kesselflicker, übersetzt von M. Rapp.

—— Kunst über alle Künste Ein bös Weib gut zu machen, deutsche Bearb. von Taming of the Shrew, aus dem Jahre 1672. Neu herausg. mit engl. Original und Anmerk. von Reinhold Köhler. 8. Berlin 1864.

Figure 17: Detail from Page 68 of Thimm's *Shakspeariana from 1564–1864* (1872), Text D, digitized by the Internet Archive.

eight editions about *Shrew* would not be captured by Voyant, meaning that even if play titles were always translated the same way, Thimm's *Shakspeariana* (unlike other enumerative bibliographies) does not lend itself to using Voyant to quantify which plays were most written about.

While using Voyant on imperfect OCR does not allow us to draw quantitative conclusions, it can help guide our inquiries. As Rockwell and Sinclair write, "Interpretive tools focus on the particularity of the work and its poetic or rhetorical language. Rather than show a theory of textuality, they assist the reader to interpret the meaning of a text or to follow a text's rhetorical structure. These tools augment reading rather than replace it" (Rockwell and Sinclair 2016, 17). When I saw, for instance, that Voyant surfaced more than fifty instances of "thlr" and "sgr," which are abbreviations for "thaler" and "silbergroschen," German currency, it led me to dive deeper into the text. Voyant automatically removes single letters from the stopwords, which means the indications of British currency would be suppressed from the results, which are written in L.s.d notation (pounds.shilling.pence; for more on this, see Hitchcock 2023). Having seen that Thimm offers prices for some of the German volumes, I then returned to Thimm with an eye to prices across the volume: he offers British books for sale in pounds, shillings, and pence, and French books for sale in francs and centimes. Thimm does not provide the price for every book he lists, but he does offer a note "To Shakspearian Collectors": "The Publisher begs to enform Libraries, and Collectors of Shakspeariana, that he has great facilities for supplying any of the books mentioned in the Catalogue English as well as Foreign" (Thimm 1872, [119]). The facing-page German translation emphasizes only his ability to source English materials (Thimm 1872, [120]). Voyant Tools, then, surfaces one of the audiences of this book: "Shakspearian Collectors." Thimm's inclusion of the cost of books is partly a sales pitch: "he has great facilities for supplying" (Thimm 1872, [119]) the volumes listed in this bibliography, should a collector be interested.

The first step of this analysis of Thimm's *Shakspeariana* with Voyant Tools was to remove mentions of Shakespeare—but Thimm's bardolatry pours through regardless. Voyant surfaced recurring uses of the word "great" (fifty times)—and, as I could tell from the contexts panel, these were not in listed bibliographical citations, but in prose. Returning to the text, these are almost entirely in Thimm's preface and introductory materials, expressing such emphatic praise as "so great a genius as Shakspeare" (53) and monikers such as "the great dramatist" (53) and "the great bard" (54). When we turn to related words, we see Thimm discussing "the greatness of Shakspeare" (5) and "Shakspeare's greatness" (54). Thimm's admiration and passion for Shakespeare shines through in this bibliography, though his prose accounts for only a fraction

of it. Thimm's language, as explored with Voyant, mirrors the effort to compile this bibliography: this volume is a labour of love for Shakespeare and his works.

To conclude, I turn to some much better OCR of the second edition of Thimm's bibliography, provided by the Austrian National Library (Österreichische Nationalbibliothek), Text K (**Table 2**; while the OCR of the blackletter German title page is imperfect, the Roman type text is well captured). Using the same stopword list, we can compare top word frequencies to determine the quality of Text D's OCR (**Table 3**). While the Internet Archive OCR missed some words, it was usually fewer than 10. (One notable exception is the word "übersetz," where Text D separated "übersetz" and "ubersetz," counting 78 without the umlaut.)

Term	Count in ANL edition (Text K)	Count in Internet Archive edition (Text D)	Difference in count
edition	339	340	−1
lond	297	291	6
london	293	286	7
vols	266	256	10
leipzig	210	208	2
1864	210	182	28
übersetzt	201	120	81
hamlet	184	181	3
paris	170	172	−2
plays	153	149	4
berlin	147	146	1
notes	130	126	4

Table 3: Comparison of most frequent words in Text K OCR with the OCR from Text D.

The terms brought to the fore by both digital texts are by and large, the same, which suggests that text analysis is possible with moderately dirty OCR, and with some of the imperfect OCR provided by major online resources. Of course, not all OCR across these resources (or even, as we saw above, in the same repository) is of comparable quality. Given the uneven OCR quality, it is key for researchers to know both the physical text and the digital text *before* undertaking digital text analysis. Text analysis on OCR'd texts, counterintuitively, cannot tell us about the text without, as Rockwell and Sinclair

advocate, moving from the digital text to the tool and back. This is not interpretation by a computer; it is, as Rockwell and Sinclair emphasize with their book's subtitle, "computer-assisted interpretation."

Conclusions

The conclusions to this paper are both methodological and interpretive. Methodologically, we have seen that you can apply Voyant Tools to pre-existing OCR from major online repositories and learn about your text. We considered the caveats that need to be taken into account before undertaking text analysis on imperfect OCR and explored how imperfect OCR can still be usable enough to guide research questions (though its usability will vary greatly depending on its quality).

As a foray into using Voyant on a multilingual text, we were able to undertake some basic text analysis. This is because Thimm's bibliography focused on, as his title page advertised, England, Germany, and France: to apply this method to a truly international bibliography, you could analyze different sections according to their language. The contents of *The World Shakespeare Bibliography* (*WSB*), the largest bibliography of Shakespeare studies today, could be fruitfully analyzed in this regard because they offer English descriptions of their contents in the annotations and most titles are translated into English. The *WSB* data could also be analyzed by other digital means beyond text analysis by quantifying the information stored in their taxonomy tags such as language of publication challenge. (The *WSB*'s taxonomy tags also solve the problem of titles translated in different ways).

When it comes to historical bibliographies, there are also many directions to be taken. Can we apply this text analysis to multiple bibliographies (such as those mentioned at the outset of this article or in Estill, forthcoming) or are their contents or organizational structure too diverse? Given that these bibliographies often cover large swathes of overlapping material, would they have to be taken separately, or can we envision a large-scale project that identifies their contents and tracks how each work appears in different bibliographies, akin to existing analyses of literary anthologies? Could textual analysis be fruitfully applied to periodical bibliographies, such as the "Shakespeare Bibliographie" that appeared in *Shakespeare Jahrbuch* for over a century? (For more on this bibliography and other periodical bibliographies, see Craig et al. 2026.) Can we extract the data from these bibliographies in a more sophisticated way (bearing in mind that this work will take time, money, and effort)? For instance, printed historical bibliographies offer quasi-structured data with their entries: if these were reimaged and OCR'd to a high standard of accuracy, could we extract the information

with other tools that would allow us to make quantitative claims about the state of a discipline? Kyle Dase, for instance, shows how visualizations allow us to explore John R. Roberts's *John Donne: An Annotated Bibliography of Modern Criticism* (Dase 2024).

As Shawn Graham, Ian Milligan, and Scott Weingart suggest, Voyant is "'a gateway drug' when it comes to textual analysis" (Graham, Milligan, and Weingart 2015, 85)—bibliographies could benefit from more sophisticated textual analysis. Graham, Milligan, and Weingart continue by praising Voyant as "arguably the best research portal in existence" (Graham, Milligan, and Weingart 2015, 85); here we see that its use extends to secondary, enumerative materials. Voyant is also often used in the classroom: we could extend its pedagogical use to thinking about bibliography and literary studies. And, of course, we can apply these techniques to bibliographies beyond Shakespeare and even beyond literary studies, though each new text will need to be considered carefully in light of Voyant's affordances to determine the claims that can be made.

Future work could consider publication dates of scholarship and editions, mapping places of publication, republications, and how these bibliographies relate to sale catalogues, library catalogues, and other early documentation. Using digital text analysis on a bibliography can give us a snapshot of the scholarly preoccupations of an era.

Beyond the methodological considerations outlined here, this article has offered a number of interpretive conclusions about Thimm's *Shakspeariana from 1564–1864*. This research has underscored the prominence of Shakespeare translation in the nineteenth century and its dissemination. This analysis considered where editions, translations, and scholarship were being published, which led us to the importance of the Meyer-Döring translations and publications, a topic which bears further study. Working with Thimm's bibliography in Voyant emphasized that how we list (and cite) scholarship can marginalize or erase people's contributions (such as Dorothea Tieck's or Caroline Schelling's) while centralizing other voices. Considering the given names in Thimm's *Shakspeariana* confirmed Falstaff's outsized reputation.

Bibliographies are not texts that were meant to be read from front to back, and so invite computer-assisted reading strategies. Applying digital text analysis to bibliographies is a fruitful way to analyze the state of scholarly discourse. Ultimately, experimenting with digital tools to analyze early bibliographies will help us better understand the history and foundations of our scholarship.

Appendix A. Additional examples of automatic text recognition (ATR)

These are examples of optical character recognition outputs from the Internet Archive page image "Sketch of the Progress of Shakspearian Criticism" (p. 1) of Text C of Thimm's bibliography.

A1. From Adobe Acrobat Professional

> SKETCHO F THEP ROGRESOSF SHAKSPEAnICARNI TICISM,
> AXD OF THE GRADUALA PPRECIA.TIO:ONF" SHAKSPEARE
> IN
> ENGLAND.
> The history of Shakspearian criticism is one which goes hand in hand with that of. the general literary and critical art of England: nay, Shakspeare's works would seem to have been particularly designed to test the march of English intellect. It wil.l therefore be necessary to glance at the successive publications of his works, in order to show the effect they produced on English writers.
> The separate plays of the great dramatist were issued during his life-time; in what consecutive order it is now impossible to say; though certain it ·is· that Shakspeare himself could never have seen them, even separately, through the press. They appeared in a corrupt state from the beginning; for, being printed and published as acting plays, they were altered, corrected and "improved" · by both actors and managers.

A2. From Transkribus

> S O PoEss O SAsPEA Cs,
> AXD OF TE CRADAL APPRECAToX o SAxsPEARE
> ENGLAND.
> The history of Shakspearian criticism is one which goes hand in hand with that of. the general literary and critical art of England: nay, Shakspearé's works would seem to have been particularly designed to test the march of English intellect. It will therefore be necessary to glance at the successive publications of his works, in order to show the effect they produced on English writers.
> Tle separate plays of tle greut dramatist were issued during his life-time; in wlat consecutive order it is now impossible to say; though certain it is that Shakspeare himself could never have seen them, even-separately, through the press. They appeared in a corrupt state from the beginning; for, being printed and published as acling plays, they were altered, corrected and 'improved" by both actors and managers.

Acknowledgments

An early version of this paper was first presented at the Canadian Society for Digital Humanities (CSDH) annual conference in 2024. Thanks especially to Harvey Quamen for his thoughtful feedback at the CSDH meeting. I would like to thank Heidi Craig, Kris L. May, and Dorothy Todd for their

collaboration and feedback as we wrote *Collaboration, Technologies, and the History of Shakespearean Bibliography*, the project from which this research question emerged, and especially to Heidi, whose feedback was timely and invaluable. Thank you to Christa Bader-Reim at the Austrian National Library. Thanks also to the journal's blind peer reviewers for their constructive and detailed feedback and to Christa Avram for her meticulous copyediting.

Competing interests

The author has no competing interests to declare.

Contributions

Editorial

Section Editor

Davide Pafumi, The Journal Incubator, University of Lethbridge, Canada

Copy Editor

Christa Avram, The Journal Incubator, University of Lethbridge, Canada

Layout Editor

A K M Iftekhar Khalid, The Journal Incubator, University of Lethbridge, Canada

References

Barnett, Tully. 2020. "Public-Private Partnerships and the Digitization of the Textual and Cultural Record." *Pop! Public. Open. Participatory*, no. 2 (October). Accessed August 2, 2025. https://doi.org/10.54590/pop.2020.009.

Bitter, Janelle. 2024. "Text Mining Bibliographic Metadata for Inclusivity: Analyzing Most Frequent Words in Titles, Summaries, and Subjects." *Library Resources & Technical Services* 68 (4). Accessed August 2, 2025. https://doi.org/10.5860/lrts.68n4.8329.

Bjerring-Hansen, Jens, Ross Deans Kristensen-McLachlan, Philip Diderichsen, and Dorte Haltrup Hansen. 2022. "Mending Fractured Texts: A Heuristic Procedure for Correcting OCR Data: 6th Digital Humanities in the Nordic and Baltic Countries Conference, DHNB 2022." *CEUR Workshop Proceedings* 3232: 177–186. Accessed August 2, 2025. https://ceur-ws.org/Vol-3232/paper14.pdf.

Burchardt, Jørgen. 2023. "Are Searches in OCR-Generated Archives Trustworthy? An Analysis of Digital Newspaper Archives." *Jahrbuch für Wirtschaftsgeschichte/Economic History Yearbook* 64 (1): 31–54. Accessed August 2, 2025. https://doi.org/10.1515/jbwg-2023-0003.

Christy, Matthew, Anshul Gupta, Elizabeth Grumbach, Laura Mandell, Richard Furuta, and Ricardo Gutierrez-Osuna. 2017. "Mass Digitization of Early Modern Texts with Optical Character Recognition." *Journal on Computing and Cultural Heritage* 11 (1): 1–25. Accessed August 2, 2025. https://doi.org/10.1145/3075645.

Cooke, Nathalie, and Ronny Litvack-Katzman. 2024. "Open *Times*: The Future of Critique in the Age of (Un)Replicability." *International Journal of Digital Humanities* 6 (1): 71–85. Accessed August 2, 2025. https://doi.org/10.1007/s42803-023-00081-y.

Cordell, Ryan. 2017. "'Q i-jtb the Raven': Taking Dirty OCR Seriously." *Book History* 20 (1): 188–225. Accessed August 2, 2025. https://doi.org/10.1353/bh.2017.0006.

———. 2019. "Why You (A Humanist) Should Care About Optical Character Recognition." *Programmable Type* (blog), January 19. Accessed August 2, 2025. https://ryancordell.org/research/why-ocr/.

———. 2020. "Speculative Bibliography." *Anglia* 138 (3): 519–531. Accessed August 2, 2025. https://doi.org/10.1515/ang-2020-0041.

Craig, Heidi, and Laura Estill. 2022. "Browse as Interface in Shakespeare's Texts and the World Shakespeare Bibliography Online." In *The Routledge Handbook of Shakespeare and Interface*, edited by Clifford Werier and Paul Budra, 218–233. Routledge.

Craig, Heidi, Laura Estill, Kris L. May, and Dorothy Todd. 2026. *Collaboration, Technologies, and the History of Shakespearean Bibliography*. Cambridge University Press.

Dase, Kyle. 2024. "Love's Short Day: Romance and Illumination in the 'Light Sequence' of John Donne's Poems." *Huntington Library Quarterly* 87 (1): 105–127. Accessed August 3, 2025. https://doi.org/10.1353/hlq.2024.a949376.

Do, Thao, Dinh Phu Tran, An Vo, and Daeyoung Kim. 2025. "Reference-Based Post-OCR Processing with LLM for Precise Diacritic Text in Historical Document Recognition." arXiv:2410.13305v3. Accessed August 3. https://doi.org/10.48550/arXiv.2410.13305.

Doleschal, Mareike. 2017. "Shakespeare in German." *Shakespeare Birthplace Trust* (blog), October 3. Accessed August 3, 2025. https://www.shakespeare.org.uk/explore-shakespeare/blogs/shakespeare-german/.

Dolmaya, Julie McDonough. 2023. *Digital Research Methods for Translation Studies*. Taylor & Francis.

Estill, Laura. 2025. Data for "Thimm's Shakspeariana from 1564–1864 and Stopwords for Digital Text Analysis [Version 1]." St. Francis Xavier University Borealis Dataverse. Accessed September 16. https://doi.org/10.5683/SP3/XTZ7MH.

———. Forthcoming. "A Bibliography of Online Open-Access Bibliographies of Shakespeare Scholarship (1814–1920)."

Estill, Laura, and Dominic Klyve. 2016. "Writing About Shakespeare: 1960–2010." *Journal of Open Humanities Data* 2: e3. Accessed August 3, 2025. https://doi.org/10.5334/johd.4.

Estill, Laura, Dominic Klyve, and Kate Bridal. 2015. "'Spare Your Arithmetic, Never Count the Turns': A Statistical Analysis of Writing about Shakespeare, 1960–2010." *Shakespeare Quarterly* 66 (1): 1–28. Accessed August 3, 2025. https://www.jstor.org/stable/24778623.

Gaby, Rosemary. 2019. "Critical Reception." In *Henry IV, Part 1*, edited by Rosemary Gaby. Internet Shakespeare Editions. Accessed August 3, 2025. https://internetshakespeare.uvic.ca/doc/1H4_CriticalSurvey/complete/index.html.

Ghorbaninejad, Masoud, Nathan P. Gibson, and David Joseph Wrisley. 2023. "Right-to-Left (RTL) Text: Digital Humanists Plus Half a Billion Users." In *Debates in the Digital Humanities 2023*, edited by Matthew K. Gold and Lauren F. Klein, 47–73. University of Minnesota Press. Accessed August 3, 2025. https://dhdebates.gc.cuny.edu/read/debates-in-the-digital-humanities-2023/section/51ad75ab-76a4-43f3-81f8-0b31e350ca9c#ch03.

Google Books. 2025. Record: "*Shakspeariana from 1564 to 1864: An Account of the Shakespearian Literature of England, Germany, France and Other European Countries during Three Centuries, with Bibliographical Introductions* by Franz J. L. Thimm, 1890." Accessed August 12. https://books.google.ca/books?id=23JeuAAACAAJ.

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. Imperial College Press.

Gregory, Kate, Lauren Geiger, and Preston Salisbury. 2022. "Voyant Tools and Descriptive Metadata: A Case Study in How Automation Can Compliment Expertise Knowledge." *Journal of Library Metadata* 22 (1–2): 1–16. Accessed August 3, 2025. https://doi.org/10.1080/19386389.2022.2030635.

Gupta, Anshul, Ricardo Gutierrez-Osuna, Matthew Christy, Boris Capitanu, Loretta Auvil, Liz Grumbach, Richard Furuta, and Laura Mandell. 2015. "Automatic Assessment of OCR Quality in Historical Documents." *Proceedings of the AAAI Conference on Artificial Intelligence* 29 (1): 1735–1741. Accessed August 3, 2025. https://doi.org/10.1609/aaai.v29i1.9487.

Hartel, Rita, and Alexander Dunst. 2018. "How Good Is Good Enough? Establishing Quality Thresholds for the Automatic Text Analysis of Retro-Digitized Comics." In *MultiMedia Modeling (MMM 2019)*, edited by Ioannis Kompatsiaris, Benoit Huet, Vasileios Mezaris, Cathal Gurrin, Wen-Huang Cheng, and Stefanos Vrochidis, 662–671. Accessed August 3, 2025. https://doi.org/10.1007/978-3-030-05716-9_59.

Hepokoski, James A. 1983. *Falstaff*. Cambridge University Press.

Hill, Mark J., and Simon Hengchen. 2019. "Quantifying the Impact of Dirty OCR on Historical Text Analysis: Eighteenth Century Collections Online as a Case Study." *Digital Scholarship in the Humanities* 34 (4): 825–843. Accessed August 3, 2025. https://doi.org/10.1093/llc/fqz024.

Hitchcock, Tim. 2023. "Currency, Coinage and the Cost of Living: Pounds, Shillings & Pence, and Their Purchasing Power, 1674–1913." *The Proceedings of the Old Bailey*, version 9.0. Accessed August 3, 2025. https://www.oldbaileyonline.org/about/coinage.

Horenberg, Lisa. 2023. "Voyant Tools' Little Outing: How a Text Reading and Analysis Environment Can Help Literary Translators." In *Computer-Assisted Literary Translation*, edited by Andrew Rothwell, Andy Way, and Roy Youdale, 203–218. Routledge.

Jansohn, Christa. 1995. "The Making of a National Poet: Shakespeare, Carl Joseph Meyer and the German Book-Market in the Nineteenth Century." *The Modern Language Review* 90 (3): 545–555. Accessed August 3, 2025. https://doi.org/10.2307/3734314.

Kanerva, Jenna, Cassandra Ledins, Siiri Käpyaho, and Filip Ginter. 2025. "OCR Error Post-Correction with LLMs in Historical Documents: No Free Lunches." arXiv:2502.01205v1. Accessed August 3. https://doi.org/10.48550/arXiv.2502.01205.

Larson, Kenneth E. 1987. "The Origins of the 'Schlegel-Tieck' Shakespeare in the 1820s." *The German Quarterly* 60 (1): 19–37. Accessed August 3, 2025. https://doi.org/10.2307/407156.

Mandell, Laura, and Elizabeth Grumbach. 2015. "The Business of Digital Humanities: Capitalism and Enlightenment." *Scholarly and Research Communication* 6 (4), 1–9. https://doi.org/10.22230/src.2015v6n4a226.

Martin, Shawn. 2007. "EEBO, Microfilm, and Umberto Eco: Historical Lessons and Future Directions for Building Electronic Collections." *Microform and Imaging Review* 36 (4): 159–164. Accessed August 3, 2025. https://repository.upenn.edu/entities/publication/da5a9541-a4ec-4fd4-903e-06fbe0c627d5.

McCarthy, John A. 2018. "The 'Great Shapesphere.' An Introduction." In *Shakespeare as German Author: Reception, Translation Theory, and Cultural Transfer*, edited by John A. McCarthy. 1–74. Brill Rodopi.

McGowan, Bethany S. 2021. "Using Text Mining Tools to Inform Search Term Generation: An Introduction for Librarians." *portal: Libraries and the Academy* 21 (3): 603–618. Accessed August 3, 2025. https://preprint.press.jhu.edu/portal/sites/default/files/21.3mcgowan.pdf.

Miller, A. 2018. "Text Mining Digital Humanities Projects: Assessing Content Analysis Capabilities of Voyant Tools." *Journal of Web Librarianship* 12 (3): 169–197. Accessed August 3, 2025. https://doi.org/10.1080/19322909.2018.1479673.

Milligan, Ian. 2013. "Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010." *The Canadian Historical Review* 94 (4): 540–569. Accessed August 3, 2025. https://doi.org/10.3138/chr.694.

Newman, Jane O. 2011. *Benjamin's Library: Modernity, Nation, and the Baroque*. Cornell University Press; Cornell University Library. Accessed August 3, 2025. https://doi.org/10.7298/3bfx-8b79.

Paulin, Roger. 2016. *The Life of August Wilhelm Schlegel, Cosmopolitan of Art and Poetry*. Open Book Publishers. Accessed August 3, 2025. https://doi.org/10.11647/obp.0069.

Pinche, Ariane, and Pauline Spychala. 2024. "ATR Step 1: Getting Started with Automatic Text Recognition." *Automatic Text Recognition* (blog), April 15. Accessed August 3, 2025. https://harmoniseatr.hypotheses.org/274.

Pujante Ángel-Luis and Juan F. Cerdá, eds. 2015. *Shakespeare en España: Bibliografia Anotada Bilingue | Shakespeare in Spain: An Annotated Bilingual Bibliography*. Universidad de Murcia and Universidad de Granada.

Quiring, Ana. 2024. "Fingerprints of British Book History: A Feminist Labor History of EEBO." *Digital Humanities Quarterly* 18 (1). Accessed August 3, 2025. https://www.digitalhumanities.org/dhq/vol/18/1/000718/000718.html.

Rockwell, Geoffrey, and Stéfan Sinclair. 2016. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. MIT Press. Accessed August 3, 2025. https://doi.org/10.7551/mitpress/9522.001.0001.

Roger, Christine, and Roger Paulin. 2010. "August Wilhelm Schlegel." In *Voltaire, Goethe, Schlegel, Coleridge*, edited by Roger Paulin, 92–127. Vol. 3 of *Great Shakespeareans*. Bloomsbury Publishing.

Sinclair, Stéfan, and Geoffrey Rockwell. 2025a. "Stopwords." Voyant Tools Help. Accessed August 13. https://voyant-tools.org/docs/tutorial-stopwords.html.

———. 2025b. "Languages." Voyant Tools Help. Accessed August 13. https://voyant-tools.org/docs/tutorial-languages.html.

———. 2025c. "Corpus Collocates." Voyant Tools Help. Accessed August 13. https://voyant-tools.org/docs/tutorial-corpuscollocates.html.

Smith, Bruce R. 1991. “Reading Lists of Plays, Early Modern, Modernist, Postmodern.” *Shakespeare Quarterly* 42 (2): 127–144. Accessed August 3, 2025. https://doi.org/10.2307/2870544.

Smith, Christian. 2021. “Translating Orchids: Rhizomes in German Shakespeare Translation: Case Study: Caroline Schlegel and the Schlegel-Tieck Translation.” In *The Shakespearean International Yearbook*, edited by Ton Hoenselaars, Tom Bishop, Stephen O’Neill, and Alexa Alice Joubin, 231–243. Routledge.

———. 2023. “Translation and Influence: Dorothea Tieck’s Translations of Shakespeare.” *Borrowers and Lenders: The Journal of Shakespeare and Appropriation* 11 (2): 1–29. Accessed August 3, 2025. https://doi.org/10.18274/HCFF8259.

Smith, David A., and Ryan Cordell. 2018. *A Research Agenda for Historical and Multilingual Optical Character Recognition*. Historical and Multilingual OCR. NU Lab, Northeastern University. Accessed August 3, 2025. https://ocr.northeastern.edu/report/.

Sohail, Muhammad Abdullah, Salaar Masood, and Hamza Iqbal. 2024. “Deciphering the Underserved: Benchmarking LLM OCR for Low-Resource Scripts.” arXiv:2412.16119v1. Accessed September 8, 2025. https://doi.org/10.48550/arXiv.2412.16119.

Stott, Douglas W. 2009. “Caroline and the Translation of Shakespeare.” Accessed August 3, 2025. https://www.carolineschelling.com/caroline-and-shakespeare/.

Tanselle, G. Thomas. 1989. *A Rationale of Textual Criticism*. University of Pennsylvania Press.

Thimm, Franz J. L. 1865. Introduction to *Shakespeariana from 1564 to 1864: An Account of the Shakespearian Literature of England, Germany and France during Three Centuries, with Bibliographical Introductions*. London.

———. 1872. Introduction to *Shakespeariana from 1564 to 1864: An Account of the Shakespearian Literature of England, Germany and France during Three Centuries, with Bibliographical Introductions*. 2nd ed. London.

Thomas, Alan, Robert Gaizauskas, and Haiping Lu. 2024. “Leveraging LLMs for Post-OCR Correction of Historical Newspapers.” In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024*, edited by Rachele Sprugnoli and Marco Passarotti, 116–121. European Language Resources Association; International Committee on Computational Linguistics. Accessed August 3, 2025. https://aclanthology.org/2024.lt4hala-1.14/.

Trettien, Whitney Anne. 2013. “A Deep History of Electronic Textuality: The Case of *English Reprints Jhon Milton Areopagitica*.” *Digital Humanities Quarterly* 7 (1). Accessed August 3, 2025. https://dhq.digitalhumanities.org/vol/7/1/000150/000150.html.

Veninga, Martijn E. B. 2024. “LLMs for OCR Post-Correction.” Master’s thesis, University of Twente. Accessed September 2, 2025. https://web.archive.org/web/20241108161417/http://essay.utwente.nl/102117/1/Veninga_MA_EEMCS.pdf.

Verhaar, Peter. 2022. “Towards a Conceptualisation of Replication in the Digital Humanities.” *Txt* 8: 95–108. Accessed August 3, 2025. https://scholarlypublications.universiteitleiden.nl/access/item%3A3465914/download.

Verheusen, Astrid, Hans van Dormolen, and Lotte Wilms. 2011. “Digitising Surrogates: Scanning from Microfilm.” Zenodo. https://doi.org/10.5281/zenodo.3621360.

Voyant Tools (@voyanttools). 2025. “Stopwords.” GitHub. Accessed August 12. https://github.com/voyanttools/trombone/tree/master/src/main/resources/org/voyanttools/trombone/stopwords.

Wajer, Merlijn. (2020) 2023. “OCR at the Internet Archive with Tesseract and hOCR.” Internet Archive. Accessed August 3, 2025. https://archive.org/developers/ocr.html.

WorldCat. 2025. Catalogue record: “Shakspeariana from 1564 to 1864: An Account of the Shakespearian Literature of England, Germany, France and Other European Countries during Three Centuries, with Bibliographical Introductions. 3rd ed. / edited, rev. & brought down to 1890 by C.A. Thimm.” OCLC 9939391. Accessed September 2. https://search.worldcat.org/title/9939391.