

Chapter Title: Parsing the Special Characters of African American Print Culture: Mary Ann Shadd and the Limits of Search

Chapter Author(s): Jim Casey

Book Title: Against a Sharp White Background

Book Subtitle: Infrastructures of African American Print

Book Editor(s): Brigitte Fielder, Jonathan Senchyne

Published by: University of Wisconsin Press. (2019)

Stable URL: <https://www.jstor.org/stable/j.ctvgs08p1.9>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

University of Wisconsin Press is collaborating with JSTOR to digitize, preserve and extend access to *Against a Sharp White Background*

Parsing the Special Characters of African American Print Culture

*Mary Ann Shadd and the * Limits of Search*

Jim Casey

The ongoing remediation of newspapers into digital surrogates is proving a remarkable mixed boon. More than ever, scholars enjoy online access to a dizzying volume of rare and hard-to-find newspapers. Where previous generations required the time and money to travel the country scrounging for files of newspapers, today we may be just a click or two away from websites where we can run keyword searches on hundreds, millions, or even billions of words. In the process, a variety of subfields of periodicals studies have sprung up, all eager to mine these enormous repositories for new materials and new ways to understand our literary, cultural, and print histories.

Across these fields, digital repositories have become central, even crucial. Yet because the processes for creating and developing these digital resources are rarely transparent, a growing number of scholars have begun to scrutinize the politics of those processes. From the archives, as Benjamin Fagan and others have noted, the selection of newspapers for digitization often reflects racial and other social inequalities.¹ Once digitized, many commercial databases are prohibitively expensive, restricting access to those researchers at wealthy institutions.

These vital conversations have only just begun to contend with the full breadth of necessary questions, already inspiring a wave of academic journal issues and projects focused on the distinct challenges of periodicals in digital environments.²

Those journals and projects join decades of work to create searchable digital resources. While many of these questions reach back to the rise of microfilm in the early twentieth century, today groups such as the Text Encoding Initiative (TEI) and the Text Creation Partnership (TCP) offer robust foundations for creating accurate digital editions of historical texts. At larger scales, it becomes necessary to rely on optical character recognition (OCR) tools that attempt to guess the words that appear on a digitized page. OCR tools have long struggled with the thin paper and faint inking of many historical newspapers, yet important strides are being made by groups such as the Early Modern OCR Project along with crowdsourcing platforms such as the National Library of Australia's Trove and the Zooniverse.³ Poor-quality OCR may not be solved just yet, but those groups and platforms point to reliable long-term solutions for developing searchable historical newspapers.

Critiques of coverage and OCR, however, can account only partly for the impact on historical research of search engines. If conversations at the intersection of print culture and digital studies have tended to focus on the integrity of digital repositories, the modes of access enabled by search play a distinctly adjacent and still-invisible role. Databases of digital surrogates of historical newspapers are daisy chains of associated technologies. Each link in the chain introduces its own contours, its own affordances and limits, inviting us to interact with these digital artifacts in highly structured ways.⁴ Faced with a range of ways to interact with a database, what is our first step when we wade into a database that, like *Chronicling America*, holds more than ten million pages? For most of us, research suggests, the first step in the research process is to "Just Google It."⁵ The volume of these databases, even as they afford the ability to read and browse widely, preserves the formidable challenge of what some call the "Great Unread."⁶ While a small number of scholars have attempted to employ algorithms to read for broader patterns in these databases, the resources and expertise for such methods elude most researchers working today. For many, the primary response to the challenge of the Great Unread is a keyword search. As search tools have become a basic component of historical research in the early twenty-first century, it bears stressing that there is no such thing as a neutral search engine.

This essay is about the limits of search in research on early African American print cultures. Specifically, I present the case of Mary Ann Shadd, the earliest-known African American woman to edit a newspaper. Shadd signed her name

to her columns with an asterisk. The asterisk was a way for Shadd to navigate the complex of racism and sexism that disrupted her editorship of her weekly newspaper, the *Provincial Freeman* (1853–57). In our current search engine technologies, that asterisk is forbidden from having any semantic meaning. The asterisk is a special character, prevented from having semantic meaning of its own, in order to help search engines operate more efficiently. The case of Mary Ann Shadd's asterisk is a cautionary tale not only against depending on keyword search in research but also on the fault lines of race and gender that shaped historical print culture and infuse information retrieval technologies.

This essay takes up those asterisks and the complex task of reading them in the nineteenth century and searching for them in the twenty-first. In the nineteenth century, Shadd's asterisks highlight an encoded form of editorial expression using punctuation that suggests the tensions between exposure in print and self-effacement. In the twenty-first century, asterisks take on new meanings within search engines that erase the typographical marks as meaningless—beyond what we can search, read, or locate.

**An Unprecedented Editor: Mary Ann
Shadd and the *Provincial Freeman*,
1853–54**

Mary Ann Shadd founded the *Provincial Freeman* in the spring of 1853 when, after her first few years living in Canada, she had realized the need for a platform of her own. She emigrated to Canada in 1851 in the wake of the Fugitive Slave Act along with many other northern free and fugitive African Americans. Shadd was born in Wilmington, Delaware, to Abraham and Harriet Shadd, who ran a successful small business and participated actively in the Philadelphia-area, middle-class community of activists, abolitionists, and Colored Convention organizers. Shadd was one of many of her peers dedicated to building such institutions as schools, conventions, and the black press. Her first stop was in a small town called Windsor, across the border from Detroit, in what was then known as Canada West. Shadd had been trained as a teacher and had some experience as an educator, so she opened a school to teach the children of the many people arriving from the United States. Windsor was the endpoint for many on the Underground Railroad, but arriving into Canada West guaranteed little in the way of financial or other certainties. Very few of the recent emigrants from the United States could afford to pay Shadd much for their children's education. Instead, she began to rely on the American Missionary Association, a white-led philanthropic organization, for the \$125 it cost her

annually to keep the school open. Her funding from the AMA, however, did not last long, a break she long suspected was the result of her increasingly bitter rivalry with Henry Bibb and his newspaper, the *Voice of the Fugitive*.⁷ Initially on friendly terms, the two had found themselves on opposite sides of the acrimonious debates about the direction of the coalescing broader community of Canada West's black population.⁸

That population included multiple, overlapping communities. Some people had already been living in Canada for years prior to the Fugitive Slave Act. Others came from northern American states, where they had lived as free or fugitive citizens with deeply curtailed civil rights. Many others came from the South, where they had lived as enslaved human beings.⁹ Along with the poverty facing many of them upon their arrival into Canada, a signal struggle for early black communities in Canada West was the challenge of social ties for so many whose previous and current living situations were sources of intense personal pain, loss, and trauma. Almost all of them were refugees from the United States, yet few of them shared much else in common. Bibb's and Shadd's disputes in this view are less a contest between two large personalities than a set of collective debates within an emerging and heterogeneous population straining to build new lives in Canada West.

Within those collective debates, Shadd soon found herself the object of personal attacks. As rumors began to circulate in 1852–53 that Shadd and others were planning to start a paper to rival the *Voice of the Fugitive*, Bibb published a number of sexist articles attacking Shadd. One article referred to her as “a designing individual whose duplicity is sufficient to prove a genealogical descent from the serpent that beguiled mother Eve in the Garden of Eden.”¹⁰ The personal attacks only got worse. The need was clear: Shadd needed a space in print to refute these attacks. She had already tried to combat these attacks in public venues, but the physical performances before mixed-gender audiences exposed her to even harsher criticism. As Carla Peterson observes, Shadd's trials “dictated that Shadd Cary devise other modes of representation more suitable to her talents that would also draw the public gaze away from any ‘blot or blemish’ that it might imagine recorded on her body.”¹¹

Absent Editor, Invisible Editor: Shadd Finds the *Provincial Freeman*

And so in March 1853 Shadd managed to publish a trial issue of a newspaper named the *Provincial Freeman*. She listed many men's names on the masthead of the first issue, including Samuel Ringgold Ward as the editor and the Reverend Alexander McArthur as the corresponding editor. Given Ward's extensive

experience in the press, his name on the masthead as the ostensible editor lent the *Freeman* instant credibility.¹² Another seven names appear as the Committee of Publication,¹³ followed by instructions for correspondents that “letters must be addressed, *Post-paid*, to MARY A. SHADD, Windsor, Canada West; Rev. J. B. Smith and J. Baker, Travelling Agents.” Shadd had omitted her name as the paper’s editor. Instead, an editorial column titled “Introductory” by Samuel Ringgold Ward appeared on the front page, despite an article on the backside of that page noting that he was hundreds of miles away when the paper went to press. If that did not belie the actual editorship of the paper, Ward’s “Introductory” column showed little enthusiasm for the new venture, declaring his reluctance to accept the editorial post and his plans to vacate soon. Ward was soon bound for England on a speaking tour; he remained in England for more than a year. He continued to travel widely in England, Canada, and the United States throughout the life of the *Freeman*.

A year after the trial issue, Shadd began issuing the *Freeman* regularly with a masthead nearly unchanged except for the addition of a white British printer named John Dick. Dick was a journeyman printer who had moved across the Atlantic to work for Frederick Douglass at the birth of the *North Star* in 1847. Along with his work on the *North Star*, Dick’s politics are conveyed by a few of the other texts he printed, including the proceedings of the Seneca Falls women’s rights convention and the 1848 national Colored Convention.¹⁴ With his past work for women’s rights, against slavery, and in newspaper publishing, John Dick was likely a welcome addition to the fledgling efforts led by Shadd to relaunch the *Provincial Freeman* in 1854.

A journeyman printer’s experience would have contributed much to the day-to-day affairs of the *Provincial Freeman* in many areas, including the organization of collaborative editorial writing. In the 1840s and 1850s the vast majority of editorial writing in the United States used the collective “we” and went unsigned. Most periodicals only had one or two editors, so readers could usually detect who had written what. There were a few rare and notable exceptions, and those usually came only on newspapers with large editorial departments. The *New York Tribune* began to have editors initial their columns in 1849. The *North Star* used a similar practice, as with the April 7, 1848, issue, which featured editorials by F.D. (Frederick Douglass), J.D. (John Dick), and M.R.D. (Martin R. Delany). The system of editorial initials was not always consistent, even in the *Tribune* or *North Star*, but served as a useful shorthand for readers to differentiate a host of contributing editors and writers.

The *Provincial Freeman* adapted that organizational system for its unusual situation of having both an absent editor and an invisible editor. When the printer, John Dick, contributed a column, he signed his writings with the letter

Figure 11. Example of Mary Ann Shadd's asterisk signature in an untitled article, *Provincial Freeman*, May 27, 1854, 2. Courtesy of INK—Our Digital World Newspaper Collection.

The irregularity of *other papers*, is of no benefit to ours, we have already had ample proof, and we do not wish to have an unfounded charge of neglect on our part, made against us also. *

At the late anniversary of the American Anti-Slavery Society, held in New York, the following resolution was adopted. It shows at least, that the members of that Society continue courageous and hopeful, and intend to persevere in their labors for the enfranchisement of the enslaved, through the regeneration of the moral sentiment of the American people. An arduous task, truly! D.

D. For the first eight months of 1854, that letter was the paper's only overt editorial attribution, save for the several dozen columns marked at their conclusion with an asterisk. The *Provincial Freeman* published at least thirty-nine editorial columns with an asterisk. These columns have a distinctive prose style, they are rife with subjunctive clauses, and they include a running focus on emigration and women's rights. That combination of qualities corresponds to Mary Ann Shadd's previous writings and her many public speeches. I follow Jane Rhodes in concluding that these asterisks could only be the mark of Shadd's invisible editorship, a mark that Shadd would deploy in 1854 for making quotidian announcements, for airing controversial opinions, and as a personal signifier within limited social circuits.¹⁵ The asterisk was a complex usage of the affordances of nineteenth-century newspaper editorship.

Some of the articles signed with an asterisk were relatively quotidian. As in the snippet above (figure 11), Shadd used the asterisk for a number of articles that did not ordinarily require any attribution. Along with a lament about the late mails in the spring weather, she published others like the July 7, 1854, piece that read, in its entirety: "Several frame buildings beginning at the corner of Queen and Yonge streets, were destroyed by fire on Tuesday last *" (figure 13). This and other articles did not necessarily require any rhetorical protection afforded by the asterisk. For example, it is hard to imagine the possible negative consequences of publishing anonymous advertisements for traveling agents (April 22, 1854, and May 13, 1854) or to announce preparations for the First of

August celebrations in Toronto and Hamilton, Canada West (June 24, 1854 and July 1, 1854). The use of the asterisk may have signified an implicit endorsement in notices about local revival meetings (April 22, 1854) or the receipt of the *American Phrenological Journal* (July 15, 1854), but such conclusions are difficult to support with so few examples. It is entirely possible that readers hardly noticed the small symbols appended to some of the more ephemeral notes in the *Freeman*.

For wider audiences, other editorials used the veiled authority of the asterisk to treat much more controversial subjects. Throughout its publication the *Freeman* was staunchly pro-emigration, perhaps never more than in an editorial on June 3, 1854, in which Shadd declared opponents of emigration “as guilty, we verily believe, as Batchelder or any other slaveholder or slave-catcher in the land. *”¹⁶ Considering that the editorial was a commentary on the news that a brother of James W. C. Pennington had been kidnapped and taken south to be enslaved, Shadd’s asterisk licensed a tone and topic that she might have hesitated to broach under her own name.¹⁷ That proved true a number of times. In that same issue in June 1854, Shadd penned an article attacking Julia Griffiths and some of the abolitionists for organizing a fund-raiser in Toronto to benefit *Frederick Douglass’ Paper* while ignoring the *Freeman*’s more local and immediate needs. “But will not Miss Griffith,” the editorial asks, “leave a few coppers behind?”¹⁸ The reply came—if only the *Freeman* had asked, the Toronto Anti-Slavery Society would have been happy to host an event. They were less inclined, though, after the attack in print. Shadd used the asterisk, impolitic or not, as a way to access the full authority of the editorial voice. It was precisely the opaque quality of the asterisk, its very deficit of information, that enabled her to speak frankly without fear of direct personal reprisal. Self-effacement helped her move beyond the print and political spaces typically afforded to a person of her race and gender.

Whether Shadd deployed the asterisk for quotidian or controversial ends, readers could track the symbol from week to week. The asterisks were at once unknown and recognizable, opening up the potential for readers to accumulate their own senses of the marks appearing underneath so many of the *Freeman*’s editorials. By accruing meaning through their periodicity, their patterns and rhythms over time, Shadd’s asterisks fit within a much larger body of semi-anonymous writing in the early black press. That body of writing stretches across much of the nineteenth century, from articles signed with pen names such as Ethiop (William J. Wilson) and Communipaw (James McCune Smith) in *Frederick Douglass’ Paper* to later writing by Ida B. Wells under the name Iola. Although no studies currently exist on pseudonyms in nineteenth-century African

American periodicals, what is at least clear is that the use of such pseudonyms was not a neutral rhetorical gesture. For Shadd, the asterisk helped foreclose her exposure to any racist or sexist retaliation while simultaneously asserting the heft of her editorial presence.

As a typographical pseudonym, Shadd's asterisks likely signaled much more transparently within the circuits of her personal acquaintances. Along with her readers on fund-raising trips, Shadd interacted frequently with activists, writers, editors, and religious leaders across Canada West and the north-eastern United States. That William Still, of the prominent Still family, would serve as a regular correspondent for the *Freeman* attests to the reach of her personal networks as a woman who had lived and worked in Delaware, New York, Pennsylvania, and New Jersey before emigrating. More locally, Henry Bibb and his allies in Canada West had encountered Shadd in any number of public gatherings. They may have guessed at the identity of the person writing behind the asterisks. Shadd also traveled constantly throughout Canada West, gathering subscriptions, raising money, and writing travel reports for the *Freeman*. Reports published quickly after her visits gave subscribers clues to the identity of the person using the asterisk.

In those fluid meanings, Shadd drew on a much wider array of expressive uses of asterisks in antebellum print culture. As her father, Abraham Shadd, had been an agent for *Freedom's Journal*, she may well have had opportunity to read *David Walker's Appeal*, which, as Marcy Dinius explains, used a variety of typographical symbols to powerful effect.¹⁹ Along with the manicules (or pointing fist) in the *Appeal*, Walker unsettled the form of the pamphlet through asterisks that called the eyes of the reader or the voice of the person reading aloud back and forth between the primary texts and the many footnotes. Walker's asterisk was a transgressive character that bridged the main discussion and the marginal interjections. Shadd may have gleaned another use for her asterisks from Margaret Fuller in the *New York Tribune*. As the *Tribune* was the most frequently reprinted newspaper in the *Freeman* under Shadd's editorship, she was likely familiar with the hundreds of earlier columns in the mid-1840s that Fuller had signed with what she called her "star." Both Fuller and Shadd added the asterisk at the end of their editorial columns. Where Fuller used the asterisk to make herself more visible as a budding celebrity in the *Tribune*, Shadd inverted that usage to make herself more invisible as the editor of the *Freeman*. Still others, including Frank Leslie and James Redpath, used asterisks as placeholders for content deemed too odious or tantalizing for their audiences—as the mark of excised language. These examples are only a few of the conspicuous

and expressive uses of the asterisk in the antebellum press. It was a potent single character, a character capable of sliding meanings while it circulated across different audiences.

Search Engines as Social Texts

Shadd's asterisks gesture to a much larger possible literary history of editorial punctuation in the nineteenth century. I say "possible" because today those punctuation marks are being erased by the technologies employed to digitize the archives of early African American print culture. Shadd's asterisk has become doubly invisible in search-assisted databases. No matter how Shadd's contemporaries may have deciphered her mark, today the asterisk is parsed by search engines as a wildcard. The wildcard is one of a set of so-called special characters set aside by search engines as instruments for optimizing the speed and accuracy of information storage and retrieval. Owing to their need for these special characters, search engines disallow the asterisk as a result for any searches. Racial and gender fault lines that originated in print culture are perpetuated in the racial politics of search engines. It is, as Michel-Rolph Trouillot writes, a digital example of "formulas of erasure" that arise in the precise moments when historical facts are recorded, simultaneously excluding the unthinkable entities from the fields of power set up around recognized personal names.²⁰

The case of Shadd and her asterisks shows that search engines are social texts. Search engines encode the priorities of their creators into their operations, as in the case of the Lucene search system, which provides access to the repository of the *Provincial Freeman* in the Accessible Archives (AA) website. The name Lucene refers to search engines derived from a library of computer code written in the language Java by Doug Cutting in the late 1990s. In 2001 Lucene was taken over by the nonprofit Apache Software Foundation. Subsequent developers integrated it with web application software such as Solr, Tomcat, and Elasticsearch.²¹ The popularity of Lucene Apache is hard to overstate. It is used in thousands of the most highly trafficked websites today, from Twitter to Wikipedia and many more. Even those corporations that build their own search engines, such as Facebook and Google, share Lucene's basic principles for processing language. The broad impact of Lucene makes it an important technology to understand. A focus on Lucene is a reminder that every search engine contains its own set of ideas about language, even if those ideas are largely illegible without technical fluencies that remain uncommon in disciplines for the study of historical language, literature, and culture.

We do not need to be novelists to analyze novels; we do not need to be programmers to take a critical view of these technologies. We can approach them as users and ask valid questions. The following section explains the concepts and process of Lucene search engines. The intent is to provide a set of questions and vocabulary that humanists can use to critique our estranged dependence on search engines.

Digitization, or Adapting Archival Sources into Digital Texts

The process of creating a searchable database of archival sources begins with facsimiles. Figures 12 and 13 are facsimiles of two brief articles that Shadd published in the *Provincial Freeman*. The textual information in these facsimiles then needs to be converted into digital texts. Although many companies and organizations create digital texts by using OCR tools, AA is unusual in that all documents have been transcribed in double-key entry. Double-key entry means that two individual transcriptions were created by hand and then merged. That process allows AA to claim an accuracy rate of 98 percent based on a single dissertation study conducted by Wesley Raabe at the University of Virginia in 2006.²² While those claims are more than a decade old and merit reconsideration, the double-keyed transcriptions remove OCR tools as a relevant factor in the status of punctuation in the AA databases. Once transcribed, all the characters and words should be search eligible. Any alphanumeric character can be entered as a search query to retrieve the two articles in figures 12 and 13. The articles assuredly appear in the April 22, 1854, and July 1, 1854, issues of the *Provincial Freeman*, yet a search for the asterisks on those or any other dates returns only an error message: “Parsing Error.”

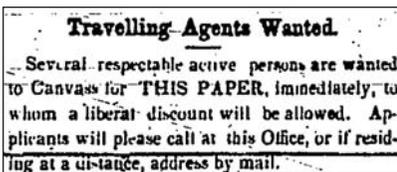


Figure 12. Sample snippet, “Travelling Agents Wanted,” *Provincial Freeman*, April 29, 1854, 2. Courtesy of INK—Our Digital World Newspaper Collection.

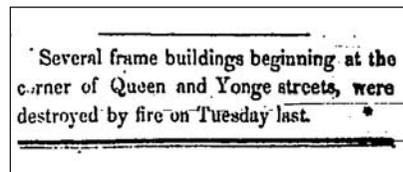


Figure 13. Sample snippet, untitled, *Provincial Freeman*, July 1, 1854, 2. Courtesy of INK—Our Digital World Newspaper Collection.

Lucene Indexing: From Text to Terms

The elaborate workflow of a Lucene search engine (see figure 14) begins with a process called tokenization. Tokenization breaks up the text and extracts single, independent tokens, or terms. This process uses filters, called analyzers, to determine which combinations of letters and numbers can be extracted as distinct terms. There are many different kinds of analyzers. Each has its own settings and produces certain kinds of terms, but the most commonly used is the StandardAnalyzer. The Lucene StandardAnalyzer converts all words to lowercase and removes all punctuation and diacritical marks.²³ The result of running the StandardAnalyzer on the facsimile texts in figures 12 and 13 would be as follows:

Table 1. StandardAnalyzer

documentID	Text
1	several respectable active persons are wanted to canvass for this paper immediately to whom a liberal discount will be allowed applicants will please call at this office or if residing at a distance address by mail
2	several frame buildings beginning at the corner of queen and yonge streets were destroyed by fire on tuesday last

In this filtering, Lucene interprets the texts on users' behalf. Users can still read or browse the full text, but tokenization extracts only the terms that satisfy its rules. Those terms are filtered a second time in the process of tokenization against what is called a stop word list. The purpose of stop words is to filter out those terms that appear frequently in almost every single document, including around thirty terms in the StandardAnalyzer, including *and*, *are*, *as*, *but*, *for*, *the*, and *with*.²⁴ Removing some stop words can improve search results quite a bit, but selecting which words to include or exclude is an idiosyncratic decision akin to a set of principles that would guide a documentary editing project. With the *Freeman*, a stop word list is unimportant, because AA does not use a stop word list out of sensitivity to the "uniqueness of the documents in the Accessible Archives databases, which in many instances use language no longer in common usage."²⁵ The composition of the stop word list is one place where

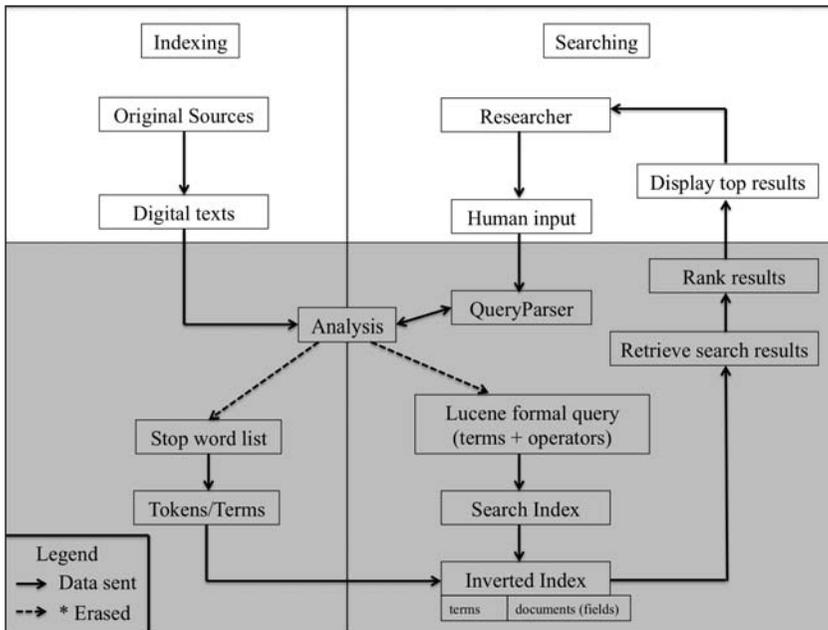


Figure 14. Workflow diagram of Lucene search.

humanists can and should ask questions about what words are filtered out and why.

Once extracted and filtered, the terms are assembled by Lucene into a unit called “documents,” a technical term that does not correspond to the vernacular sense of a document. Rather, a document is better understood as a collection of fields or as buckets of terms. A document may contain many fields. In AA the fields are named in a radio button to the right of the search box: full text, title, author, and “book publisher only.” The full-text field for the *Freeman* snippets would be the same set of terms after being parsed by the StandardAnalyzer. Researchers can also direct the system to look for a query only in a certain kind of field, such as searching for “Shadd” in the author field. While the technical notions of documents and fields can seem arbitrary, they are vital to understand because they are the stored information that queries search. A full-text search does not actually go into all the transcriptions of all the articles from all the newspapers. Rather, a full-text search scans all these documents and their fields.

Once each article has been tokenized, filtered, and adapted into a document with multiple fields, the system needs to organize everything into an

index. Index is here again a technically specific word. What Lucene calls an index may be more familiar to humanities scholars as a concordance. Like a concordance, a Lucene index provides a list of which terms appear in which documents. Lucene's inverted index maps terms to their appearances in any document. Table 2 shows a sample Lucene inverted index:

Table 2. Inverted Index for Documents 1 and 2

Term	documentID
several	[1], [2]
persons	[1]
by	[1], [2]
yonge	[2]
street	[2]

Most Lucene instances also enrich the term listings with additional information, such as length, location on the page (to enable fuzzy searching), synonyms, and frequency of a term in a document. A Lucene index may remove the common suffixes of words to allow matching between queries and words with similar roots, such as for “heads,” “headings,” “headed,” and so on.²⁶ As those types of information do not directly influence the handling of the asterisk, I omit them in this explanation. The advantages of this inverted index is that the system can store which words appear in multiple documents without having to store multiple copies of the text associated with each article. Inverted tables are then aggregated further into a tree-like graph structure that ensures the information storage can be easily maintained and expanded with additional tables without creating the need to recompute the entire index each time. Using those structures, Lucene makes the storage of textual information more memory efficient by avoiding the need to store copies of the full text of every article where the word “several” appears. Instead, the system can use those references to pull up links to relevant documents in the list of search results.

Such a view of the Lucene inverted index as a concordance suggests some new ways to think about searchable repositories of historical newspapers. A Lucene index is not an archive but a para-archive, a complex system for organizing, storing, and retrieving information abstracted from print or digital

sources. Because an index is an abstraction of the texts, it is a threshold that structures our encounters with surrogates of historical print objects. A search does not provide direct access to an archive. Searches engage a table of information abstracted from digital surrogates of printed objects. These layers underscore the role of faith implicit in these search acts—faith in the communities of computer scientists, programmers, and information professionals who contributed to the source code of Lucene and other search systems. Every technical development reflects the priorities set by a group of people about the balance of access, accuracy, and completeness. Their priorities become the practical structures, then, for these para-archives, which determine what kinds of expressions are legible and may have value.

Lucene Parsing and Special Characters

Once the index has been built to adequately store the parsed tokens in all the right fields and documents, it is ready for researchers to access it through a search box on a user interface through queries. When a researcher enters a query, Lucene modifies the string of letters or numbers into a formal syntax structured to find hits in the index intuitively and accurately. This process helps users query a system without needing to memorize any arcane syntax for formal logical expressions. That systemic generosity is enforced by a set of operations known as a query parser. There are many different varieties of query parsers. Each parser enables a different kind of search query by interpreting the search string. The range of options for advanced searches provides clues about what kinds of parsers are active in a given search engine.²⁷ AA appears to use the most common package, which in the latest release of Lucene 6.0.0 is called `QueryParser`.²⁸

`QueryParser` begins analyzing the string of characters by sorting them into terms and operators. `QueryParser` deliberately replicates the process used to create the index. But where the initial analysis for the index looks only for terms, this round of analysis breaks a query into both terms and operators. On AA those operators can be expressed by clicking buttons or by entering special characters: + - & | ! () { } [] ^ “ ~ * ? : \ / . Because these operators give explicit and clear instructions to the system, the special characters can be powerful tools for adept users. That power, however, comes at the cost of the asterisk being able to have the status of a term, a status that would make it findable and therefore have meaning in the search act. Instead, the analysis performed by the `QueryParser` acts as a translator from human natural language into Lucene syntax. Helpfully, AA supplies information about the results of this translation

act at the top of most search pages. Thanks to this information, I can see the translation of a query for mentions of Mary Ann Shadd and the *Freeman* in other newspapers that do not mention her father, Abraham, as (((freeman)) AND ((shadd))) NOT ((abraham))). This is a basic example, but it demonstrates the conversion of a researcher's language into a formal expression. In contrast, the QueryParser rejects any queries that contain only a single asterisk. The system replies with the message: "Lucene Parsing Error." The translating function of the QueryParser, then, is directly responsible for erasing Mary Ann Shadd's asterisk. The bounds of allowable translations by the QueryParser are one more place where researchers can and should ask critical questions: How does a system parse my questions into formal expressions?

Ranking Search Hits

The QueryParser rejects the asterisk in part because Lucene uses the special characters to organize the search results into a ranked list. A search for a common term can bring up millions of hits. Search engines depend on the logic of special operators to handle such large volumes efficiently. Access to thousands of articles in a digital repository may seem thrilling, but most researchers can meaningfully use only a few hundred search results. The need for efficiency is why AA often supplies only the first two hundred results for a search and limits the total number of hits to fifty thousand.²⁹ Search engine developers have to balance the volume of documents recalled with the precision of the documents in a list of results,³⁰ even at the cost of excluding those who could only express their voices through asterisks.

Costly Searches: The Politics of Parsing African American Print Culture

The erasure of asterisks in search engines is not inevitable. In the jargon of information retrieval, allowing a user to search for a lone asterisk is "costly." Forbidding such costly searches makes sense for a few reasons. Not only does it bring up far more hits than anyone would want, it also consumes significant amounts of computer memory. Memory used for a single search may be trivial, but it encumbers a system scaled up for hundreds, thousands, or millions of users. These searches also run on physical computers. Even leaving aside the energy efficiencies and fossil fuels involved in the physical computers, these searches slow down the system for other users. A system bogged down by a lone wildcard search starts to lag, cascading the distributed rhythms of search-assisted

research. Finally, along with the material concerns of memory and lags, search technology today is overwhelmingly led by profit-seeking companies who make staggering fortunes by selling intelligence about what appears in their indices to advertisers. Information about how and what users find in Google's index is lucrative, spawning an entire industry of search engine optimization for exactly that purpose. Given that a wildcard search effectively requests all the information stored in a given index, a wildcard search theoretically would provide access to silicon bonanzas. Even though Google and Facebook, among others, use indexing and searching tools different from those used by Lucene, they still share Lucene's distaste for the lone asterisk.

Parsing away the asterisk is an unintended result of resource allocation and neglect of marginalized voices in historical print. Companies and groups that erase the asterisk control most of the culture industry in the early twenty-first century: Google, Twitter, Netflix, iTunes, Facebook, and Amazon. The same is true for many of the central resources for American literary history: Accessible Archives, JSTOR, the MLA Bibliography on EBSCOhost, and Gale Cengage's 19th Century U.S. Newspapers.³¹ Gale Cengage at least offers an error message for the lone asterisk that speaks as a comment on Mary Ann Shadd's own difficulties: "Your search term has too few leading characters." While all these sites deny the validity of Shadd's asterisk, some do parse the single character as a synonym for the term "asterisk." This synonym parsing is active on Wikipedia, Yahoo, and Bing, among others. Still others generously allow the lone asterisk to serve as a wildcard search, pulling up every document in the index on such sites as Chronicling America, the Digital Public Library of America, Umbra, Readex, and the Text Creation Partnership's release of the EBBO, ECCO, and EVANS texts. While the synonyms or full-access wildcards are not exactly the same as the QueryParser's erasure of Mary Ann Shadd's asterisk, they still encode the * as an operator and not a term. The asterisk remains unsearchable. No one is going to abandon search tomorrow, but the erasure of Shadd's asterisks may point to a broader project. Just as African American emigrants deliberated the terms of their culture in Canada West in the 1850s through the *Provincial Freeman*, so might we too learn a great deal from Mary Ann Shadd in this the adolescence of our digitized print.

Notes

1. See, for example, recent critical attention to the construction of digital periodical resources: Benjamin Fagan, "Chronicling White America," *American Periodicals: A Journal*

of *History & Criticism* 26, no. 1 (2016): 10–13; James Mussell, *The Nineteenth-Century Press in the Digital Age* (New York: Palgrave Macmillan, 2012); Paul Fyfe, “An Archaeology of Victorian Newspapers,” *Victorian Periodicals Review* 49, no. 4 (2016): 546–77; Ryan Cordell, “‘Q i-jtb the Raven’: Taking Dirty OCR Seriously,” *Book History* 20 (2017): 188–225. Those writings are joined by efforts to develop countervailing digital projects such as Recovering the US Hispanic Literary Heritage project, led by Nicolás Kanellos and Carolina Villarroel, and the Digital Colored American by Eurie Dahn and Brian Sweeney.

2. Ryan Cordell et al., “Forum: Digital Approaches to Periodical Studies,” *American Periodicals: A Journal of History and Criticism* 26, no. 1 (2016): 1–24; “Editing Modernism in Canada,” *Editing Modernism in Canada*, editingmodernism.ca, accessed April 1, 2016; “Modernist Journals Project,” *Modernist Journals Project*, <http://www.modjourn.org/>, accessed April 1, 2016.

3. “eMOP,” *Early Modern OCR Project*, <http://emop.tamu.edu>, accessed April 15, 2016; “Home—Trove,” *Trove—National Library of Australia*, <http://trove.nla.gov.au/>, accessed April 15, 2016; “Text Creation Partnership,” *Text Creation Partnership*, textcreationpartnership.org/, accessed April 15, 2016.

4. Lisa Gitelman, *Paper Knowledge: Toward a Media History of Documents* (Durham, NC: Duke University Press, 2014).

5. Max Kemman, Martijn Kleppe, and Stef Scagliola, “Just Google It—Digital Research Practices of Humanities Scholars,” ArXiv e-print, September 10, 2013, <http://arxiv.org/abs/1309.2434>.

6. Margaret Cohen, *The Sentimental Education of the Novel* (Princeton, NJ: Princeton University Press, 1999), 23.

7. The *Voice of the Fugitive* is available online through OurOntario’s Our Digital World Newspaper Collection at <http://ink.ourdigitalworld.org/vf>. The *Voice* is another title in the antebellum black press that deserves much greater attention.

8. Throughout this essay my accounts of Mary Ann Shadd’s life, work, and importance draw heavily on the excellent biography by Jane Rhodes, *Mary Ann Shadd Cary: The Black Press and Protest in the Nineteenth Century* (Bloomington: Indiana University Press, 1998). Mary Ann Shadd married Thomas Cary in early 1855. As most of the events described in this chapter fall before her marriage, I refer to her as Mary Ann Shadd. Both versions are common and equally acceptable.

9. Robin W. Winks, *The Blacks in Canada: A History* (Montreal: McGill-Queen’s University Press, 1997).

10. Rhodes, *Mary Ann Shadd Cary*, 73.

11. Carla L. Peterson, *Doers of the Word: African-American Women Speakers and Writers in the North (1830–1880)* (New Brunswick, NJ: Rutgers University Press, 1998), 103.

12. In 1853 few African American editors had more experience or stature than Samuel Ringgold Ward. As Ward wrote in his editorial for the sole issue of the *Aliened American* in 1853, “I am requested to edit one paper, to act as corresponding editor of another, and, I am a regular contributor to a third” (April 9, 1853, 2). Along with the

Freeman and the *American*, both started in 1853, Ward had helped edit the *True American* (1846–48); the *Northern Star* and *Colored Farmer* (ca. 1846–49); and the *Impartial Citizen* (1849–53).

13. In the copies that I examined through Accessible Archives and OurOntario, the digital reproduction from microfilm made this section impossible to read.

14. Colored National Convention, *Report of the Proceedings of the Colored National Convention Held at Cleveland, Ohio, on Wednesday, September 6, 1848* (Rochester, NY: Printed by John Dick, at the North Star Office, 1848), <http://udspace.udel.edu/handle/19716/17046>; *Report of the Woman's Rights Convention, Held at Seneca Falls, N.Y., July 19th & 20th, 1848* (Rochester, NY: Printed by John Dick, at the North Star Office, 1848), <https://www.nps.gov/wori/learn/historyculture/report-of-the-womans-rights-convention.htm>.

15. See Rhodes's chapter "We Have 'Broken the Editorial Ice'" in *Mary Ann Shadd Cary* along with 24on27.

16. James Batchelder was a notorious US Marshal with a reputation for kidnaping black men and women in northern cities to be sold back into slavery. Batchelder died in 1854 after being stabbed while trying to keep a formerly enslaved man named Anthony Burns imprisoned in Boston.

17. James W. C. Pennington was a prolific speaker, historian, editor, minister, abolitionist, and Colored Conventions leader. Born enslaved, he emancipated himself and became a prominent minister in the New York and Connecticut areas. Pennington traveled widely in Europe, writing an 1849 book, *The Fugitive Blacksmith*, and receiving an honorary doctorate from the University of Heidelberg.

18. Mary Ann Shadd, "A Bazaar in Toronto for 'Frederick Douglass' Paper,' &c.," *Provincial Freeman*, June 3, 1854.

19. Marcy J. Dinius, "'Look!! Look!!! At This!!!!': The Radical Typography of David Walker's 'Appeal,'" *PMLA* 126, no. 1 (2011): 55–72.

20. Michel-Rolph Trouillot, *Silencing the Past: Power and the Production of History* (Boston: Beacon Press, 2012), 96, 115.

21. Richard Lawrence, "Lucene Frequently Asked Questions," Jakarta Project, February 2, 2002, <https://web.archive.org/web/20020204075100/http://www.lucene.com:80/cgi-bin/faq/faqmanager.cgi?file=chapter.general&toc=faq#q1>.

22. J. D. Thomas, "The Double-Keyed Full Text Difference," *Accessible Archives Blog*, December 2, 2012; Wesley Raabe, "Harriet Beecher Stowe's *Uncle Tom's Cabin*: An Electronic Edition of the *National Era* Version" (PhD diss., University of Virginia, 2006).

23. Note that the StandardAnalyzer works for English-language terms. The documentation for Lucene 6.0.0 (released April 2016) lists thirty-five language-specific analyzers. http://lucene.apache.org/core/6_0_0/analyzers-common/index.html.

24. Manu Konchady, *Building Search Applications: Lucene, LingPipe, and Gate* (Oakton, VA: Mustru Publishing, 2008), 29.

25. User manual, Accessible Archives, n.p.

26. The foundational expression of this idea is in Martin F. Porter, "An Algorithm for Suffix Stripping," *Program* 14, no. 3 (1980): 130–37.

27. One of the notable contributions of the Google Search engine is the use of a single box for all operations. Compare the single Google box to the user interface on AA, which allows at least a handful of boxes, buttons, and menus. There is no perfect design, as developers must balance adding advanced features with the burden of learning how to use those features.

28. Although Lucene is a self-contained Java library, it needs to be integrated with a server platform for users to be able to access it. AA currently uses Apache Tomcat, an open source web server. The web server does not have any direct bearing on the status of asterisks, so I omit any discussion of it here.

29. User manual, Accessible Archives, n.p.

30. Amit Singhal, "Modern Information Retrieval: A Brief Overview," *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 24, no. 4 (2001): 35–42.

31. Because the MLA Bibliography disregards the asterisk, the character * in the title of this essay will be unsearchable.

