Algorithms of Oppression

How Search Engines Reinforce Racism

Safiya Umoja Noble

NEW YORK UNIVERSITY PRESS
New York

5

The Future of Knowledge in the Public

Student protests on college campuses have led to calls for increased support of students of color, but one particular request became a matter of national policy that led to a threat to the Library of Congress's budget in the summer of 2016. In February 2014, a coalition of students at Dartmouth College put forward "The Plan for Dartmouth's Freedom Budget: Items for Transformative Justice at Dartmouth" (the "Freedom Plan"),[1] which included a line item to "ban the use of 'illegal aliens,' 'illegal immigrants,' 'wetback,' and any racially charged term on Dartmouth-sanctioned programming materials and locations." The plan also demanded that "the library search catalog system shall use undocumented instead of 'illegal' in reference to immigrants." Lisa Peet, reporting for *Library Journal*, noted,

> The replacement of the subject heading was the culmination of a two-year grassroots process that began when Melissa Padilla, class of 2016, first noticed what she felt were inappropriate search terms while researching a paper on undocumented students at Dartmouth's Baker-Berry Library in 2013. While working with research and instruction services librarian Jill Baron, Padilla told LJ [*Library Journal*], she realized that nearly every article or book she looked at was categorized with the subject heading "Illegal aliens."[2]

The Dartmouth College librarians became deeply engaged in petitioning the Library of Congress. According to Peet, "Baron, DeSantis, and research and instruction services librarian Amy Witzel proposed that the students gather documentation to prove that 'Illegal aliens' is not a preferred term, and to find evidence that better terms—such as 'Undocumented immigrant,' which was their initial suggestion for a replacement—were in common use. At that point news organizations such as the Associated Press, *USA Today*, ABC, the *Chicago Tribune*,

and the *LA Times* had already committed not to use the term 'Illegal' to describe an individual."[3] Though unsuccessful in 2015, the librarians' case to the Library of Congress had gained traction, and the librarian and professor Tina Gross at St. Cloud State University began organizing caucuses and committees in the American Libraries Association, including the subject analysis committee, social responsibilities round table, and REFORMA, which advocates for library services for Latinos and those who speak Spanish. Social media campaigns ensued, organized under the Twitter hashtags #DropTheWord and #NoHumanBeingIsIllegal.[4] By March 29, 2016, Dartmouth College's student-led organization the Coalition for Immigration Reform, Equality (CoFired) and DREAMers announced in a press release that after a two-year battle, in partnership with campus librarians and the American Libraries Association, "the Library of Congress will replace the term 'illegal aliens' with 'noncitizens' and 'unauthorized immigrants' in its subject headings."[5]

"Illegal Alien" Revisited

The struggle over reclassifying undocumented immigrants was part of a long history of naming members of society as problem people. In many ways, this effort to eliminate "illegal alien" was similar to the ways that Jewish people were once classified by the Library of Congress as the "Jewish question," later to be reclassified in 1984 as "Jews," and Asian Americans were once classified as the "Yellow Peril."[6] Control over identity is political and often a matter of public policy. Almost as soon as the successful change was approved, House Republicans introduced HR 4926 on April 13, 2016, also known as the "Stopping Partisan Policy at the Library of Congress Act," sponsored by Rep. Diane Black (R-TN). In essence, the bill threatened the Library's budget, and Black suggested that the effort to change the Library of Congress Subject Headings (LCSH) was a matter of "caving to the whims of left-wing special interests and attempting to mask the grave threat that illegal immigration poses to our economy, our national security, and our sovereignty."[7]

The battle over how people are conceptualized and represented is ongoing and extends beyond the boundaries of institutions such as the Library of Congress or corporations such as Alphabet, which owns and manages Google Search. Jonathan Furner, a professor of information

studies at UCLA, suggests that information institutions and systems, which I argue extend from state-supported organizations such as the Library of Congress to the Internet, are participating in "legitimizing the ideology of dominant groups" to the detriment of people of color.[8] His case study of the Dewey Decimal Classification (DDC) system, for example, underscores the problematic conceptualizations of race and culture and efforts to "deracialize" the library and classification schemes.[9] Furner offers several strategies for thinking about how to address these issues, using critical race theory as the guiding theoretical and methodological model. I believe these strategies are of great value to thinking about the information studies issues at hand in this research:

- admission on the part of designers that bias in classification schemes exists, and indeed is an inevitable result of the ways in which they are currently structured;
- recognition that adherence to a policy of neutrality will contribute little to eradication of that bias and indeed can only extend its life; [and]
- construction, collection and analysis of narrative expressions of the feelings, thoughts, and beliefs of classification-scheme users who identify with particular racially-defined populations.[10]

While the web-indexing process is not the same as classification systems such as DDC, the application of the theoretical model is still valid for thinking about conceptualizing algorithms and indexing models that could actively intervene in the default normativity of racism and sexism in information resources.

Problems in Classifying People

The idea of classification as a social construct is not new. A. C. Foskett suggests that classificationists are the products of their times.[11] In the work of Nicholas Hudson of the University of British Columbia on the origins of racial classification in the eighteenth century, he suggests that during the Enlightenment, Europeans began to construct "imagined communities," citing Benedict Anderson's term.[12] He says, "This mental image of a community of like-minded individuals, sharing a 'general

will' or a common national 'soul,' was made possible by the expansion of print-culture, which stabilized national languages and gave wide access to a common literary tradition."[13] Classification systems, then, are part of the scientific approach to understanding people and societies, and they hold the power biases of those who are able to propagate such systems. The invention of print culture accelerated the need for information classification schemes, which were often developing in tandem with the expansion of popular, scholarly, and scientific works.[14] Traces of previous works defining the scientific classification of native peoples as "savage" and claims about Europeans as the "superior race," based on prior notions of peoples and nations, began to emerge and be codified in the eighteenth century. Extensive histories have been written of how racial classification emerged in the eighteenth and nineteenth centuries in North America as a paradigm of differentiation that would support the exclusion of native and African people from social and political life.

By the nineteenth century, the processes involved in the development of racial classification marked biological rather than cultural difference and were codified to legally deny rights to property ownership and citizenship. These historical practices undergird the formation of racial classification, which is both assumed and legitimated in classification systems. Without an examination of the historical forces at play in the development of such systems, the replication and codification of people of African descent into the margins goes uncritically examined. This process can be seen in knowledge organization that both privileges and subordinates through information hierarchies such as catalogs and classification systems. The field of library science has been implicated in the organization of people and critiqued for practices that perpetuate power by privileging some sectors of society at the expense of others.

Traditional library and information science (LIS) organization systems such as subject cataloging and classification are an important part of understanding the landscape of how information science has inherited and continues biased practices in current system designs, especially on the web.

Opportunities abound for the interdisciplinarity of LIS to extend more deeply into cultural and feminist studies, because these social science fields provide powerful and important social context for information about people that can help frame how that information is organized

and made available. To date, much of the attention to information organization, storage, and retrieval processes has been influenced and, more importantly, funded by scientific research needs stemming from World War II and the Cold War.[15] The adoption of critical race theory as a stance in the field would mean examining the beliefs about the neutrality and objectivity of the entire field of LIS and moving toward undoing racist classification and knowledge-management practices. Such a stance would be a major contribution that could have impact on the development of new approaches to organizing and accessing knowledge about marginalized groups.

If the information-retrieval priority of making access to recorded information efficient and expedient is the guiding process in the development of technical systems, from databases to web search engines, then what are the distinguishing data markers that define information about racialized people and women in the United States? What have primarily been missing from the field of information science, and to a lesser degree library science, are the issues of representation that are most often researched in the fields of African American studies, gender studies, communications, and increasingly digital media studies. Information organization is a matter of sociopolitical and historical processes that serve particular interests.

A Short History of Misrepresentation in Classifying People

In order to understand how racial and gender representations in Google Search express the same traditional bias that exists in other organizational systems, an overview of how women and non-Whites have been historically represented in information categorization environments is in order. The issue of misrepresentations of women and people of color in classification systems has been significantly critiqued.[16] Hope A. Olson, an associate dean and professor at the School of Information Studies at the University of Wisconsin, Milwaukee, has contributed among the most important theories on the social construction of classification that many of us in the field assign to our students as a way of fostering greater awareness about the power that library, museum, and information professionals hold. Those who have the power to design systems—classification or technical—hold the ability to prioritize

hierarchical schemes that privilege certain types of information over others. An example of these biases include the cataloging of people as subjects in the Library of Congress Subject Headings (LCSH), which serve as a foundational and authoritative framework for categorizing information in libraries in the United States. The LCSH have been noted to be fraught with bias, and the radical librarian Sanford Berman details the ways that this bias has reflected Western perspectives:

> Since the first edition of *Library of Congress Subject Headings* appeared 60 years ago, American and other libraries have increasingly relied on this list as the chief authority—if not the sole basis—for subject cataloging. There can be no quarrel about the practical necessity for such labor-saving, worry-reducing work, nor—abstractly—about its value as a global standardizing agent, as a means for achieving some uniformity in an area that would otherwise be chaotic. . . . But in the realm of headings that deal with people and cultures—in short, with humanity—the LC list can only "satisfy" parochial, jingoistic Europeans and North Americans, white-hued, at least nominally Christian (and preferably Protestant) in faith, comfortably situated in the middle- and higher-income brackets, largely domiciled in suburbia, fundamentally loyal to the Established Order, and heavily imbued with the transcendent, incomparable glory of Western Civilization.[17]

Eventually the LCSH abolished labels such as "Yellow Peril" and "Jewish Question" or made substitutions in the catalog, changing "Race Question" or "Negroes" to "Race Relations" and "Afro-Americans,"[18] but the establishment of such headings and the subsequent decade-long struggles to undo them underscored Berman's point about Western racial bias. (In fact, it was Berman who led the field in calling for antiracist interventions into library catalogs in the 1970s.) Patriarchy, like racism, has been the fundamental organizing point of view in the LCSH. The ways in which women were often categorized was not much better, with headings such as "Women as Accountants" in lieu of the now-preferred "Women Accountants"; women were consistently an aberration to the assumed maleness of a subject area.[19]

Furthermore, efforts at self-identity from the perspective of marginalized and oppressed groups such as the Roma or Romanies cannot es-

cape the stigmatizing categorization of their culture as "Gypsies," even though their "see also" designation to "rogues and vagabonds" was finally dropped from the LCSH.[20] A host of other problematic naming conventions including "Oriental" instead of "Asian" and the location of Christianity at the top of the religious hierarchy, with all deviations moving toward the classification of "Primitive," suggests that there is still work to be done in properly addressing and classifying groups of people around identity.[21] Olson says, "the problem of bias in classification can be linked to the nature of classification as a social construct. It reflects the same biases as the culture that creates it."[22] These types of biases are often seen in offline information practices where conquest is a means of erasing the history of one dynasty or culture by the subsequent regime.[23] Olson's research has already shown that classifications reflect the philosophical and ideological presumptions of dominant cultures over subordinate cultures or groups. For example, in traditional Dewey Decimal Classification (DDC), over 80% of its religion section is devoted exclusively to Christianity, even though there are greater numbers of other religious texts and literature.[24] Olson points to the Library of Congress Classification (LCC) and its biases toward North American and European countries in volumes on the law, with far fewer allocations of space for Asia, Eurasia, Africa, the Pacific area, and Antarctica, reflecting the discourse of the powerful and the presumption of marginality for all others.[25]

In this respect, Olson reminds us that the ordering of information provided in classification schemes "tends to reflect the most mainstream version of these relationships" because "classificatory structures are developed by the most powerful discourses in a society. The result is the marginalization of concepts outside the mainstream."[26] In other words, the most mainstream (e.g., White, heterosexual, Christian, middle-class) controlling regimes in society will privilege themselves and diminish or subdue all others in the organization of what constitutes legitimate knowledge. When we inherit privilege, it is based on a massive knowledge regime that foregrounds the structural inequalities of the past, buttressed by vast stores of texts, images, and sounds saved in archives, museums, and libraries. Certainly, classification systems have some boundaries and limits, as they are often defined in whole by what is included and what is excluded.[27] In the case of most library databases in the United States, Eurocentrism will dominate the canons of knowledge. Knowledge management reflects the same social biases that exist in society, because human beings are at the epicenter of information curation. These practices of the past are part of the present, and only committed and protracted investments in repairing knowledge stores to reflect and recenter all communities can cause a shift toward equality and inclusion in the future. This includes reconciling our brutal past rather than obscuring or minimizing it. In this way, we have yet to fully confront our histories and reconstitute libraries and museums toward reconciliation and reparation.

Search engines, like other databases of information, are equally bounded, limited to providing only information based on what is indexed within the network. Who has access to provide information in the network certainly impacts whether information can be found and surfaced to anyone looking for it. Olson's research points to the ways that some discourses are represented with more power, even if their social classifications are relatively small:

> In North American society, taking away women, African Americans, Hispanic Americans, French Canadians, Native peoples, Asian Americans, lesbians and gay men, people with disabilities, anyone who is not Christian, working class and poor people, and so forth, one is left with a very small "core." An image that shows the complexity of these overlapping categories is that of a huge Venn diagram with many sets limited by Boolean ANDs. The white AND male AND straight AND European AND Christian AND middle-class AND able-bodied AND Anglo mainstream becomes a very small minority . . . , and each set implies what it is not. The implication of this image is that not every person, not every discourse, not every concept, has equal weight. Some discourses simply wield more power than others.[28]

Arguably, if education is based in evidence-based research, and knowledge is a means of liberation in society, then the types of knowledge that widely circulate provide a crucial site of investigation. How oppressed people are represented, or misrepresented, is an important element of engaging in efforts to bring about social, political, and economic justice.

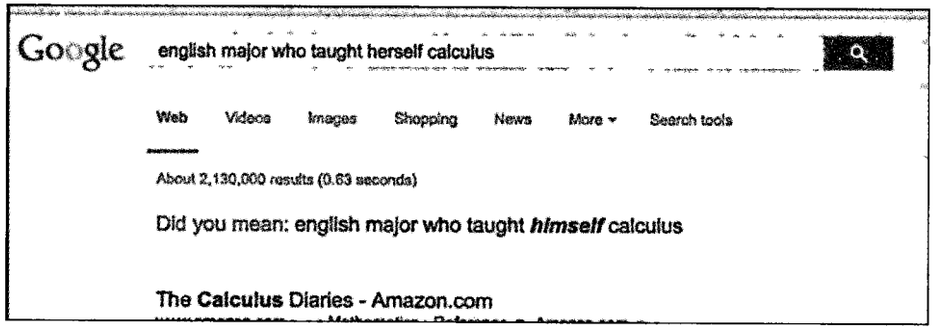


Figure 5.1. Google autocorrects to "himself" rather than "herself." Search sent to me by a colleague, June 16, 2016.

We have to ask ourselves what it means in practical terms to search for concepts about gender, race, and ethnicity only to find information lacking or misrepresentative, whether in the library database or on the open web. Olson's notion that cultural metaphor is the basis of the construction of classification systems means these cultural metaphors are profoundly represented in the notions of the "Jewish Question" or the "Race Question." These subject headings suggest both an answer and a point of view from which the problems of Jews and race are presupposed. Simply put, to phrase "Jewish" or "race" as a question or problem to be answered suggests a point of view on the part of the cataloger that is quite different from how a Jewish person or a racialized person might frame themselves. It is here that the context and point of view of library and information science professionals who are responsible for framing people and communities as "problems" and "questions" is important. By examining the ways that Black people specifically have been constructed in the knowledge schemes, the African American studies professor and philosopher Cornel West aptly describes the positionality of how this community is depicted in the West:

> Black people as a problem-people rather than people with problems; black people as abstractions and objects rather than individuals and persons; black and white worlds divided by a thick wall (or a "Veil") . . . ; black rage, anger, and fury concealed in order to assuage white fear and anxiety; and black people rootless and homeless on a perennial journey to discover who they are in a society content to see blacks remain the permanent underdog.[29]

The library scholar Joan K. Marshall points to the way this idea was expressed in the Library of Congress when "N*ggers" was a legitimate subject category, reflecting the "social backgrounds and intellectual levels" of users, concretizing oppressive race relations.[30] Difference, in the case of the Library of Congress, is in direct relation to Whiteness as the norm. No one has made this more clear than Berman, whose groundbreaking work on Library of Congress Subject Headings has forever changed the field. He notes that in the case of both Jews and the representations of race, these depictions are not without social context:

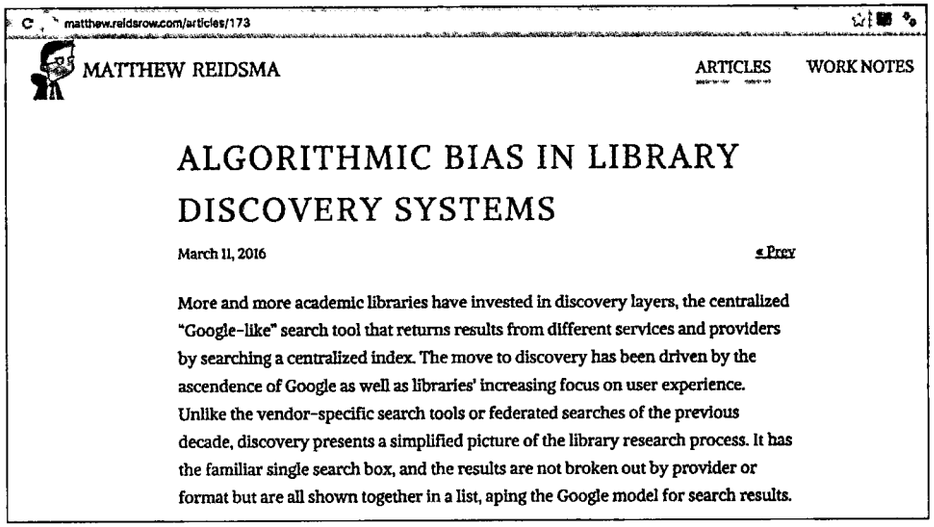
> For the image of the Jew to arouse any feelings, pro or con, he [*sic*] had to be generalized, abstracted, depersonalized. It is always possible for the personal, individual case to contradict a general assertion by providing living, concrete proof to the contrary. For the Jews to become foils of a mass movement, they had to be converted into objectified symbols so as to become other than human beings.[31]

In the case of Google, because it is a commercial enterprise, the discussions about its similar information practices are situated under the auspices of free speech and protected corporate speech, rather than being posited as an information resource that is working in the public domain, much like a library. An alternative possibility could be that corporate free speech in the interests of advertisers could be reprioritized against the harm that sexist and racist speech on the Internet could have on those who are harmed by it. This is the value of using critical race theory—considering that free speech may in fact not be a neutral notion but, rather, a conception that when implemented in particular ways silences many people in the interests of a few.

The disclaimer by Google for the problem of searching for the word "Jew" leading to White supremacist, Holocaust-denial web results is surprisingly similar to the construction of Jewish identity in the LCSH. Both systems reflect the nature of the relationship between Jewish and non-Jewish Europeans and North Americans. This is no surprise, given that hyperlinking and indexing are directly derived from library science

citation-analysis practices. This linkage between the indexing practices of the World Wide Web and the traditional classification systems of knowledge structures such as the Library of Congress is important. Both systems rely on human decisions, whether given over en masse to artificial intelligence and algorithms or left to human beings to catalog. The representation of people and cultures in information systems clearly reflects the social context within which the subjects exist. In the case of search engines, not unlike cataloging systems, the social context and histories of exploitation or objectification are not taken into explicit consideration—rather, they are disavowed. What can be retrieved by information seekers is mediated by the technological system—be it a catalog or an index of web pages—by the system design that otherizes. In the case of the web, old cataloging and bibliometric practices are brought into the modern systems design.

Library science scholars know that bibliographic and naming controls are central to making knowledge discoverable.[32] Part of the issue is trying to understand who the audience is for knowledge and naming and organizing information in ways that can be discovered by the public. Berman cites Joan Marshall's critiques of the underlying philosophy of the Library of Congress's subject-cataloging practices and the ways that they constitute an audience through organizational bias, wherein a "majority reader" is established as a norm and, in the case of the Library of Congress, is often "white, Christian (usually Protestant) and male."[33] Indeed, these scholars are taking note of the influence that categorization systems have on knowledge organization and access. What is particularly important in the interrogation of these marginalizing information-management systems is Berman's reference to the Algerian psychologist Franz Fanon's articulation of the mechanics of cultural "brain washing" that occurs through racist cataloging practices.[34] Berman underscores that the problems of racial representation and racism are deeply connected to words and images and that a racist worldview is embedded in cataloging practices that serve to bolster the image and domination of Western values and people (i.e., White, European, and North Americans over people of African descent). The library practitioner Matthew Reidsma gave a recent gift to the profession when he blogged about library discovery systems, or search interfaces, that are just as troubled as commercial interfaces. In his blog post, he details the limitations of


matthew.reidsrow.com/articles/173

MATTHEW REIDSMA ARTICLES WORK NOTES

ALGORITHMIC BIAS IN LIBRARY DISCOVERY SYSTEMS

March 11, 2016 « Prev

More and more academic libraries have invested in discovery layers, the centralized "Google-like" search tool that returns results from different services and providers by searching a centralized index. The move to discovery has been driven by the ascendence of Google as well as libraries' increasing focus on user experience. Unlike the vendor-specific search tools or federated searches of the previous decade, discovery presents a simplified picture of the library research process. It has the familiar single search box, and the results are not broken out by provider or format but are all shown together in a list, aping the Google model for search results.

Figure 5.2. A call to the profession to address algorithmic bias in library discovery systems by Matthew Reidsma attempts to influence the field of information studies. Source: Reidsma, 2016.

databases, the kinds of gender biases that are present in discovery tools, and how little innovation has been brought to bear in resolving some of the contradictions we know about.[35]

I sought to test the call that Reidsma made to the profession to interrogate library information management tools by conducting searches in a key library database. I looked in the largest library image database available to academic libraries, ArtStor, and found troublesome practices of metadata management there too. Undoubtedly, these kinds of cataloging stances can be evaluated in the context of the field of library science, which is largely averse to teaching and talking about race and the White racial gaze on information. I have published several articles with colleagues about the challenges of teaching about race in the library and information studies classroom and the importance of integrating theory and training of information workers around issues of social justice in the profession. I interpret these kinds of cataloging mishaps as a result of the investment of the profession in colorblind ideology. Unable and unequipped to think through the complexities of systems of racialization, the profession writ large struggles to find frameworks to

Figure 5.3. Search in ArtStor for "black history" features the work of a series of European and White American artists, March 2, 2016. The first result is work by Thomas Waterman Wood.

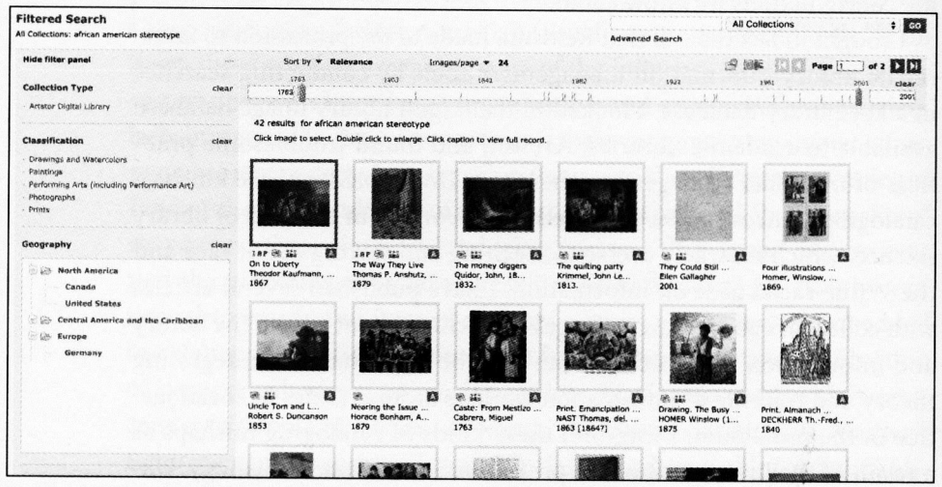


Figure 5.4. *On to Liberty*, an oil on canvas by Theodore Kauffman, a German painter, is the first item under "African American stereotype."

Figure 5.5. A satirical piece by the artist Damali Ayo and her online piece *Rent-A-Negro*, which is a critique of liberal racial ideologies that tokenize African Americans. The work is cataloged as "racism."

think critically about the long-term consequences of misidentification of people and, in this case, concepts about works of art.

Search as a Source of Reality

Indeed, problematic results in ArtStor are just a small window into a long and troubled history of misrepresentation in library subject cataloging and classification systems, which are faithful reflections of the problematic representations in mainstream U.S. culture. Our ability to recognize these challenges can be enhanced by asking questions about how technological practices are embedded with values, which often obscure the social realities within which representations are formed. The interface of the search engine as a mechanism for accessing the Internet is not immune, nor impartial, to the concerns of embedded value systems. Search is also more than the specific mathematical algorithms and deep-machine learning developed by computer scientists and software engineers to index upward of a trillion pages of information and move some from the universal data pile to the first page of results on a computer screen. The interface on the screen presents an information reality,

while the operations are rendered increasingly invisible.[36] The media and communications scholar Alex Galloway destabilizes the idea that digital technologies are transparent, benign windows or doors providing a view or path to somewhere and in themselves insignificant—the digital interface is a material reality structuring a discourse, embedded with historical relations, working often under the auspices of ludic capitalism, where a kind of playful engagement of labor is masked in vital digital media platforms such as Google.[37] Search does not merely present pages but structures knowledge, and the results retrieved in a commercial search engine create their own particular material reality. Ranking is itself information that also reflects the political, social, and cultural values of the society that search engine companies operate within, a notion that is often obscured in traditional information science studies.

Further, new digital technologies may constitute containers for old media discourses, and the web interface (such as a plain Google search box) is a transitional format from previous media forms.[38] Certainly in the case of digital technology such as commercial search engines, the interface converges with the media itself. Commercial search, in the case of Google, is not simply a harmless portal or gateway; it is in fact a creation or expression of commercial processes that are deeply rooted in social and historical production and organization processes. John Battelle, who has carefully traced the history of Google, describes search as the product of our needs and desires, aggregated by companies:

> Link by link, click by click, search is building possibly the most lasting, ponderous, and significant cultural artifact in the history of humankind: the Database of Intentions. The Database of Intentions is simply this: the aggregate results of every search ever entered, every result list ever tendered, and every path taken as a result. . . . This information represents the real-time history of post-Web culture—a massive clickstream database of desires, needs, wants, and preferences that can be discovered, subpoenaed, archived, tracked and exploited for all sorts of ends.[39]

Undoubtedly, search is also pivotal in the development of artificial intelligence. In many ways, Google Search is an attempt to use computer science as a basis for sorting and making decisions about the relevance and quality of information rather than human sorting and web-indexing practices—practices that search engine companies such as Yahoo! and those of the past invested in heavily and that were both expensive to implement and limited and less responsive in real time.[40]

Providing Context for Information about People

In a narrow sense, information is a series of signals and messages that can be expressed through mathematics, algorithms, and statistical probabilities. In a broader sense, however, Tefko Saracevic, a professor emeritus of information science at Rutgers, suggests that information is constituted through "cognitive processing and understanding."[41] There is a pivotal relationship between information and users that is dependent on human understanding. It is this point that I want to emphasize in the context of information retrieval: information provided to a user is deeply contextualized and stands within a frame of reference. For this reason, it is important to study the social context of those who are organizing information and the potential impacts of the judgments inherent in informational organization processes. Information must be treated in a context; "it involves motivation or intentionality, and therefore it is connected to the expansive social context or horizon, such as culture, work, or problem-at-hand," and this is fundamental to the origins of information science and to information retrieval.[42] Information retrieval as a practice has become a highly commercialized industry, predicated on federally funded experiments and research initiatives, leading to the formation of profitable ventures such as Yahoo! and Google, and a focus on information relevance continues to be of importance to the field. Information science is essentially deeply entwined with the history of library science and has primarily been concerned with the collection, storage and retrieval, and access to and use of information. Saracevic notes that "the domain of information science is the transmission of the universe of human knowledge in recorded form, centering on manipulation (representation, organization, and retrieval) of information, rather than knowing information."[43] This foregrounds the ways that representations in search engines are decontextualized in one specific type of information-retrieval process, particularly for groups whose images, identities, and social histories are framed through forms of systemic domination. Although there is a long, broad, and historical context for

addressing categorizations, the impact of learning from these traditions has not yet been fully realized.[44]

Attention to "the universe of human knowledge" is suggestive for contextualizing information-retrieval practices this way, leading to inquiries into the ways current information-retrieval practices on the web, via commercial search engines, make some types of information available and suppress others. The present focus on the types of information presented in identity-based searches shows that they are removed from the social context of the historical representations and struggles over disempowering forms of representation. These critiques have been levied toward other media practices such as television and print culture. Whether human beings believe that the information delivered in search is relevant has consistently been the basis of judgment about information quality,[45] but what is underdiscussed is that retrieval of information in commercial platforms such as web-based search engines is not unique to the individual searcher. A web-based commercial search engine does not entirely "know" who a user is, and it is not customizing everything to our personal and political tastes, although it is aggregating us to people it thinks are similar to us on the basis of what is known through our digital traces.

Finding Culturally Situated Information on the Web

The field of LIS is significantly engaged in information classification and organization work, which can inform the framework for thinking about developing ICTs that are focused on surfacing prioritized results, such as the search engine. Critical race theory in this process of developing information-organization tools is of great value, particularly when thinking about the phenomenon of excessive recall of documents on the web that are irrelevant or decontextualized. Responses to the kinds of problematic biases in large commercial search engines are part of the growing motivation behind a host of culturally situated search engines that are emerging, particularly Blackbird (www.blackbirdhome.com), a Mozilla Firefox browser designed to help surface content of greater relevance to African Americans. Blackbird has been met with mixed reviews, from support and interest to wholesale rejection.[46] In any case, organizations and individuals are responding to the limits of

traditional commercial search engines through the development of such search engines. Identity-focused websites, a combination of web-based browsers and web directories, are emerging to prioritize the interests of specific communities on the basis of the human-curated practices of the past and can be seen in search engines such as BlackWebPortal (www.blackwebportal.com); GatewayBlackPortal (www.gatewayblack.com), which is based on international models such as JGrab, a Jewish search engine; BlackFind.com (www.blackfind.com); and Blackbird. Sites such as Jewogle (www.jewogle.com), which serves as an online encyclopedia of the accomplishments of Jewish people; Jewish.net (http://jewish.net/), which is used to "search the Jewish Web"; JewGotIt (www.jewgotit.com); and Maven Search (www.maven.co.il), which catalogs over fifteen thousand Jewish websites, have emerged in the hundreds, some tongue in cheek, across religion, culture, and national origin. Much of this is a response of communities that are seeking control over relevant content and representation, as well as access to quality information within racial or group identity.

One of the fundamental challenges for these culturally situated search engines is the way in which they make visible the contradictions and biases in search engines, which André Brock discusses in relationship to Blackbird. He notes that "Blackbird's efforts to foreground African American content were seen as an imposition on the universal appeal of the internet, highlighting the perception of the browser as a social structure limited by Black representation."[47] Brock's work indicates that though there is a demand for culturally relevant Internet browsing that will help surface content of interest to Black people, its value works against norms on the web, making it less desirable.

Reproducing Social Relations through Information Technologies

Online racial disparities cannot be ignored because they are part of the context within which ICTs proliferate, and the Internet is both reproducing social relations and creating new forms of relations based on our engagement with it. Technologies and their design do not dictate racial ideologies; rather, they reflect the current climate. As users engage with technologies such as search engines, they dynamically co-construct content and the technology itself.[48] Online information and content

available in search is also structured systemically by the infusion of advertising revenue and the surveillance of user searches, which the subjects of such practices have very little ability to reshape or reformulate. Lack of attention to the current exploitative nature of online keyword searches only further entrenches the problematic identities in the media for women of color, identities that have been contested since the advent of commercial media such as broadcast, print, and radio. Noticeably absent in the discussions of search is the broader social and technical interplay that exists dynamically in the way technology is increasingly mediating public access to information, from libraries to the search engine.

Now, more than ever, a new conception of information access and quality rooted in historical, economic, and social relations could have a transformational effect on the role and consequences of search engines. It is my goal through this research to ensure that traditionally underrepresented ideas and perspectives are included in the shaping of the field—to surface counternarratives that would allow for a questioning of the normalization of such practices. Rather than prioritize the dominant narratives, Internet search platforms and technology companies could allow for greater expression and serve as a democratizing tool for the public. This is rendered impossible with the current commercial practices.

What we need are public search engine alternatives, united with public-interest journalism and librarianship, to ensure that the public has access to the highest quality information available.

6

The Future of Information Culture

In March 2010, the U.S. Federal Communications Commission (FCC) put forward its ten-year broadband plan, wherein it called for high-speed Internet to become the common medium of communications in the United States.[1] This plan still governs the information landscape. The FCC envisioned that the Internet, as the common medium, would potentially displace current telecommunications and broadcast television systems as the primary communications vehicle in the public sphere. According to the report, "almost two-thirds of the time users spend online is focused on communication, information searching, entertainment or social networking."[2] The plan called for increased support for broadband connectivity to facilitate Americans' ability to access vital information, and it is focused on infrastructure, devices, accessibility, and connectivity. However, the plan made no mention of the role of search engines in the distribution of information to the public, with the exception of noting that the plan itself will be archived and made available in perpetuity in the archives of the Internet. Primary portals to the Internet, whether connecting through the dwindling publicly funded access points in libraries and schools or at home, serve as a gateway to information and cannot be ignored. Access to high-quality information, from journalism to research, is essential to a healthy and viable democracy. As information moves from the public sphere to private control by corporations, a critical juncture in the quality of information available and the public's ability to sift and use it is at stake, as the noted political economist Herbert Schiller forewarned:

> The American economy is now hostage to a relatively small number of giant private companies, with inter-locking connections, that set the national agenda. This power is particularly characteristic of the communication and information sector where the national cultural-media agenda is provided by a very small (and declining) number of integrated private

My hope is that the public will reclaim its institutions and direct our resources in service of a multiracial democracy. Now, more than ever, we need libraries, universities, schools, and information resources that will help bolster and further expand democracy for all, rather than shrink the landscape of participation along racial, religious, and gendered lines. Information circulates in cultural contexts of acceptability. It is not enough to simply want the most accurate and credible information to rise to the top of a search engine, but it is certainly an important step toward impacting the broader culture of information use that helps us make decisions about the distribution of resources among the most powerful and the most disenfranchised members of our society.

In short, we must fight to suspend the circulation of racist and sexist material that is used to erode our civil and human rights. I hope this book provides some steps toward doing so.

NOTES

INTRODUCTION

1. Matsakis, 2017.
2. See Peterson, 2014.
3. This term was coined by Eli Pariser in his book *The Filter Bubble* (2011).
4. See Dewey, 2015.
5. I use phrases such as "the N-word" or "n*gger" rather than explicitly using the spelling of a racial epithet in my scholarship. As a regular practice, I also do not cite or promote non–African American scholars or research that flagrantly uses the racial epithet in lieu of alternative phrasings.
6. See Sweney, 2009.
7. See Boyer, 2015; Craven, 2015.
8. See Noble, 2014.
9. The term "digital footprint," often attributed to Nicholas Negroponte, refers to the online identity traces that are used by digital media platforms to understand the profile of a user. The online interactions are often tracked across a variety of hardware (e.g., mobile phones, computers, internet services) and platforms (e.g., Google's Gmail, Facebook, and various social media) that are on the World Wide Web. Digital traces are often used in the data-mining process to profile users. A digital footprint can often include time, geographic location, and past search results and clicks that have been tracked through websites and advertisements, including cookies that are stored on a device or other hardware.
10. "Kandis" is a pseudonym.
11. See H. Schiller, 1996.

CHAPTER 1. A SOCIETY, SEARCHING

1. See UN Women 2013.
2. See Diaz, 2008; Segev, 2010; Nissenbaum and Introna, 2004
3. See Olson, 1998; Berman, 1971; Wilson, 1968; and Furner, 2007.
4. See Daniels, 2009, 2013; Davis and Gandy, 1999.
5. See Halavais, 2009, 1–2.
6. See Angwin et al., 2016.
7. See O'Neil, 2016, 8.
8. See Levin, 2016.
9. See Kleinman, 2015.

27. A comprehensive timeline of Edward Snowden's whistleblowing on the U.S. government's comprehensive surveillance program is detailed by the *Guardian* newspaper in MacAskill and Dance, 2013.
28. Tippman, 2015.
29. Kiss, 2015.
30. Goode, 2015.
31. See Robertson, 2016.
32. Ibid.

CHAPTER 5. THE FUTURE OF KNOWLEDGE IN THE PUBLIC

1. See the plan at "The Plan for Dartmouth's Freedom Budget: Items for Transformative Justice at Dartmouth," *Dartblog*, accessed August 9, 2017, www.dartblog.com/Dartmouth_Freedom_Budget_Plan.pdf.
2. Peet, 2016.
3. Ibid.
4. Ibid.
5. Qin, 2016.
6. Sanford Berman documents the sordid history of racist classification in the Library of Congress in his canonical work *Prejudices and Antipathies* (1971). A follow-up to his findings was written thirty years later by Steven A. Knowlton in the article "Three Decades since *Prejudices and Antipathies*: A Study of Changes in the Library of Congress Subject Headings" (2005).
7. Peet, 2016.
8. Furner, 2007, 148.
9. Ibid., 147.
10. Ibid., 169.
11. See Olson, 1998.
12. See Anderson, 1991, 37–46.
13. Hudson, 1996, 256; Anderson, 1991.
14. The first documented evidence of print culture is attributed to Chinese woodblock printing. See Hyatt Mayor, 1971, 1–4.
15. See Saracevic, 2009.
16. See Berman, 1971; Olson, 1998.
17. Berman, 1971, 15.
18. Ibid., 5.
19. See ibid.; Palmer and Malone, 2001.
20. See Berman, 1971, 5.
21. Ibid.
22. Olson, 1998, 233.
23. Ibid., 234
24. Ibid., 234–235.
25. Ibid., 235.
26. Ibid.

27. See Cornell, 1992; Olson, 1998.
28. See Olson, 1998, 237.
29. See C. West, 1996, 84.
30. See Berman, 1971, 18.
31. Ibid., citing Mosse, 1966.
32. See Wilson, 1968, 6.
33. Berman, 1971, 19, citing Marshall, personal communication, June 23, 1970.
34. Ibid., 20.
35. Reidsma, 2016.
36. See Galloway, 2008.
37. See Galloway, Lovink, and Thacker, 2008.
38. See Galloway, 2008.
39. Battelle, 2005, 6.
40. See Brin and Page, 1998a.
41. Saracevic, 1999, 1054.
42. Ibid.
43. Saracevic, 2009, 2570.
44. See Bowker and Star, 1999.
45. See Saracevic, 2009.
46. See Brock, 2011.
47. Ibid., 1101.
48. See Fuchs, 2008.

CHAPTER 6. THE FUTURE OF INFORMATION CULTURE

1. See Federal Communications Commission, 2010
2. Ibid.
3. H. Schiller, 1996, 44.
4. Cohen, 2016.
5. See McChesney and Nichols, 2009; H. Schiller, 1996.
6. See Harris-Perry, 2011; hooks, 1992.
7. Arreola, 2010.
8. See the website of the Society of Professional Journalists Code of Ethics, www.spj.org.
9. See Darnton, 2009; Jeanneney, 2007.
10. See Jeanneney, 2007.
11. See *Authors Guild v. Google*, Case 1:05-cv-08136-DC, Document 1088, November 14, 2013.
12. Darnton, 2009, 2.
13. See Search King v. Google, 2003.
14. See Dickinson, 2010.
15. Ibid., 866.
16. Ibid.
17. Ingram, 2011.