

# The difficulty of neutrality

by Karen Coyle

## Introduction

With the increased awareness of the effects of social media on the actual information that people encounter, comes a renewed questioning of the role that libraries play in supporting civil society. Private services like the search engines (represented here with Google), media services (YouTube), and even the non-commercial information source of Wikipedia show various degrees of content bias. It was the United States election cycle of 2016 that demonstrated most plainly that the argument «it's just an algorithm» doesn't guarantee the neutrality of purpose. The counter argument made by many librarians that libraries are neutral was once accepted as true, but more recently has been questioned by some librarians.

## The miseducation of Dylann Roof

On June 17<sup>th</sup>, 2015, a young Southern American man entered the historic Emanuel African Methodist Episcopal Church in Charleston, South Carolina, and was welcomed as the only white person at a prayer meeting. One hour later he produced a gun and killed nine of the worshippers.

The Southern Poverty Law Center (SPLC), a non-profit organization that investigates racist and other hate organizations, did an analysis of Dylan Roof and his road to radical white supremacy<sup>1</sup>. Roof left ample evidence of this in a manifesto in which he described his process of discovery, which, according to the SPLC, took a wrong turn when he typed «black on white violence» into the Google search box. The result of that search turned up a number of extremist sites with misleading information about the threat of African-Americans to members of the white race. In a video, SPLC states that the Google algorithm formed a kind of feedback loop, such that as Roof continued to search for information he was profiled increasingly to see racist literature. The SPLC video then says: «This is a fundamental problem that Google must address if it is truly to be the world's library»<sup>2</sup>.

The SPLC may have significant expertise in law, poverty, and racism, but at least in this video it shows itself to hold some of the same mistaken ideas about Google and about libraries that prevail in the general population. It is this kind of thinking that leads some towns to cut back on library budgets, or cut out their funding alto-

KAREN COYLE, Digital libraries consultant, Berkeley, United States, e-mail [kcoyle@kcoyle.net](mailto:kcoyle@kcoyle.net).

Websites last accessed on: March 20<sup>th</sup>, 2018.

1 Southern Poverty Law Center, *Google and the miseducation of Dylann Roof*, January 18<sup>th</sup>, 2017, <<https://www.splcenter.org/20170118/google-and-miseducation-dylann-roof>>.

2 Southern Poverty Law Center, *The miseducation of Dylann Roof*, <<https://youtu.be/qB6A45tA6mE>>.



gether, because they think that «it's all on the Internet». The actual picture is complex with many facets, from defining what most people mean by the Internet, to the role of search engines and algorithms that select content for users, to the position of libraries (primarily the publicly accessible ones) in the conveyance of culture.

### Algorithms

In 1999 Larry Page and Sergei Brin, while students at Stanford University, developed a new type of search engine that made calculations as to the importance of retrieved items by using weighted values that treated inter-site links much like citations in academic articles<sup>3</sup>. Each link served as a kind of recommendation from one document to another, with heavily cited documents gaining the highest ranks:

PageRank is a global ranking of all web pages, regardless of their content, based solely on their location in the Web's graph structure. Using PageRank, we are able to order search results so that more important and central Web pages are given preference. In experiments, this turns out to provide higher quality search results to users. The intuition behind PageRank is that it uses information which is external to the Web pages themselves - their backlinks, which provide a kind of peer review<sup>4</sup>.

Google of course claims that its algorithms are neutral. It also makes the point that it uses algorithms and is not making editorial decisions. This latter is at least in part in response to US law which affords data carriers immunity from liability if they treat all bits passing through their network equally, without selecting for the content of the message or charging different users different transaction prices. If they do discriminate based on the content, then they become potentially responsible for messages that are illegal, with particular emphasis on intellectual property rights<sup>5</sup>.

We know, however, that Google does make human-based changes to its algorithms. In 1998 Google modified its algorithms to lower the visibility of online pornography<sup>6</sup>. Before that, a search on a surprising number of terms would show results for pornography web sites; this meant that children doing searches for Disney characters like Bambi or toys like Barbie were being directed to pornography in which actors used those names. Google made this change as a business decision, not a political or social one, because they felt that the service would not be acceptable if it became known for presenting pornography to unwitting users.

**3** Sergey Brin; Lawrence Page, *The anatomy of a large-scale hypertextual web search engine*, «Computer networks and ISDN systems», 30 (1998), n. 1-7, p. 107-117.

**4** Lawrence Page [et. al.], *The PageRank citation ranking: bringing order to the web*. Technical Report, Stanford InfoLab, 1999, p. 15, <<http://ilpubs.stanford.edu:8090/422/>>.

**5** The Digital millennium copyright act (DMCA) is a 1998 United States law that sets out the rules governing copyrights when the materials are digital and are accessed over a network. Although primarily focused on copyright, the law provides general exceptions to liability for online service providers who do not themselves provide or manage the content of the files that pass over its network. For this reason, networked providers such as Google, YouTube, Facebook, Twitter, and others are reluctant to admit of editorial practices regarding the content.

**6** Ken Auletta, *Googled: the end of the world as we know it*. New York: Penguin Books, 2010, p. 56.

## NOTE E DISCUSSIONI

In addition, Google makes decisions for business reasons, and these are presented as improvements to the ranking algorithms<sup>7</sup>. Some of this is in response to the growing business of 'search engine optimization', where experts who have studied the effects of PageRank help online sites meet the criteria to appear high in the ranked order. Some of this work can be considered scamming the system by artificially creating the conditions that result in higher ranking such as creating thousands of links to a page to raise its PageRank value. Google claims to make upwards to 500 changes to its ranking algorithm each year, but reveals few details of the algorithm's functioning as this is considered a trade secret<sup>8</sup>.

It's pretty clear that if Google could lower the ranking of pornography it could also lower the ranking of some segment of racist sites. For all we know, however, Google has indeed done this already. Even if many searches using the same terms as found in those sites would not reveal racist sites on the first screen of results, like pornography you could indeed find racism online with the right search terms. Google (and other search engines) wants you to find what you are looking for; its algorithm attempts to manage false hits but if content exists on the Internet it tries to make it findable. It is this quality that brings people back to the search engine time and again, and because the search engine is funded by advertising that is correlated to search terms and results, it is this repeat business that is in Google's financial interest.

Perhaps in the early days of Google's PageRank it was accepted as socially and politically and factually neutral, but the myth of the neutral algorithm has been shattered in a number of studies and thought-pieces that have revealed that algorithms often contain the prejudices of their creators, and always reflect a human decision-making process, regardless of how much mathematical or statistical expertise is applied. In particular, the goals of the algorithm must be taken into account<sup>9</sup>: who benefits from the decisions? What is the desired outcome, either explicit or implicit? Is the outcome of the algorithm predictable? Can it be 'scammed'? Whose values does it reflect?

One particularly difficult area is that of machine-learning. Machines 'learn' from data sets that represent a series of decisions about content. A notorious failure of machine learning took place in 2015 in the Google Photos application. The application did automatic tagging of the content of photographs based on an algorithm that had supposedly undergone extensive learning. Unfortunately, the algorithm tagged a group of photographs of young African-Americans with the tag «gorillas»<sup>10</sup>. Presumably the algorithm's training set had included animals but had not included sufficient examples of African-American faces so that it could learn to recognize them with the same accuracy as white faces. Based on similar problems of training algorithms without a representative sample, some facial recognition programs do

**7** Claire Cain Miller, *Google tweaks algorithm to push down low-quality sites*, «New York Times», February 25<sup>th</sup>, 2011, <<https://bits.blogs.nytimes.com/2011/02/25/google-tweaks-algorithm-to-push-down-low-quality-sites/>>.

**8** Steven Levy, *Exclusive: how Google's algorithm rules the web*, «Wired», February 22<sup>nd</sup>, 2010, <[https://www.wired.com/2010/02/ff\\_google\\_algorithm/all/1/](https://www.wired.com/2010/02/ff_google_algorithm/all/1/)>.

**9** Cathy O'Neil, *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown Publishers, 2016, p. 19-23.

**10** Sara Wachter-Boettcher, *Technically wrong: sexist apps, biased algorithms, and other threats of toxic tech*. New York: W. W. Norton & Company, 2017, p.129.

not accurately recognize non-white faces<sup>11</sup>, and there is a demonstration of a «racist hand dryer» that does not turn on for dark hands<sup>12</sup>. While it seems that text should be less of a problem because less interpretation is need of the contents, search results on texts also reveal serious problems. Author Safiya Noble recounts her experience in looking for ideas for a daughter's playgroup: she typed in «black girls» and was offered an insulting array of highly sexualized sites, sites which differed significantly from those that one would see as results for a search query «white girls». Continuing her discovery, she found that searches for business-appropriate attire showed white, middle-class images almost exclusively, and that African-American images were often found to be tagged as examples of improper clothing choices, even when the models were dressed in suitable attire<sup>13</sup>.

One might assume that library systems do not suffer these same problems. Yet as library systems increase their use of search completion suggestions and make recommendations, the same types of algorithmic embarrassments can be encountered. Matthew Reisma reported on his experience in a commercial library software package that linked to Wikipedia pages based on bibliographic search results. The system suggested the Wikipedia page for «workforce stress» as its recommendation following searches on «women in the workforce»<sup>14</sup>. After reporting this to the vendor, the algorithm was changed.

The number of examples of prejudicial algorithms is very large, and these are only a few recent ones, ones that should be enough to refute any suggestions that algorithms are neutral. They are not only no more neutral than the people who create them, they also are very good at hiding the complexity of their models making it difficult to examine the underlying assumptions that are fed into thousands of calculations over many millions of information resources.

Is neutrality the right question?

Returning to the story about Google's role in the radicalization of Dylann Roof, we have to ask what went wrong? The user typed in a query, and Google retrieved and prioritized sites and documents that corresponded to that query. In this case it does not appear that the algorithm showed prejudice; in fact, one could easily argue that, at least from the user's point of view, the search engine responded perfectly. The actual meaning of the SPLC complaint is that the algorithm did not distinguish between true and false information, nor between socially acceptable and unacceptable messages. Given the number of documents that Google indexes (measured at least in the trillions), making a judgment about truth for the mass of pages would clearly be a difficult undertaking, if possible at all. However, it has been shown that Google is able to monitor the content of pages when required. In countries that regulate the intellectual content that their citizens are allowed to access, Google finds ways to comply with these regulations.

**11** Steve Lohr, *Facial recognition is accurate, if you're a white guy*, «New York Times», February 9<sup>th</sup>, 2018, <<https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html>>.

**12** T4runs, *Racist hand dryer*, <<https://www.youtube.com/watch?v=8E09Xdrvf-E>>.

**13** Safiya Umoja Noble, *Algorithms of oppression: how search engines reinforce racism*. New York: New York University Press, 2018.

**14** Matthew Reidsma, *Auditing algorithms*. In: "Code4lib 2018 Annual Conference" (Washington, DC, February 15<sup>th</sup>, 2018), <<https://mreidsma.github.io/talks/code4lib/>>.

## NOTE E DISCUSSIONI

In the United States free expression is covered by the first amendment, which prohibits the government from regulating speech except in very narrow circumstances. Non-governmental agencies, such as corporations, are not bound by the same prohibition. However, asking a company as powerful as Google to regulate information access would create a single point of control that is in the hands of a company whose primary interest is in satisfying the needs of its advertisers and whose selection process is not transparent. As Siva Vaidhyanathan so succinctly stated, we searchers are not Google's customers, we are the product that Google provides to the companies that advertise on its service<sup>15</sup>.

### The comparison to libraries

The comparison of an Internet search engine to a library should not be viewed as a choice between different search algorithms. The difference begins long before any programming is applied to the available content. Why would a library not have provided the searcher with the racist materials that allegedly transformed Dylann Roof into a white supremacist and a killer? It is not sufficient to say that the library would not have selected those materials for its collection; the library is not the first step in the selection process - that role is performed long before the library selection process by the publishing industry. Libraries purchase and provide access to what can be described as 'regularly published materials', meaning those that are published by a group of known and reliable content providers. Large libraries, especially research libraries, often accept all publications coming from a particular group of publishers, proof that they accept the quality-control role of the publishing industry. It is only rarely that libraries accept self-published material, and that is usually because it has some particular local or scientific interest.

The Internet, on the other hand, is almost entirely a platform for self-publishing, meaning that, at least in many countries, no editorial decisions are made to determine what documents are provided over the network. Google's PageRank algorithm is not the same as editorial oversight because no materials are entirely eliminated by the ranking algorithm, they are just relegated to lower-ranked pages which are rarely seen.

At the beginning of the World wide web there was an initial attempt to treat the Internet like other information sources, making selections based on quality and reliability. Sites like AltaVista<sup>16</sup> and Yahoo began to create human-assigned indexing to the contents of the web, which at that time was many times smaller than it is today. Libraries also tried to view the Internet as they would library materials. In the late Nineties, a research division of the OCLC launched a project called Cooperative online resource cataloging to add records to library catalogs for selected web sites and to generate web pages for libraries that mixed library-owned materials and web resources of particular importance to the library and its institutional context<sup>17</sup>. The project was short-lived, ending not long after the first meeting of participants in 1999. This was around the same time that the Google founders, Page and Brin, were

<sup>15</sup> Siva Vaidhyanathan, *The Googlization of everything*. Berkeley: University of California Press, 2010. p. 3.

<sup>16</sup> Web pages from earlier days can often be found at the Internet archive, such as: <<http://web.archive.org/web/19990208011304/http://www.altavista.com:80/>>.

<sup>17</sup> Thomas B. Hickey, *CORC-cooperative online resource catalog*, June 11<sup>th</sup>, 2003, <<http://worldcat.org/arcviewer/1/OCC/2003/06/11/0000003480/viewer/file1.html>>.

developing the first versions of their search engine, with the assumption that the only viable solution to indexing the contents of the web would require powerful computing environments. Without these search engines there would be little access to materials on the web, we would be returned to the very early days of Internet access, when one had to know *a priori* what sites to look at in hopes of finding material because there was little or no ability to perform searches. This favored the sites of companies and organizations already known the analog world, and rarely would non-mainstream content be accessed.

### Neutrality

Because services like Google and even YouTube are often compared, either favorably or unfavorably, to libraries, it would be instructive to know what speakers mean when they use libraries in that comparison. Unfortunately, a definition of 'neutral' or even 'library' is rarely offered, and the assumptions about libraries that exist in the minds of the speakers are not available to us. One hears librarians declaring the superiority of libraries due to their 'neutrality', although, the term 'neutrality' is rarely defined, and it is not always possible to extrapolate from the arguments to a supposed definition. An example of this was a debate scheduled at the American Library Association meeting in February of 2018<sup>18</sup>. Panelists spoke pro or con neutrality, but there was little actual debate because the parties were using such different assumptions about the meaning of the term 'neutrality' that they did not hold a definition in common. The two dominant definitions of neutrality were:

- that the library is neutral because it treats all qualified members of the defined user community equally, without discrimination based on race, age, religion, or social status. This definition addresses the libraries obligation to users, but does not touch on the question of the contents of the library;
- that the library is neutral because it provides access to the widest possible range of information sources, without itself promoting any particular school of thought or point of view.

The first argument, that of treating users equally, is not a strong one. As public institutions, for libraries there are laws that govern service obligations. As one speaker mentioned, a librarian must serve even those members of the public with whom she disagrees, such as neo-Nazis or religious extremists, as long as their behavior does not violate rules set in place to protect the library and other users. Those adhering to this definition of neutrality would probably say that had Dylann Roof entered a library with the same question on black on white violence, the librarian would be obligated to attempt to find materials to answer his query. The librarian would probably not be breaching neutrality by offering materials with varying viewpoints, but could not promote one view over another.

The second argument appears to be the stronger one, but it does not hold up to analysis. In this argument the library carefully selects materials that support a wide variety of views resulting in a kind of information commons where users can learn and come to conclusions on their own. However, this is true only within a somewhat narrow range of socially and scientifically accepted materials. Libraries, as mentioned above, generally carry only materials issued from reputable publishing hous-

**18** Amy, Carlton. *Are libraries neutral? President's Program tackles heavy subject from multiple angles*. American Libraries, February 12, 2018. <<https://americanlibrariesmagazine.org/blogs/the-scoop/are-libraries-neutral/>> Accessed. March 7, 2018



## NOTE E DISCUSSIONI

es. This does not include ‘fringe’ publications, or if some are included in the collection they are carefully selected for the interests of the immediate community. By limiting the collection solely to those materials from publishers deemed ‘serious’, the library becomes a mere reflection of current culture, which includes class and racial biases, among others.

There is also no guarantee of neutrality in the output of mainstream publishing. While in theory libraries select for reliability and accuracy, there is a great deal of popular literature that is at best speculative in nature, including ‘self-help’ materials that give unscientific lifestyle advice; any number of conflicting dietary theories; and dubious claims relating healing and spirituality. There are also many resources offering business advice that may or may not provide the desired results. However, the library can defend its choices as being within the normal culture of discourse, thereby leaning on the publishing industry as the actual arbiter of content. This is what many mean when they refer to library neutrality; yet, this simply amounts to a form of passive acceptance of the dominant culture represented by the publishing industry.

There are topical areas around which libraries and their communities discuss the library collection. Some topics, such as pornography, are almost uniformly considered out of bounds, as are materials that are deemed to be socially divisive, such as hate speech. But even with these there is no bright line of distinction. Some public libraries in the United States carried Madonna’s book *Sex*<sup>19</sup>, although in a number of cases community members found it objectionable and asked that it be removed from open shelves if not from the library altogether, leading to significant struggles between libraries and their communities<sup>20</sup>. Episodes like this show that while some library material selection is routine and non-controversial, selection itself will lead to choices that are not neutral.

The conference debate that brought us to this point contained a number of threads that do not fit neatly into either of the two stated viewpoints. Brought up at that debate, and elsewhere, is the view of activist librarians who feel that their purpose is to support the community against the very dominant culture that oppresses minority members of society<sup>21</sup>. The American Library Association through its policies and lobbying efforts fights challenges against books that address controversial subjects such as homosexuality, drugs, and sexuality, especially in books written for younger readers<sup>22</sup>. Clearly advocating for the topics that make some community members uncomfortable while they provide support to others is a non-neutral position. It is also pointed out that librarians themselves are more representative of the dominant society than of society as a whole<sup>23</sup>, with the acknowledgement that US librarians

**19** Madonna; Steven Meisel; Fabien Baron, *Sex*. New York: Warner Books, 1992.

**20** *Free-speech champion Gordon Conable dies*, «American Libraries», January 18<sup>th</sup>, 2005, <<https://americanlibrariemagazine.org/free-speech-champion-gordon-conable-dies/>>.

**21** R. David Lankes, *My remarks on library neutrality for the ALA midWinter president’s panel*, «R. David Lankes», February 11<sup>th</sup>, 2018, <<https://davidlankes.org/my-remarks-on-library-neutrality-for-the-ala-midwinter-presidents-panel/>>; *Questioning library neutrality: essays from progressive librarians*, edited by Alison Lewis. Duluth: Library Juice Press, 2014.

**22** American Library Association, *Top ten most challenged books lists*, <<http://www.ala.org/advocacy/bbooks/frequentlychallengedbooks/top10>>.

**23** Myrna Morales; Em Claire Knowles; Chris Bourq, *Diversity, social justice, and the future of libraries*, «Libraries and the Academy», 14 (2014), n. 3, p. 439-451; April Hathcock, *White librarianship in blackface: diversity initiatives in LIS*, January 28<sup>th</sup>, 2018, <<http://www.inthelibrarywiththleadpipe.org/2015/lis-diversity/>>.

are overwhelming white and middle-class<sup>24</sup>. If 'neutrality' means not questioning the dominant culture, then it also means a passive acceptance of all of the biases inherent in that culture, and this is especially problematic for librarians who serve those minority communities.

### Conclusion

Google and other search engines provide a view of the Internet that has social and political impact, although analysis of that impact is extremely difficult due to complexity and secrecy. What we do know is that the commercial search engines have an interest in maximizing advertising revenue. Libraries have a social basis in their funding, but this does not guarantee that there is no bias in their contents and services. Unlike search engines, however, libraries can be open to studies of their decision-making. It remains to be seen where an understanding of library bias will lead us. The first step is to question the idea of 'neutrality'. This is still a long way from accepting a role of social and political activism on the part of libraries and librarians, as the image of cultural neutrality in libraries is deeply embedded in the profession.

Ai fini della citazione, utilizzare esclusivamente il testo in lingua italiana, che presenta il DOI, la paginazione, l'abstract e gli altri dati ufficiali.

When citing, please always refer to the Italian translation only, complete with DOI, page numbers, abstract and other data.

[Karen Coyle, *La neutralità difficile*.

AIB studi, vol. 58 n. 2 (maggio/agosto 2018), p. 257-265. DOI 10.2426/aibstudi-11788]

<sup>24</sup> Angela Galvan, *Soliciting performance, hiding bias: whiteness and librarianship*, «In the library with the lead pipe», January 28<sup>th</sup>, 2018, <<http://www.inthelibrarywiththeleadpipe.org/2015/soliciting-performance-hiding-bias-whiteness-and-librarianship/>>; Todd Honma, *Trippin' over the color line: the invisibility of race in library and information studies*, «InterActions», 1 (2005), n. 2, <<https://escholarship.org/uc/item/4njow1mp>>.