Democratizing Search? From Critique to Society-oriented Design.

Bernhard Rieder

Democratizing Search? From Critique to Society-oriented Design.

Bernhard Rieder

Since 1995, when Engineers at the *Digital Equipment Company* introduced *AltaVista*, the first large scale search engine for the World Wide Web, many things have changed. The Web in 2009 is the dominant stage for all things related to information and communication. It hosts, on a single technological platform, a wide variety of activities that were formerly distributed over many different channels. With 1.5 billion users, over a trillion pages and a diversity of services ranging from the simple display of text-based information to elaborate online applications that combine databases and sophisticated multimedia, the contemporary Web is an informational behemoth, and central to a capitalist economy whose mode of value production is shifting from industrial to *cognitive*[1]. Without any kind of built-in indexing or cataloguing system, it falls to search engines to make the rather unruly structure of the Web manageable for human users. These complex tools are surprisingly easy to use: a search query composed of one or several search terms will bring up an ordered list of pages that contain the specified words. It is no wonder that search engines are among the most popular services on the Internet.[2]

This central role of search engines has brought a considerable amount of critical attention to these technical artifacts and, more recently, to the companies that build them. Despite its reassuring (informal) corporate motto – "don't be evil" – *Google Inc.*[3], the undisputed market leader, has somewhat replaced *Microsoft* as the favorite target for scrutiny and critique by Internet researchers and activists. What initially seemed to be a purely technical matter concerning only computer and information scientists has slowly been engulfed in a rich debate[4] that includes social, political, cultural, economic, and even philosophical

---

1 The concept roughly holds that capitalism, after passing through a mercantile and an industrial phase, has entered a new stage where the production of wealth is increasingly based on the production of immaterial goods. Theoretical work on the concept of "cognitive capitalism" has gathered around the journal *Multitudes*, driven by authors like Yann Moulier Boutang, Maurizio Lazzarato, Antonio Negri, Paolo Virno, and others.

2 Deborah Fallows, "Search Engine Users," Pew Internet & American Life Project, published January 23, 2005, http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf.

3 In this article, I distinguish between *Google search* (http://www.google.com) a Web search engine and *Google Inc.*, the company that owns and operates this service. *Google Inc.* offers many other products, such as online office applications, video sharing sites, an email service, an online social network, etc.; it generates virtually all of its revenue through advertisements placed on its own sites as well as on thousands of partner sites.

4 Eszter Hargittai, "The Social, Political, Economic, and Cultural Dimensions of Search engines: An Introduction," *Journal of Computer-Mediated Communication* 12, no. 3 (2007), http://jcmc.indiana.edu/vol12/issue3/hargittai.html.

considerations. Search is increasingly theorized in terms of representation[5] and power[6]; matters of privacy, equality, plurality, and commercialization are taken into account. At the lower levels of normative reasoning, however, a lot of work remains to be done. This article will therefore consider two problems. First, while much of the critique directed against search engines relies on a set of particular values, the discourse on search is rarely very explicit when it comes to anchoring these values in larger normative frameworks; the arguments put forward may therefore at times appear unconnected. Second, both design and policy recommendations often remain, as a result, vague and unspecific. This is due in part to the fact that expert technical knowledge on Web search remains elusive in the humanities; C. P. Snow's "two cultures" remain separate for the time being. But it is also related to the first problem: without a clear normative standpoint it is difficult to formulate suggestions how search *ought* to work. After a short summary of the most common points of criticism directed at Web search and an account of the two dominant values embodied in current general-purpose search technology, popularity and convenience, I will therefore develop a rather explicit normative position that advocates plurality, autonomy, and access as alternative guiding principles for both policy and design. The concluding recommendations are probes into the question of how these principles may be applied in practice.

1 Web search as normative technology

As the question mark in the title suggests, "democratizing search" is by no means a trivial matter and this essay is necessarily informed by the values and limitations of its author. It is meant as a contribution to the political debate on Web search; the necessary starting point, then, is a closer look at the arguments that are commonly made.

1.1 Common critique

Most strikingly, the discussion surrounding web search has been resolutely *critical*. But although many of the central issues were been raised in two early papers,[7] the stabilization of clear lines of critique was a longer process. Today, we can identify three subject areas that have caused the most contention:

The fact that search engines habitually store personal search data for their many users has lead to an interest in matters of *privacy* and its counterpart, *surveillance*. Search engine companies

---

5 Lucas Introna and Helen Nissenbaum, "Shaping The Web: Why The Politics of Search Engines Matters," *Information Society* 16, no .3 (2000): 169-185; Susan Gerhart, "Do Web Search Engines Suppress Controversy," *First Monday* 9, no. 1 (2004), http://www.firstmonday.org/issues/issue9_1/gerhart/index.html.

6 Bernhard Rieder, "Networked Control: Search Engines and the Symmetry of Confidence," *International Review of Information Ethics* 3 (2005), http://www.i-r-i-e.net/inhalt/003/003_rieder.pdf; Theo Röhle, "Machtkonzepte in der Suchmaschinenforschung," in *Die Macht der Suchmaschinen / The Power of Search Engines* ed. Marcel Machill and Markus Beiler (Köln: Herbert von Halem Verlag,2007), 127-142.

7 Hartmut Winkler, "Suchmaschinen. Metamedien im Internet," *Telepolis*, Mars 12, 1997, http://www.heise.de/tp/r4/artikel/1/1135/1.html; Introna and Nissenbaum, 2000

cooperating with totalitarian regimes on repression and censorship, data retention (and erasure) policies, and the problem of cross-database profiling have been some of the main points of critique. However important, these problems are not specific to search engines but concern all systems that store personal data. I will therefore leave them aside.

When in the nineties search engine companies started to include paid links in their result lists, *advertisement policy* quickly became a major issue, which lead to the 2002 disclosure recommendations by the US Federal Trade Commission (FTC), which demanded that "any paid ranking search results are distinguished from non-paid results"[8]. The tension between public interest and commercial gain continues to be a permanent issue,[9] albeit typically in the context of the last subject area.

The question of *ranking* – the ordering of search results – has certainly been the most popular topic of debate. Winkler noted early on that search engines divide the Web into "main and side roads"[10] and Introna & Nissenbaum argued that they reshape the Web in favor of commercial interests by directing "prominence" to big sites rather than small ones. [11] Hindman et al. coined the term "Googlearchy" to describe a Web that is dominated by a few highly ranked sites.[12] The question of hierarchy is at the core of search and this article will concentrate on the matter.

There are, however, two additional issues that run through all of the three subject areas:

First, critics commonly point out the lack of *transparency* when in comes to privacy, advertisement, and ranking. Modern IT systems are largely black boxes – or "black foam"[13], as I have suggested elsewhere;[14] we can never know for sure what actually happens behind the interface, which renders informed critique quite difficult. Disclosure of technical specifications and corporate policy has consequently been the most frequent demand by researchers and activists. Second, with *Google Inc.* controlling an overwhelming fraction of

---

8 FTC Bureau of Consumer Protection, "Commercial Alert Complaint Letter", FTC, http://www.ftc.gov/os/closings/staff/commercialalertattatch.shtm.

9 Alejandro Diaz, "Through the Google Goggles: Sociopolitical Bias in Search Engine Design," in *Web Search: Multidisciplinary Perspectives*, ed. Amanda Spink and Michael Zimmer (Berlin: Springer, 2008), 11-34.

10 Winkler, 1997

11 Introna and Nissenbaum, 2000

12 Matthew Hindman and Kostas Tsioutsiouliklis and Judy A. Johnson, "'Googlearchy': How a Few Heavily-Linked Sites Dominate Politics on the Web" (paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, Illinois, April 3-6, 2003).

13 The basic difference is that with a box we know, at least, where the system starts and where it ends. Modern IT is generally so intertwined that we can often no longer determine the limits of the system. E.g., we can only guess at the level of interaction and data exchange between the different services offered by *Google Inc*.

14 Rieder, 2005

the search market[15] and the crucial online advertisement business – the most important revenue stream for free online services – the question of *monopoly* is raised with increased frequency.

Behind each of these five points lurks a normative debate that is framed by the political conflicts of contemporary capitalism, including the questions of how to balance public and private interest, how to ensure equal rights to participation and expression, how to govern multinational companies, and how to keep markets open to competition. Established political positions are not left behind when discussing Web search and the question is inseparable from larger issues concerning the politics of IT. At the same time, Web search is part of a specific research tradition, the field of *information retrieval* (IR), and in order to understand the normative dilemmas we face, we have to consider some of its specifics.

1.2 General-purpose search

The field of IR emerged over the course of the 20th century as an answer to the problems associated with the acceleration of information production in all sectors of western society. While the first half of the 20th century was marked by library innovators like Paul Otlet and increasingly finer-grained classification systems, the emergence of automatic sorting devices and general-purpose computers opened up new directions in thinking about information. The work by pioneers like Hans Peter Luhn and others led up to Gerard Salton's monumental book "Automatic Information Organization and Retrieval" in 1968, which firmly defined computer algorithms as the logical way to meet the problem of finding information. But IR was developed at a time when information was mostly stored in well-structured form in thematic databases. Many of the conceptual ambiguities associated with Web search come from the fact that data on the Web is neither structurally nor thematically consistent.

The Web can be understood as an address space for document access. In principle, every unit of information is accessible via a *Uniform Resource Locator* (URL)[16]. At least the *surface Web*[17] is, from a purely technical perspective, a homogenous space of informational resources, stored in the form of HTML documents. Web search, by which I mean the *general-purpose* type epitomized first by *AltaVista* and then *Google search*, is a technical artifact that treats the Web as this uniform address space and does not make any distinctions between different types of information beyond abstract categories such as file type, date, or language. Search engines use a "one size fits all" approach to brokering access to a wide variety of

---

15 In March 2008 *comScore* announced that 80% out of all searches in Europe are conducted on sites owned by *Google Inc*.: comScore, "comScore Releases March 2008 European Search Rankings," http://www.comscore.com/press/release.asp?press=2208.

16 This holds true not only for static content (e.g. http://www.example.com/example.html) but also for sites that generate pages from a database (e.g. http://www.example.com/example.php?pid=354).

17 The part of the Web accessible by following links; information that can be found only by querying search fields is part of the *deep Web*.

resources; contextual precision, e.g. the distinction between shopping for a book and looking for a summary, has to be provided by the user in the form of additional search terms.

This universality is in fact one of the major difficulties when it comes to conceptualizing Web search in non-technical terms. Framing the Web purely in terms of topological (a network of documents and links), syntactical (documents as containers of markup language), and statistical (word occurrences) structure radically departs from our human habit of ordering information into subject matters, areas of activity, context, and so on. We would intuitively concur that a restaurant address, the latest gossip on a Hollywood starlet, the price of a laptop, a blog post on weight loss, a scientific paper, and a newspaper article on some political candidate's stance on international politics do not embody the same *type* of information, that they belong to different domains of existence. We will probably also agree that each one of these items can be linked to a set of distinct activities – from diversion to research – that might imply specific acts of decision-making such as settling on where to have dinner or who to vote for in an election. We might even come to a rough consensus on the importance or triviality of each one of these items. These (informal) levels of differentiation belong to what Clifford Geertz defined as *culture*, "webs of significance", order based on *meaning*, not on statistics and graph theory.[18] I do not want to argue that these methods do not capture certain semantic dimensions; neither that research into semantic techniques is necessarily a dead end. But the fact remains that current general-purpose search engines treat the Web as a single whole and use the same techniques to search and rank pieces of information that pertain to a wide variety of domains. Their relative agnosticism to meaning is their brand of *objectivity* – which still implies a particular cultural logic and normative orientation.

1.3 Popularity and convenience

In 1963, after a series of preliminary projects, Eugene Garfield published the first edition of the *Scientific Citation Index* (SCI), a complete index to the 1961 volume of 613 scientific journals containing an ordered list of over 1.4 million citations. Since then, the SCI has allowed scientists to search for relevant papers by following the "trails of association" – a term coined by Vannevar Bush – established through the scientific practice of citation. Starting with any given publication, a scientist can easily find all the papers citing it and techniques like co-citation analysis (two articles citing the same sources may be thematically related) turn the SCI into a powerful tool for information retrieval.[19]

The SCI is of interest here for several reasons. First, it anticipated an important development in the history of the Web, namely the move away from hand-selected and classification based directories like *Yahoo* towards the completely automatic search methods epitomized first by *AltaVista* and now by *Google search*. Its arguments were almost identical to those of the later search engine: that there was too much information to be handled by human editors, that manual classification was slow and expensive, and that controlled vocabularies were

---

18 Clifford Geertz, *The Interpretation of Cultures* (New York: Basic Books, 1973), 5.

inflexible, burdensome, and ultimately subjective. Second, the SCI represented a real paradigm-shift away from content-based organization towards topological analysis built on graph theory. Web search engines later made a similar transition: *AltaVista*'s ranking was still mainly based on document properties, meaning that a search terms frequency of occurrence, position in the document, presence in the URL, etc. would be the dominant factors in ranking. According to Page et al. , "[i]t is obvious to try to apply standard citation analysis techniques to the web's hypertextual citation structure",[20] and *AltaVista* did in fact count the links that pointed toward a document but did not attribute a dominant role to the link-factor. The success of *Google search* however was mostly built on an explicit link-topological method, where every citation is attributed a specific "weight" (*PageRank*) depending on the "importance" of the site it comes from, which is itself based on a recursive calculation of the whole graph. A third point about the SCI is that its "impact factor" (a citation count for papers, individual scientists or institutions based on the SCI), which has become the dominant means for evaluating scientific productivity, has long been a target of the same sort of criticism that is now directed at search ranking. Researchers have argued that quality is not the same as notoriety and that citation-based ranking might stifle innovation by installing a star system and reducing diversity.[21]

This equation of importance with popularity is indeed the central critique of link analysis as the dominant method of ranking search results – *bias* in search engines is generally understood in this sense.[22] Instead of favoring a particular opinion, political party or company, the worldview embedded in link analysis is much more abstract and, in a sense, delegates ranking *to the Web itself*, as the links that will determine *PageRank* or other topological measures have not been placed by the search engine but the people who create websites, blogs, and other types of online content. It is not surprising that *Google*'s own rhetoric strongly relies on "democratic" imagery and equates links with "votes"[23]. The search engine then functions as a mere (vote) counting mechanism and the company used to pride itself that "no human involvement" taints this process, "which is why users have come to trust Google as a source of objective information"[24]. Ranking is nonetheless targeted at emulating

---

19 The SCI is now part of the *Web of Science* database, maintained by *Thomson Scientific*.

20 Larry Page et al., "The PageRank Citation Ranking: Bringing Order to the Web," Technical Report, Stanford InfoLab, 1999, http://ilpubs.stanford.edu:8090/422/

21 Robert Adler and John Ewing and Peter Taylor, "Citation Statistics," Report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS), published June 12, 2008, http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf

22 Introna & Nissenbaum, 2000

23 Google Inc., "Technology Overview," http://www.google.com/corporate/tech.html

24 The phrasing has changed in mid 2008 probably in preparation for the *SearchWiki* feature that allows users to manually move sites to the top. The referred passage can still be found on the Internet Archive at

human agents' judgments and the closer the equation between the machine view and the users' appreciations, the higher the "quality" of results. Search engine companies therefore employ teams of human evaluators that test changes in algorithms and decide whether they are beneficial or not.[25] A study by Pan et al. suggests that most users are willing to place great trust in the competence of this process.[26]

Generally speaking, link analysis turns the power-law link structure of the Web, where a small number of hubs dominate a large number of scarcely linked sites [cf. Hindman et al. 2003] into a measure of importance. The underlying principle has been called "cumulative advantage", "preferential attachment" or "Matthew effect", but the consequence is simply that already well-ranked sites have a higher visibility and therefore get linked more often, leading to yet better rankings. In other words, the rich get richer. Using popularity as measure for quality is, of course, a normative decision. The *logic of the hit*, combined with the fact that search engine optimization (SEO), link campaigning, and classic marketing allow economically potent actors to skew the game in their favor is effectively responsible for both centralization[27] and commercialization[28] tendencies. But there is a second *core value* informing the current design of search engines.

The success of *Google search* has been, in part, attributed to its simple, uncluttered interface. Apart from the basic query language, the user has few means to influence the search process and ranking parameters are completely off-limits. Following the recommendations of user-oriented design – a design philosophy mostly based on cognitive psychology – the goal is to make the search process as *simple and convenient* as possible. That this is where research is continuing to head has recently been made clear by Marissa Mayer, Vice President of Search Products & User Experience at *Google Inc.*, in her definition of the "ideal search engine":

> Your best friend with instant access to all the world's facts and a photographic memory of everything you've seen and know. That search engine could tailor answers to you based on your preferences, your existing knowledge and the best available information.[29]

The goal of personalization is to make search yet more convenient by using personal search histories and session profiles to disambiguate queries. If a user has been surfing shopping sites for the last hour, a query for a book title might automatically favor bookstores over

---

http://web.archive.org/web/20071228101625/www.google.com/intl/en/corporate/tech.html.

25 Scott Huffman, "Search evaluation at Google", Official Google Blog, posted September 15, 2008, http://googleblog.blogspot.com/2008/09/search-evaluation-at-google.html.

26 Bing Pan et al., "In Google we trust: Users' decisions on rank, position, and relevance," *Journal of Computer-Mediated Communication* 12, no.3 (2007), http://jcmc.indiana.edu/vol12/issue3/pan.html.

27 Winkler, 1997

28 Introna & Nissenbaum, 2000

29 Marissa Mayer, "The Future of Search," Official Google Blog, posted September 10, 2008, http://googleblog.blogspot.com/2008/09/future-of-search.html.

informational or scholarly documents. The search experience becomes faster and simpler.

Both popularity and convenience are derived from user-oriented design principles where "reducing the cognitive effort and time costs for searchers"[30] is the main objective. Despite the "democratic" rhetoric, design decisions are based on perceived benefits to individual end-users; reasoning on the level of society is rarely coming into play. But if search engines are indeed powerful gatekeepers and therefore central social institutions, we may legitimately ask ourselves what a more "society-oriented" approach that goes beyond the values of popularity and convenience would look like.

2 Alternative guidelines

What I have called "society-oriented design"[31] is not a detached form of normative speculation but the attempt to bridge the gulf between the contextualized practice of technical creation and considerations of social benefit that go beyond efficiency, control, and material prosperity. Normative reasoning, here, is *bounded* by feasibility. Therefore, if Herbert Simon was right in 1971 when he declared *attention* a scare resource consumed by an overabundance of information, we have to recognize that ranking is not only very useful but also inevitable. If there is more than one result for a query – and most often, there are many more – the only way to bypass ranking would be to order results randomly, most probably rendering the system incredibly frustrating. As Winkler argued,[32] there is no anti-hierarchic medium – just like there is nothing outside of power (Foucault) and no freedom in detachment (Latour). Any system of ranking will favor certain sites over others; the question is which ones. The goal, then, can neither be the abolition of ranking in search, nor the design of the "perfect search engine" (Larry Page). But between these two extremes lies a series of possibilities for design and policy that do indeed merit our – however scarce – attention.

2.1 *Wikia search* as society-oriented design

When looking at some of the Web 2.0 rhetoric one might get the impression that larger social and political considerations have already found their way into the mindset of designers. The term "democracy" has indeed become omnipresent in the Web 2.0 age. It is by and large defined as the user's ability to vote – on whatsoever; from ideas to video clips and from news items to search rankings. Jimmy Wales, the co-founder of *Wikipedia*, has recently attracted a lot of attention with *Wikia Search*, a new "social" search engine that tries to make use of this simple logic in its attempt to "democratize search".[33] The idea of "social" search, where user

---

30 Pan et al., 2007

31 Bernhard Rieder, "Métatechnologies et délégation. Pour un design orienté-société dans l'ère du Web 2.0" (PhD diss., Université de Paris 8, 2006), http://tel.archives-ouvertes.fr/tel-00179980/.

32 Winkler, 1997

33 Jimmy Wales, "Free Speech, Free Minds and Free Markets" (talk at the Ford Hall Forum, Suffolk University, Boston, Massachusetts, September 11, 2008), http://fora.tv/2008/09/11/Jimmy_Wales_-

ratings replace link analysis as dominant mode of ranking, has been implemented several times over the last years – *Eurekester*, *Mahalo*, *Wink*, etc. – but the effort lead by Wales stands out not only because of its leader's celebrity but also because of the very explicit discussion of the norms and values that (should) guide its technical design. The project's internal mailing list is an interesting source because it is through this channel that four principles have been decided on:

- *Transparency* has been one of the central lines of critique for search engine providers and *Wikia Search* has decided very quickly to strive for openness on the technical level (all software is open source), on the database level (the index can be downloaded), and on the organizational level (decision-making and corporate functioning should happen in the open).

Community is directly derived from Wales' most successful project, *Wikipedia*, and implies that ranking and other tasks will rely on user labor. The goal is to attract a sufficiently large number of active users and involve them both in the task of sorting out results and collective governance of the project.

Quality would seem obvious as a goal but it is very interesting that the *Wikia Search* project explicitly aims at results that are useful to a very high number of people. The quality principle implicitly acknowledges that ranking is not merely a problem of values or worldview but embedded in users' trajectories of activity and therefore more or less *useful*.

Privacy is the second element that is directly related to common critique concerning search and other services on the Web. Here, the *Wikia Search* project is quite conscious of the fine line that exists between the very useful gathering of data and the users' right to protect their identity. The consensus that seems to emerge is to store as little user data as possible and to make user identification (which is crucial to transparency and community decision-making) an opt-in feature.

For the time being, *Wikia Search* is work in progress and it remains to be seen how the guiding principles will be applied in a more definitive version of the search engine. But there are at least two observations that can be made at this point: first, it is quite reassuring to see that a deep discussion of values can be productively brought *into* the process of technical design. This is obviously not a straightforward matter of coming up with a set of values and *implementing* them, but a rather difficult process in which technical, commercial, and ethical demands are in constant conflict. Second, when looking more closely at the guiding principles of *Wikia Search*, we find that while some common criticism of traditional search engines is taken into account, the central tenant of the project – users adding, rating, and deleting results – does neither address the problem of equality in search results, nor question popularity as the central expression of importance. *Wikia Search*, in its current stage, produces the same top-down list of results as established search engines; when looking at some of the result pages, it becomes quite obvious that the problem of spam is nowhere near solved and, more importantly, that dominant opinions in the community will lead to imbalanced results. While

researching for this article the query "John McCain" brought up a list that, below the obligatory *Wikipedia* entry, essentially pointed to sites criticizing or ridiculing the republican presidential candidate. It is entirely possible that further versions of the engine will be less easily controlled by dominant user groups but this will probably entail a fine grained system of governance, which, similar to *Wikipedia*, will have to rely on administrators, result policing, and user banning.

Wikia Search is certainly a fascinating experiment in society-oriented design but its primary concern is the *social governance of the object itself* and not the relevance of search as part of a larger *socio-technical configuration*. This has something to do with the specific interpretation of democracy as *community* that informs even the more sophisticated Web 2.0 projects.

2.2 Two concepts of democracy

Using the common sociological distinction between *community* and *society* – habitually traced to Tönnies but certainly prefigured in Durkheim's distinction between organic and mechanic solidarity – we can roughly distinguish between two ideas of a "government of the people, by the people, for the people". Without going into detail, I believe that the *democracy as community* idea can be found most explicitly in certain forms of protestant doctrine (e.g. Puritanism) and in American social thinking in the tradition of Emerson and Whitman. Highly skeptical of political institutions, it holds that humans are rational beings capable of self-governance. When people come together in good will, consensus will inevitably emerge and as Howard Zinn argues, "[t]o depend on great thinkers, authorities, and experts is […] a violation of the spirit of democracy".[34] There is, however, considerable doubt that this ideal can (or should) be applied to the governance of contemporary nation-states:

> I believe that a democratic society is not and cannot be a community, where by a community I mean a body of persons united in affirming the same comprehensive, or partially comprehensive doctrine.[35]

For political philosopher John Rawls, the pluralistic composition of modern societies demands a concept of democracy that takes into account that certain interests and opinions cannot be easily reconciled. The idea of *democracy as society* is therefore built upon complex political institutions and processes, separation of power, and guaranteed constitutional rights that protect citizens from state control. While there is considerable disenchantment with the political maneuvering and compromises that seem to dominate liberal democracy and a longing for the warmth of community, we should be highly cautious when it comes to writing off society – the coexistence of people that neither agree nor resemble each other – as a locus for democratic governance. As Vedel points out, the community ideal tends to perceive key political institutions – parties, unions, media companies – as perversions of the democratic

---

[34] Howard Zinn, *Passionate Declarations: Essays on War and Justice* (New York: Perennial, 2003), 6.

[35] John Rawls, *Justice as Fairness. A Restatement* (Cambridge MA: Harvard University Press, 2001), 3.

ideal.[36] But how can we imagine governance of very large groups without mediation? The Web 2.0 discourse provides an answer: instead of slow and intransparent institutions we now have software that allows us to scale the community ideal to the level of society; the resulting transfer of power to the companies providing these tools – which thereby become social institutions themselves – is rarely addressed.

Curiously, both understandings of democracy seem to lead to a common normative concept when it comes to Web search, namely some variant of Habermas' ideal of the public sphere as a space of egalitarian expression. The background of virtually all critique directed at ranking is the worry that the Web might lose its capacity to provide a means of expression to formerly excluded voices and to function as "a valuable collision space between official and unofficial accounts of reality".[37] Depending on the favored ideal of democracy, there are subtle but important variations. While the "community" version, which informs most Web 2.0 rhetoric (and many recent scholarly publications), insists that each voice has equal value, the "society" tradition emphasizes that "[all citizens] should have a fair chance to add alternative proposals to the agenda for political discussion".[38] Again, the second stance does not banish mediation, selection, and weighing as long as they are based on the argument made and not status, wealth, or power. Habermas not only accepts a certain level of "critical filtering" – the task of the journalist – but explicitly warns of the "fragmentation" a public sphere without central mediators would fall victim to.[39] If we see selection and hierarchization as necessary functions for democracy the whole question is how to define Rawls' "fair chance" to add an opinion to the debate. And if search engines function indeed as "media gatekeepers"[40] the first objective is to ask how their filtering is affecting this "fair chance" and, if necessary, to correct its course. The particular process that produces search results becomes less important than the question whether this process furthers "reasonable pluralism" or not. This does not exclude community as a value and deep participation as a mode of governance; but the application of voting to ranking can no longer be seen as the obvious and uncontested way to "democratize search".

Before turning to alternative recommendations, I would like to point out that the Web can be seen as something other than a *mass medium*, e.g. as a huge library, a repository for cultural representation or a tool for education. This perspective opens up alternative sources for normative reasoning, including the following:

---

36 Thierry Vedel, "L'idée de démocratie électronique. Origines, visions, questions," in *Le désenchantement démocratique*, ed. Pascal Perrineau (La Tour d'Aigues: Editions de l'Aube, 2003), 243-266.

37 Richard Rogers, *Information Politics on the Web* (Cambridge MA: MIT Press, 2004), 28.

38 John Rawls, *A Theory of Justice* (Cambridge MA: Harvard University Press, 1971); 225.

39 Jürgen Habermas, "Preisrede anlässlich der Verleihung des Bruno-Kreisky-Preises für das politische Buch 2005" (talk at Universität Wien, Austria, March 9, 2006), http://www.renner-institut.at/download/texte/habermas2006-03-09.pdf

40 Diaz, 2008

- *Ethical guidelines for librarians and documentation professionals*: Besides the usual niceties, the ethical codes of professional associations such as the European Council of Information Associations (ECIA) or the American Library Association (ALA) establish specific ethical guidelines such as confidentiality and unbridled *access to information*, which not only means rejecting censorship but actively helping users in their search.
- *Cultural diversity policy*: International declarations such as the UNESCO "Convention on the Protection and Promotion of the Diversity of Cultural Expressions", ratified in 2005 by all UN member states except for the US and Israel, expressively names cultural diversity as worth protecting and the famous *exception culturelle* clause, added in 1993 to the GATT (now WTO) agreements, allows countries to exempt cultural products from free-trade agreements. These are efforts to extend the principle of *plurality* from the level of opinion to that of cultural expression.

Education and empowerment: In the spirit of enlightenment and education reformers like Paulo Freire, we can think of the Web as a tool for empowerment and a means for individuals to gain *autonomy*. This cannot mean abandoning users to their own devices but devising "critical pedagogy" whose goal is to help users to look behind the surface of Web search.

When combining these lines of thought with the public sphere model, three principles emerge: (reasonable) *plurality*, *autonomy,* and *access*.

3 Proposals and perspectives

In the context of society-oriented design, normative reasoning is necessarily oriented towards application. Following the three principles established in the previous section, what follows is a series of proposals that try to go beyond the often heard demand for more transparency. Having more information about the inner workings of search engines would surely make informed critique of ranking mechanisms easier but even now, we have a relatively good idea how search engines produce their results.[41] Critics ask for more transparency to further social control and to impede the arbitrary exercise of power; but the problem with the current search configuration is not simply the potential for abuse but rather the one-sided focus on popularity and convenience. The following suggestions aim at reorienting the search configuration according to the principles of plurality, autonomy, and access, which, in practice, form a coherent whole.

3.1 Regulation and stimulus

While it is uncontested that search engines provide immense social utility, the overwhelming dominance of the global search market by a single company is indeed a problem. Without

---

41 There is considerable doubt however when it comes to banning and negative ranking parameters. Google search states that "[l]inks to web spammers or bad neighborhoods on the web" might reduce rank but these definitions remain vague. Cf. Google Inc., "Link schemes", http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=66356

accusing *Google Inc.* of any wrongdoings, the sheer concentration of power over visibility on the Web should give us pause. From the standpoint of the public sphere model, the Web is part of the media system and if search engines perform similar functions to media outlets, we can argue that the measures put in place by western democracies to thwart media concentration may apply here as well. Even without any malicious manipulation of search results, *Google Inc.* gains tremendously from its privileged gatekeeper status. The huge amounts of data users leave behind in the different services the company provides are invaluable market intelligence that can either be sold or used to optimize corporate strategy. Prominent links on the main search page direct users to the company's entire product line, a considerable advantage when introducing a new service, and with complete knowledge of ranking procedures *Google Inc.* has the best possible SEO a company can hope to have. As Kawaguchi and Mowshowitz argue, we need to make sure that alternatives exist and from a policy standpoint there are two strategies to do that: states can either limit the dominant actor or help the competition.[42]

Concerning the first possibility, telecommunication laws in most democracies regulate, often in precise percentages, the part of the media market a company may control and limit so called "cross-ownership", e.g. a television network acquiring a newspaper company. While such regulations may not be easily applied to the search market in letter they may be in spirit. The 2002 FTC case had already established a link between search and media legislation and the European Commission's Article 29 Working Party is actively investigating[43] matters of data protection concerning search engines. The 2004 European Union vs. *Microsoft* competition case might also provide guidelines, specifically concerning the use of the dominant position in search to control other markets. However, given the complexity of such cases we should not rely solely on constraining measures; stimulating competition in the search market might be a more viable strategy.

Again, there is precedent in the context of media law. Especially European countries grant considerable benefits to newspaper companies, either through direct financial aid (Austria, France, Spain, etc.) or via reduced taxes or fees for delivery (Germany, UK, Switzerland, etc.). The French *Quaero* (€99M) and the German *Theseus* (€90M) projects are pioneer attempts to publicly fund R&D in the field of search. These sums, despite being doubled by the private consortium partners, are of course nowhere close to the major search companies' resources and the projects, realistically, do not target general-purpose application but multimedia search (*Quaero*) and semantic approaches (*Theseus*). It is somewhat regrettable that these funds are not directly applied to areas that are central to liberal democracy. Because if we perceive the Web as a heterogeneous infrastructure we may conclude that "more

---

42 Akira Kawaguchi and Abbe Mowshowitz, "Bias on the Web," *Communications of the ACM* 45, no. 9 (2002): 56-60.

43 Article 29 Data Protection Working Party, "Opinion on Data Protection Issues Related to Search Engines," European Commission, published April 4, 2008, http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf

egalitarian and inclusive search mechanisms"[44] are perhaps not needed in all areas but only in those that pertain directly to public interest, such as the representation of politically relevant information. Why not fund research in automatic news aggregation that, instead of ranking stories by popularity, the principle behind *Google News*, tries to represent the complexity and heterogeneity of the media debate? Applying the principle of plurality to general-purpose search has its limits and as Pieter van der Linden, the coordinator of the *Quaero* project, pointed out in personal conversation, search will have to increasingly focus on niches to make progress. A public strategy for research funding should select projects not only according to industrial potential but take into account larger questions of public interest. But this is not the only field where the state can play a productive role.

3.2 Exploration and re-ranking

By furthering users' autonomy we can increase plurality, even with the general-purpose logic that is prevalent today. If *convenience* means shortening interaction between the user and search engine, *autonomy* means trying to keep the mediator from disappearing. Education can certainly promote a more conscious and competent use of search engines and there is an argument to be made for the inclusion of critical information literacy into teaching curricula. But we also need to think about *technical means* to better explore the diversity of search results. While query operators – e.g. AND, OR, NOT – are most certainly powerful tools to access a larger set of sites, they are currently rather limited and there is little or no possibility to weight ranking parameters. Alternative search engines such as *Exalead*[45] – one of the few general-purpose search engines that actually produce result lists that differ significantly from the market leader's – allow searching with regular expressions, but there remains much room for improvement in extending users' control over the search process. While using query parameters undoubtedly requires a certain expertise, there are ways to deepen the search process without making it overly complex. Two examples:

- *Clusty*[46] is a search engine that divides results into thematic clusters that users can use to navigate up to 500 results at a time. The cluster list provides a first overview over the search subject and may direct users in new directions by showing aspects that they were not aware of before.
- *TermCloud Search*[47] is a search interface designed to map a topic rather than provide the shortest way between a query and a document. Using the simple *tagcloud* principle – keywords are shown in different sizes according to relevancy – the goal is to make users aware of the concepts surrounding their query and to encourage exploration

---

44 Introna & Nissenbaum, 2000

45 http://www.exalead.com/search

46 http://clusty.com

47 http://software.rieder.fr/termcloud

rather than quick answers.

These two examples are quite simple but there is enormous potential in search technology and interface design approaches that understand search as part of the knowledge building process and actively promote learning, plurality, and user interaction. Emphasizing the mediation process instead of making it disappear can at the same time strengthen users' autonomy and promote access to deeper regions of the index. In recent years however, exploding costs have become an important roadblock for innovation because a modern search engine is not just an elaborate piece of software but also an impressive feat of datacenter design and deployment. Indexing a steadily growing Web and executing billions of searches per day is a daunting task even if many queries are purely navigational (users typing "ebay" into the browser search field to get to ebay.com). To guarantee fast response time world-wide, a company has to build a network of datacenters, with each node within physical proximity to a central Internet hub to reduce the packed travel route. Promoting experimentation and innovation will have to include measures that help dealing with these infrastructure issues.[48] On the organizational level, this can be done in several different ways but I would like to quickly explore a route that implies using established companies' infrastructure.

Applications like *Clusty* and *termCloud Search* use so called APIs (Application Programming Interfaces) to get a set of results from different search engines in machine-readable form (e.g. XML), apply some form of processing (e.g. re-ranking the results), and display the results in a customized fashion (e.g. result clusters). While this is a great way to experiment and implement ideas without having to invest in server infrastructure, there are important limitations. First, the number of results per request is rather low – *Google* serves only up to eight, *Yahoo* and *Microsoft* 50 – and while an application can load several result sets at the same time, there are practical limits; *Clusty* stops at 500, *termCloud* Search at 250 to keep response time acceptable. Second, the contractual provisions of API terms of use are strongly leaning in favor of service providers. While this is not a problem for experimentation, building a business on a search API may be a risky affair as technical details and legal specifications may be changed at will. While there are robust licensing models on the content side such as *creative commons*, developers using APIs still hope for a more standardized and accountable way to clear up legal uncertainties.

In order to resolve the technical limitations of API based experimentation and enable "deep" experimentation that not only re-ranks results but implies alternative ranking methods, "sandbox" solutions are conceivable.[49] This would mean running search applications by external developers in a protected environment ("sandbox") on datacenter infrastructure run by *Google Inc.* and other companies, allowing programmers to interface directly with the

---

48 Some authors and technologists have suggested peer-to-peer methods for solving the datacenter problem. While this is certainly an interesting direction to pursue, I remain very skeptical whether such an approach can produce the response time needed to compete with a well-maintained datacenter infrastructure. The most promising attempt, for the moment, seems to be FAROO, http://www.faroo.com.

49 Rieder, 2005

index, bypassing *PageRank* and other ranking techniques.

We can imagine several intermediate stages between the current API-based approach and a full-fledged search sandbox that vary in difficulty of implementation, economic feasibility, and their ability to effectively encourage innovation and broader access to search results. But whatever the precise approach, without proactive lawmaking any technical solution is at the mercy of service providers. French regulations on cultural diversity, which compel television companies to invest a certain percentage of their revenue in cinema production, could be seen as a rough model for thinking about how binding shared-resources requirements could look like. The project of democratizing search is difficult to imagine without at least some involvement by public authorities.

4 Conclusion

The goal of this text was to explore both the normative and practical side of how to protect the Web's capacity to be "a valuable collision space between official and unofficial accounts of reality"[50] in light of search engine mechanisms that emphasize popularity and convenience above all else. The issue is part of a larger debate on how to govern technology that is performing algorithmically, and on a very large scale, tasks that used to require human judgment such as the selection and consideration of information. When we examine mechanisms like link analysis, we find that these technical procedures do not divide the world along the same lines as we are culturally trained to do, which leads to considerable conceptual uncertainty – already at the level of analysis and even more so when it comes to normative reasoning. As we have seen, extending democratic principles to the governance of information technology is delicate business that forces us to go back to some of the basic questions of political organization. The complexity of the socio-technical configurations in search and other areas of "cultural" IT will continue to increase in the coming years, and the task of establishing clear lines of techno-normative thinking is not going to get any easier. In these pages I have tried to show how one can move from the critical analysis of technology to explicit recommendations without omitting a discussion of the political frame of reference that orients both. For this is the core of the problem: to "democratize search" we will have to incorporate a clear conceptual grasp of the technology, to reexamine our understanding of democracy, and to build bridges between these two on the levels of critique, design, and policy.

---

50 Rogers, 2004