Defining the Web: The Politics of Search Engines

Manipulated by the wealthy and the technologically adept, search engines may be presenting an increasingly distorted and limited view of the Web.

Lucas Introna
London School of Economics

Helen Nissenbaum
Princeton University

Search engines provide essential access to the Web, both to those with something to say and offer and to those who wish to hear and find. Yet many leading search engines give prominence to popular, wealthy, and powerful sites at the expense of others. Some researchers[1] have estimated that, taken individually, none of the Web search engines studied indexes more than 16 percent of the total indexable Web. Combined, the results from *all* search engines they studied increased Web coverage to only about 42 percent.

These results confirm the popular belief that the Web, at 800 million pages and growing, is almost inconceivably large, and that search engines only partially meet the desperate need for an effective means of finding things. What search engines do find they retrieve through technical mechanisms such as crawling, indexing, and ranking algorithms, and through human-mediated trading of ranking prominence for a fee.

But what about those portions of the Web that remain hidden from view? In this article, we look at how search engine developers, designers, and producers grapple with the technical limits that restrict what their engines can find. We also examine influences that may determine systematic inclusion and exclusion of certain sites, and the wide-ranging factors that dictate systematic prominence for some sites while relegating others to systematic invisibility.

Make no mistake: These are political issues.[2] What those who seek information on the Web can find will determine what the Web consists of—for them. We fear that technological limitations and commercial interests may conspire to disenfranchise those outside the mainstream and those who lack the resources or knowledge to promote their Web presence. Deprived of its diversity and impoverished by a lack of choice, a diminished Web would affect us all, individuals and institutions alike.

TECHNICAL OVERVIEW

A Web page provider who seeks search-engine recognition must focus on two key tasks: being indexed and, when its page qualifies as a valid candidate for a given search, being ranked in the top 10 to 20 results displayed.

Building an index

Being indexed is the essential first step in achieving search engine recognition. Although there are other means of reaching Web pages such as link-following and knowing or guessing universal resource locators (URLs), search engines are by far the most prominent means, especially for conducting initial exploration of a particular interest.

Search engines create a map of the Web by indexing Web pages according to keywords. From the enormous databases these indexes generate, search engines link page content through keywords to URLs. When a user who seeks information submits a keyword or phrase that best delineates the data sought, the search engine's database ideally returns a list of relevant URLs.

How then does a search engine create its database, and what does it store in it? That depends on the search engine's type. Search engines such as AltaVista, Lycos, and Hotbot use spiders, also referred to as robots or softbots, to harvest URLs automatically. In directory-based search engines, such as Yahoo and AliWeb, Webmasters and other Web page creators manually submit the vast majority of indexed pages to the search engine's editors. Their efforts are usually augmented, however, by some form of automatic harvesting.

The human touch. A directory-based search engine receives URLs from Web page creators for possible inclusion in its database. Someone who wants a page recognized by Yahoo, for example, must submit the page's URL and background information to a human editor, who reviews it and decides whether to schedule the page for indexing. The indexing software

0018-9162/00/$10.00 © 2000 IEEE

retrieves the page scheduled for indexing, then parses and indexes it according to the keywords found in the page's content. For directory-based search engines, human gatekeepers hold the keys to inclusion in their indexed databases. Given the considerable backlog caused by the continual appearance of new sites, this process can take six months from submission to inclusion.

Various criteria determine which pages get indexed. With Yahoo, for example, representatives say they use criteria of relevancy.[3] The exact nature of these criteria, however, is not widely known, publicly disseminated, or consistently applied by Yahoo's various editors. Thus, a site may be rejected without its owners being notified or given any guidance on how to overturn that decision.

Two factors seem to increase a site's chances of being listed with Yahoo:

- the number of links to and from a given site, also referred to as inlinks and outlinks; and
- how full a particular category in the directory happens to be.

Where editors feel they need more references within a category, they lower the entry barriers. Defending their approach, Yahoo representatives maintain that they list what users want, arguing that if users were not finding relevant information they would cease using Yahoo.

Along comes a spider. Being indexed by a search engine that automatically harvests URLs involves being visited by its spider. The spider usually starts its crawl from a historical list of URLs. Such lists favor documents that contain many links, such as server lists, What's New pages, and other popular Web sites. Spiders crawl the Web by automatically traversing its hypertext structure. They first retrieve a document, then recursively retrieve all documents referenced in it.

What routes these spiders follow, which sites they visit, how often, and other operational details tend to be steadfastly guarded trade secrets. At best, their guiding algorithms can be inferred from the spiders' behaviors. Junghoo Cho and colleagues have explored the nature of spider algorithms in detail.[4] Their work, highlighted in the "Page Importance Metrics" sidebar, describes the metrics that spiders commonly use to determine a Web page's importance.

Unless a query term or string steers the crawling spider, metrics must decide a page's importance. The *Backlink* metric uses a page's backlink or inlink count as an importance heuristic. The value of the backlink count equals the number of links to a specific page that appear across the entire Web. The metric derives from the assumption that the more pages that link to a given page, the greater that page's importance.

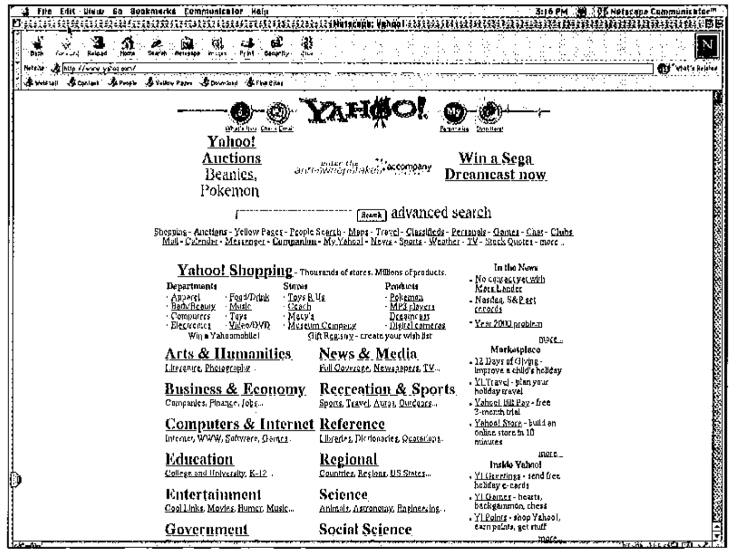


Figure 1. The Yahoo search page, with selectable search options (http://search.yahoo.com/bin/search/options/). A directory-based search engine, Yahoo draws from a database of Web pages indexed by its human editors.

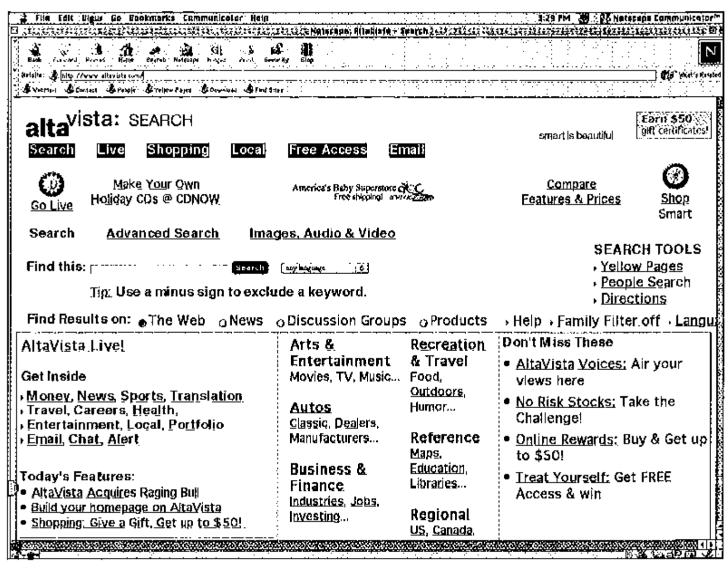


Figure 2. The AltaVista search page (http://www.altavista.com). Spiders—robots that comb the Web and record the URLs of the sites they traverse—generate AltaVista's database of Web pages.

In the cases of Excite, Hotbot, and Lycos, some evidence suggests that this metric plays a major role in determining indexing appeal. Exclusion, using this metric, is less likely for a search engine like AltaVista—which goes for massive coverage—than for its smaller and more selective competitors.

Imagine how such a metric may play out in the realm of academic research. Such research has long been underpinned by the canonical works that authors cite. We know also, however, that not all topics necessarily have canons. Further, whereas a small number of citations may make a particular work a canon in some fields, others require a vast number of citations to attain such status. Thus, the Backlink heuristic tends to gather links to large topics and fields, such as shareware computer games, in which even a tertiary site would have vastly more links than, say, those found at a localized community services Web site. Through sheer volume of backlinks, large fields determine the measure, or threshold, of importance in ways that tend to push out equally important small fields.

The *PageRank* metric exacerbates this problem. Instead of treating all links equally, this heuristic gives prominence to backlinks from pages that themselves have high backlink counts. In theory, this approach mirrors the academic practice of giving greater weight to citations from works that are considered authorities. In practice, the Backlink and PageRank metrics blindly assume that backlinks reliably indicate a Web page's importance or relevance. When page creators link to pages they deem valuable, this assumption may hold. However, many organizations actively cultivate backlinks by offering incentives such as product discounts, free software utilities, and access to exclusive information. Such incentives slant Web visibility toward those with the deepest pockets.

The *Location Metric* uses URL location information to determine a crawl's next steps: its domain type (.com, .org, .net, .edu) and whether certain keywords such as "home" appear in the URL. Presumably, programmers make such decisions when they set the crawl heuristics for a particular spider. Thus, we can conclude that spiders will target pages that

- have many backlinks, especially backlinks from other pages with high backlink counts, and
- occupy locations seen as useful or important to the crawling spider.

Another heuristic that seems to guide spiders is breadth or depth of representation. If a spider's algorithm favors breadth, it will visit more sites but index only a fraction of each. For example, spiders index only about 10 to 15 percent of large sites such as America Online or Geocities, which themselves host many smaller sites. Therefore, if your site is hosted by AOL, for example, there is a good chance it will be overlooked.

Other reasons for exclusion include presenting information on your site in non-HTML format, and the robot exclusion standard, which inserts a tag in the HTML file that specifies which robots may index a page and which ones should stay out. Most spider-based search engines do, however, allow autonomous submissions by Web masters or designers. Some software packages facilitate simultaneous submissions to search engines and automatically generate the required electronic formats.

Ranking concerns

Indexing presents but one hurdle Web page creators must clear when striving for search engine recognition. Once they succeed at being indexed, their concern shifts to ranking. Most search engines generally display up to 10 of the most relevant hits on the first page of a search's results. Knowing that few users have the time to page through more than one or two screens, Web designers jealously covet those 10 or 20 top slots.

Further, anecdotal evidence suggests that seekers are likely to look down a list, then cease looking when they find a good match for their search. A study of travel agents who use computerized airline reservations systems showed an overwhelming likelihood that they would select a flight from the first screen of search results. Such findings suggest similar behavior among Web users at large.[5]

Page Importance Metrics

Junghoo Cho and colleagues[1] have examined commonly used importance-weighting metrics in detail. According to them, Web-crawling spiders use the following algorithms to assess a page's importance. Given a Web page P, they define the importance of the page, $I(P)$, in one of four ways:

Similarity to a Driving Query. This metric uses a query term or string (Q), such as "holiday cottages," as the basic crawling heuristic. The spider need not decide the importance of a given page because the query string itself directs the spider's search.

Backlink Count. The value of $I(P)$ is the number of links to P that appear over the entire Web. We use $IB(P)$ to refer to this importance metric. Intuitively, a page P that many pages link to is more important than one that is seldom referenced.

PageRank. The $IB(P)$ metric treats all links equally: A link from Yahoo's home page counts the same as a link from some individual's personal home page. Given the Yahoo home page's much higher IB count, it makes sense to value that link more highly. Thus the PageRank backlink metric, $IR(P)$, recursively defines the importance of a page to be the weighted sum of the backlinks to it.

Location Metric. The $IL(P)$ metric asserts that the importance of page P is a function of its location, not its contents. If URL u leads to P, then $IL(P)$ is a function of u. For example, URLs that end with .com may be deemed more useful than URLs with other endings. Likewise, a URL containing the string "home" may be more interesting than other URLs. Another location metric sometimes used considers URLs with fewer slashes more useful than those with more slashes.

Reference

1. J. Cho, H. Garcia-Molina, and L. Page, "Efficient Crawling through URL Ordering," *Proc. 7th Int'l World Wide Web Conf.*, W3C World Wide Web Consortium, http://www7.scu.edu.au/programme/fullpapers/1919/com1919.htm.

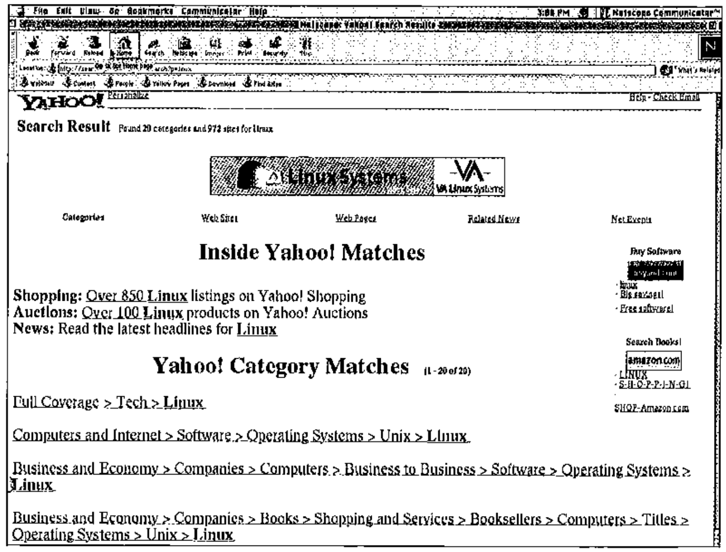
Relevancy ranking is enormously difficult. Besides the engineering challenges, experts must struggle with using a computer algorithm to approximate the complex human value of something being relevant to a person's interest. Most search-engine ranking algorithms use both the position and frequency of keywords for their heuristics: The more instances of a keyword, and the earlier in the document those instances occur, the higher the document's ranking. Other ranking schemes, like the heuristic used by Lycos, are based on *in-link* popularity. The search engine calculates the popularity score for a particular site by totaling the sites that contain links to that site. High link popularity leads to an improved ranking. As with crawl metrics, big sites determine the standard or threshold of relevance at the expense of equally relevant small sites.

The battle for ranking has generated a new discipline on how to design Web pages for greater search engine recognition, called "search engine design." It teaches design principles for optimizing a Web page's ranking, and combines these teachings with software to assess the page's ranking potential. Practices that make reasonable use of these prima facie heuristics help designers optimize their Web pages' expected ranking when they are legitimately relevant to the person doing the search.

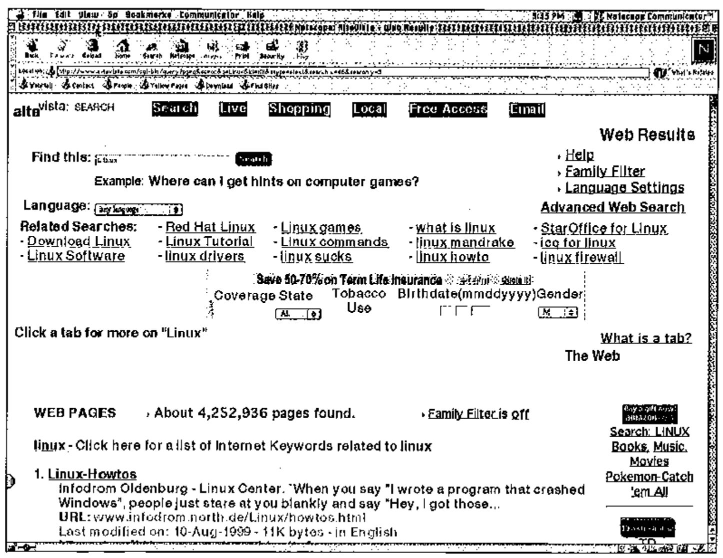
Unfortunately, the unscrupulous can use this knowledge to manipulate the ranking heuristics. Relevancy (or keyword) spamming lets Web page designers trick the algorithm into giving their pages a higher ranking. For example, ranking spammers often stuff keywords into invisible text and tiny text. Hidden from most Web users but visible to spiders, such text brims with repeated instances of keywords, thereby elevating a site's ranking relative to more scrupulous sites that restrict such keywords to legitimate usage.

This ranking warfare has created an impossible situation. Search engine operators are loath to release details of their ranking algorithms, fearing that spammers will use this knowledge to trick them. Yet ethical Web page designers legitimately need to know how to indicate relevancy to the ranking algorithm so that their pages will be listed in response to genuinely relevant searches. In addressing this problem, a team of Australian researchers has devised an ingenious method for reverse engineering the relevance ranking algorithms of various commercial search engines.[6]

Beyond the challenge of second-guessing ranking algorithms, there may yet be another, more certain, method of getting results. Some Web site producers try to find ways to buy a higher ranking—despite the indignant protests of several major search engine representatives that they do not sell search positions. Recently, however, in a much-publicized move, AltaVista and Doubleclick invited advertisers to bid for position in their top slots. Yahoo sells prominence indirectly, allowing Web owners to pay for express indexing, which moves their pages ahead in the six-month queue. Another method for buying prominence lets Web owners buy keywords that, when searched for, display both the search results and the owners' banner ads. Amazon Books, for example, has a comprehensive arrangement of this type with Yahoo, as does Barnes & Noble with Lycos.



(a)



(b)

Figure 3. Typical first-page search results for (a) Yahoo and (b) AltaVista.

Biased search results

Having established that search engines mediate much of what information seekers on the Web experience, we can predict the following:

- Seekers will most likely find large, popular sites whose designers have enough technical savvy to succeed in the ranking game.

The Importance of Domain Names

An effective system for finding and retrieving individuals' and institutions' Web sites, based on content and relevancy, may prove a constructive response to the debate currently raging over the importance of domain names. The desirability of the more intuitive and easily recognizable domain names increases in inverse proportion to the efficacy of available search mechanisms. A highly effective indexing and retrieval mechanism can thus avoid domain name assignments that favor some while penalizing others.

Some have argued that the increasingly popular portal sites provide another alternative for finding Web pages. Perhaps so. We think it highly unlikely, however, that a link established through a portal site would fail to meet the indexing criteria for search engines.

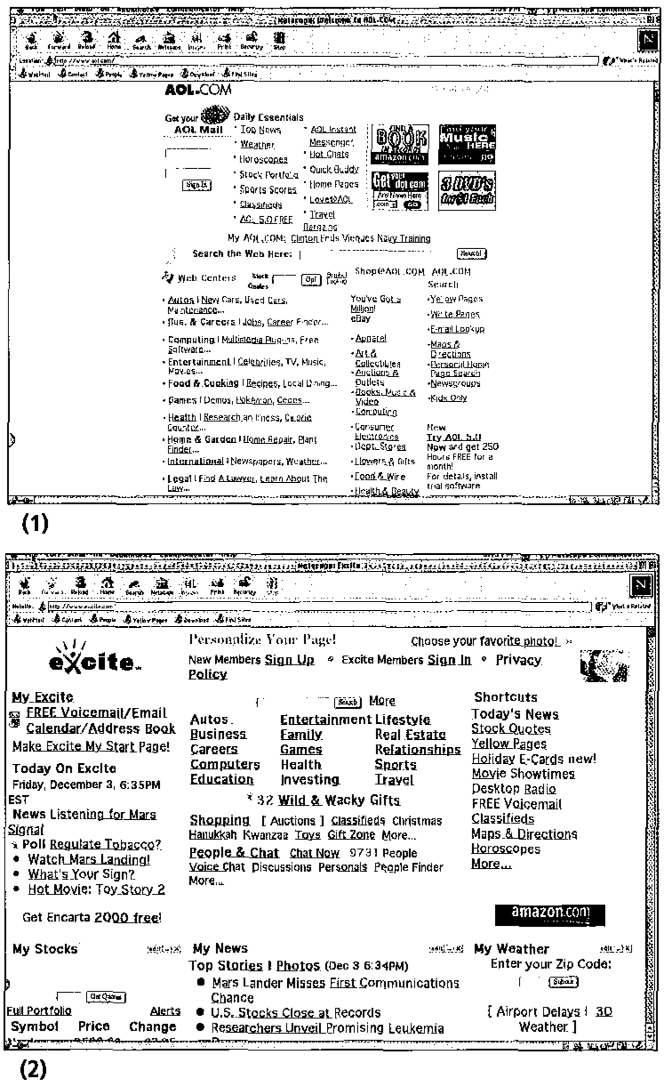


Figure A. Portal sites such as (1) AOL and (2) Excite may provide an alternative to search engines for finding Web pages.

- Sites whose proprietors can pay for various means of improving their positioning will attract even more attention.
- On the other hand, seekers will have trouble finding the less popular and smaller sites, especially if less knowledgeable professionals support those sites. When a search does yield these sites, they will tend to be listed far down in the results rankings.

The accuracy of these predictions will vary generally according to the seeker's competence, and specifically according to the keywords or phrases used to initiate a search. The search engine used will also shape the seeker's experience. Some users may consistently choose one search engine over others, while some will simply use the default engine provided by their institution or Internet service provider. Because they base their results on existing search engines and normally accomplish their task by recognizing only higher-order search keys than first-order engines, Metacrawler, Ask Jeeves, and Debriefing will likely support our predictions. Not only are most users unaware of search engines' particular biases, they seem to be unaware that they are unaware of them.

SHOULD MARKET DYNAMICS SHAPE THE WEB?

Some may argue that we should let the evolving marketplace in search engines be governed by free-market dynamics. As Yahoo's representatives have argued, users' reactions must remain the benchmark of quality: Dissatisfied seekers will defect from an inadequate search engine to competitors that index and prioritize better. Although we hesitate to ignore the wishes of users so cleanly reflected in their market choices, two main concerns prompt us to question the social implications of an Internet dominated by market dynamics.

First, the marketplace we see evolving from the current situation would fall too short of the ideal competitive free market. Defenders of the market mechanism frequently claim that participants are free to express their preferences through the choices they make among alternatives. In the case of search engines, however, most Web users lack critical information about alternatives. Relatively few users understand how search engines work and by what means they yield their results. Thus it is misleading to suggest that these users can meaningfully express preferences or exercise free choice.

Second, we believe that Web search mechanisms are too important to be shaped by the marketplace alone. The current, commercial model for search engines rewards with the greatest popularity those that cater best to majority interests. Market diversity would undoubtedly force a degree of comprehensiveness and objectivity in listings. But where is the market incentive to list sites of interest to small groups, such as esoteric collectors, specialist researchers, or the economically disadvantaged?

Commercial bias

Popularity with seekers is not, however, the only force that shapes search engine listings. Entities that wish to be found also exert a force—one subject to enormous inequality. Some enter the market already wielding vastly greater economic prowess and power than others. The rich and powerful clearly can influence search engine tendencies; their dollars can, and in some ways already do, play a decisive role in what sites a given search retrieves. The cost to a search engine of losing a small number of customers may be outweighed by the benefits of pandering to majority interests and to entities who pay for some form of enhanced visibility. We can expect, therefore, that those who wish to be found will cause at least some drift, which, in turn, would further narrow the field of what is available to seekers of information, association, support, and services. Engines that use link popularity for priority listing will be even more prone to reifying a mode of conservatism on the Web.

It may be useful to think of the Web not as a single market, but as a market of markets. When we seek, we are not interested in information per se, but rather in data related to our specific interests and needs. Seekers might be in the market for information about, for example, packaged tour holidays or computer hardware suppliers. For these markets, where we expect the demand for information to be great, we would expect the competition for recognition to be great as well. Companies would pay high prices for the keyword banners that will ensure them the top spot, and a search will generate many hits for the seekers. In contrast, there are other, significantly smaller markets for information about a rare medical condition, or about services of a local government authority or community.

Pareto's law

In this market of markets, there is likely to be little incentive to ensure the inclusion of these small markets, and only a small cost in loss of participation for their exclusion. Although we do not have empirical evidence, we would expect Pareto's law[7] to apply: A high percentage of search requests, say 80 percent for argument's sake, are directed to a small percentage, say 20 percent, of the big markets, which would be abundantly represented in search results. Only a small percentage of the search requests, say 20 percent, might be addressed to the large percentage, say 80 percent, of the smaller markets, which would be underrepresented. This scenario explains the limited incentive for inclusion and relatively low cost of exclusion.

A market enthusiast does not find this result problematic because it describes exactly what the market is supposed to do. The range and nature of choices are supposed to ebb and flow in response to the ebb and flow of market participants' wants and needs. Nevertheless, we resist this conclusion not because we are suspicious of the marketplace in general, but because maintaining the Web's variety of options is of special importance.

We think that the value of comprehensive, thorough, and wide-ranging Web access lies within the category of goods that the political theorist, Elizabeth Anderson, argues should not be left entirely if at all to the marketplace.[8] A liberal democratic society such as ours, firmly committed to various principles having to do with freedom, welfare, and autonomy, chooses to sustain various goods even if they are not sustainable via a market mechanism. Reasonable contenders include such things as public parks, artistic and historic treasures, schools, reproductive capacities, addictive drugs, and persons themselves. Such goods should be distributed not in accordance with market norms but "in accordance with public principles."[8] We see an equivalent need in the case of search engines.

Our argument against leaving search engines fully to the mercy of the marketplace is not, however, a purely formal one. The trajectory of search engine development is not wrong or politically dangerous in itself; rather it undermines a particular, normative vision of the Web in society. Those who do not share this vision are unlikely to be convinced that search engines are different in kind from, say, salad dressings and automobiles. The case that search engines are a special, political good presumes that the Web, too, is a special good.

The Web has earned its greatest approbation as a political good by functioning as a medium for intensive communication among and between individuals and groups.

THE WEB AS A PUBLIC GOOD

Search engines, functioning in the manner we've outlined, raise political concerns. They do so not simply because of how they function, but because how they function seems at odds with the compelling ideology of the Web as a particular kind of public good: a rich array of commercial, political, and artistic activity that fosters associations and communications of all kinds, and provides a virtually endless supply of information.

Over the past decade, first the Internet and now the Web have come to be perceived as a great public good. For example, when only a fraction of the US population possessed Internet access, US Vice President Al Gore promoted the vision of a Global Internet Infrastructure that promised great economic gains, global cooperation, the spread of political freedoms, and other benefits. This concept—part reality, part wishful thinking—has gripped scholars, engineers, scientists, entrepreneurs, politicians, and many others. Each sector has highlighted a particular dimension of the Web's promise, some focusing on information, some on communication, some on commerce, and so on.

The ideal Web would extend the possibilities for association, facilitate access to obscure information sources, and give voice to many of the typically unheard.

A public space

Most versions of the Web's promise assume that it contributes to the public good by serving as a special kind of public space. One characteristic that pushes something into the public realm is a lack of private ownership. The Web does seem to be public in this sense: No person, institution, or even nation wholly owns its hardware and software infrastructure. Nor, given its global distribution, does it come under the territorial jurisdiction of any existing sovereign state.[9]

The Web also fulfills some functions of other traditional public spaces such as museums, parks, beaches, and schools. It serves as a medium for artistic expression, a space for recreation, a place for storing and exhibiting items of historical and cultural importance, and a resource for education.

Yet the Web has earned its greatest approbation as a public space and a political good by functioning as a medium for intensive communication among and between individuals and groups in nearly any permutation imaginable. It is the Hyde Park Corner of the electronic age, the public square where a community may gather as a whole, or associate in smaller groups. They may talk and listen, they may plan and organize.

Such spaces, where content is regulated by only a few fundamental rules, embody the ideals of a liberal democratic society. The Web's use as public space and forum for political deliberation has fueled ongoing discussions of teledemocracy. Although scholars have reached no universal agreement about what influence the Web may have on politics, several contributors to the debate have cited cases in which the Web appears to have had a decisive impact on a specific political situation.

Douglas Kellner[10] cites the Web's role in aiding the Zapatistas struggle against the Mexican government, the Tianamen Square democracy movement, environmental activists who exposed McDonald's through the McLibel campaign, and the Clean Clothes Campaign support for Filipino garment workers' attempts to expose exploitive working conditions.

An information conduit

Above all, the Web qualifies as a public good because it conveys information. In this so-called Information Age, some consider being among the information-rich so important that, like philosopher Jeroen van den Hoven, they argue it makes sense to construe access to information as a Rawlsian "primary good," thus compelling any just society to guarantee a basic or reasonable degree of Web access to all its citizens.[11]

Growing use of the Web as a repository for information such as government documents, consumer goods, scientific and artistic works, and local public announcements lends increasing weight to this prescription. Accordingly, the Web is not a vehicle for further expanding the gap between haves and have-nots, but one for narrowing it.[12,13]

This positive view of the Internet has fueled much of the social and economic investment in the medium and its supporting technology. It has convinced both progressive politicians and those who want to appear progressive to support it with funding and political backing.[13] Idealistic computer scientists and engineers have volunteered energy and expertise toward developing and promulgating the Web's hardware and software, from the likes of Jonathan Postel, an early builder of the Internet, who worked to keep its standards open and free, to professionals and researchers who volunteer to wire schools and help build infrastructure in poorer nations.

Web creators like Tim Berners-Lee have been very much aware of these inclusive values from the start:

> The universality of the Web includes the fact that the information space can represent anything from one's personal private jottings to a polished global publication. We as people can, with or without the Web, interact on all scales. By being involved on every level, we ourselves form the ties which weave the levels together into a sort of consistency, balancing the homogeneity and the heterogeneity, the harmony and the diversity. We can be involved on personal, family, town, corporate, state, national, union, and international levels. Culture exists at all levels, and we should give it a weighted balanced respect at each level.

If trends in search engine design and function lead to a narrowing of options—either an actual narrowing or a narrowing in what can be located—the Web as the kind of public good many envisioned will be undermined.

A universal forum

So far we have discussed these effects from the seekers' perspective, as a limiting of opportunities to locate various types of information, individuals, and organizations, a narrowing of the full range of deliberative as well as recreational capabilities. Yet even more is at stake: Web access for those who would like to be found, seen, and heard. The public good of the Web lies not merely in its role as a place for seekers to find things, but as a forum for those with something to offer. Those excluded from search results because their lower ranking deprives them of attention or recognition may well offer just as much value as do those who appear on the "pages found" screen. We lose twice over in this case: first, because continuing invisibility may cause options to atrophy, thinning the field of opportunity and, second, because the Web fails to serve many who reach out for attention or connection.

The ideal Web serves all people, not just some, and not merely those in the mainstream. The Web's potential for inclusivity and breadth are precisely what energized many to think this technology would offer more than a new tool for entrenched views and powers. The ideal Web would extend the possibilities for association, facilitate access to obscure information sources and give voice to many of the typically unheard, and preserve intensive and broadly inclusive interactivity.

Many have observed that for the Web to become a democratizing technology and a public good, we must first take the question of access seriously. We agree, but would define the question in broader terms. Access is not merely a computer and a network hookup, even when coupled with the skills and know-how that enable effective use. Rather, access implies a comprehensive mechanism for finding and being found. Thus our concern with the politics of search engines—a politics that at present seems to push the Web in a direction that favors special interests at the expense of marginalizing the general public.[14]

DEMOCRATIZING SEARCH ENGINES

To ensure a Web that does not favor the wealthy, the unscrupulous, and the technologically proficient, we need more than scrutiny and discussion, we also need policy and action. We advocate a combination of regulation through public policy and value-conscious design innovation.

Regulating and restricting development of commercial search engines, however, would be neither practically appealing nor wise, and might smack of cultural elitism or paternalism. Rather, we propose that Web search engine capabilities be enhanced and refined—a prescription that echoes Amartya Sen's reaction to current economics: "It is not my purpose to write off what has been or is being achieved, but definitely to demand more."[15]

Promoting inclusiveness

As a first step, we would demand full and truthful disclosure of the underlying algorithms that govern indexing, searching, and prioritizing, stated in a way meaningful to most Web users. Although such information might help spammers, we argue otherwise. Would not the impact of spammers' unethical practices be severely dampened if both seekers and those wishing to be found became aware of the particular biases inherent in any given search engine? We believe that informing users, on the whole, will be better than maintaining the status quo. Those who favor a market mechanism may be pleased to note that disclosure would move us closer to fulfilling the criteria of an ideal competitive market in search engines.

Disclosure is a step in the right direction. But disclosure, by itself, may not sustain and enhance Web offerings in the way we would like: by retaining transparency for those less popular sites to promote inclusiveness.

The marketplace alone will not ensure such transparency. As a policy step, we should consider public support for developing more egalitarian and inclusive search mechanisms, and for research into search and metasearch technologies that would increase transparency and access. Although these and other policies promise a fairer representation of Web offerings, a second key lies in the technology itself.

We hope to inspire among designers and builders of search-engine technology the value of fairness.

Reflecting social values in system design

Philosophers have recognized the intricate connection between technology and social, political, and moral values. That technological systems may embed or embody values resonates in the social and political commentary on information technology written by engineers, philosophers, and cyberlaw experts. Translating their ideas into practice implies that we can build systems that better reflect important social values if we build them with an explicit commitment to such values. We hope to inspire among designers and builders of search-engine technology the value of fairness and the suite of values that comprise the ideology of the Web as a public good.

The two leading approaches to achieving this goal do have some drawbacks. The first, to associate search engines with particular sectors of society, may increase segmentation and diversification by drawing borders according to traditional categories such as sports, entertainment, and the arts. One problem with segmentation is that it could fragment the very inclusiveness and universality we value. Eventually, a segmented Web may merely mirror societal institutions and their baggage of asymmetrical power structures, privilege, special interests, and so forth.

The second approach calls for the development of individualized robots that search for pages based on individual criteria, then build individualized databases according to individual needs. Given the extensive overhead this option imposes, we may lack the resources to implement it: Automatic harvesting via even the existing robot population already consumes resources extravagantly enough to cause concern.

Beyond the policies and actions we've advocated, several technological developments could, in principle, help search engines build a more egalitarian portrait of the Web. These developments include improving how individual pages indicate relevance (also referred to as metadata), refining overall search engine technology, and improving Web resource presentation and visualization and metasearch technology.

Although improvements like these might, accidentally, promote particular values, they hold greatest promise as remedies to the current politics of search engines if values guide them explicitly. We urge engineers and scientists who adhere to the ideology of the Web—especially its inclusivity, fairness, and scope of representation—to pursue improvements in indexing, searching, accessing, and ranking with these values firmly in mind. Given its rapid growth and growing influence, the Web will play an increasingly important social role. The struggle to chart it and capture the attention of its information seekers is thus not merely technical but also political. ❖

Acknowledgments

This article benefited from the generous help of others. We thank the audience at the 1998 Conference on Computer Ethics: Philosophical Enquiry, where we first presented the paper upon which this article is based; Eszter Hargittai and Steven Tepper for extensive comments; Lee Giles, Brian LaMacchia, Andrea LaPaugh (and members of her graduate seminar) for astute comments and technical guidance; and Michael Cohen and Sayumi Takahashi for their able research assistance. Nissenbaum gratefully acknowledges the support of National Science Foundation grant SBR-9806234.

References

1. S. Lawrence and C.L. Giles, "Accessibility of Information on the Web," *Nature*, No. 400, 1999, pp. 107-109.
2. L. Winner, "Do Artifacts Have Politics," *Daedalus*, No. 109, 1980, pp. 121-136.
3. V. Phua, "Towards a Set of Ethical Rules for Search Engines," MSc dissertation, London School of Economics, London, 1998.
4. J. Cho, H. Garcia-Molina, and L. Page, "Efficient Crawling through URL Ordering," *Proc. 7th Int'l World Wide Web Conf.*, 1998, http://www7.scu.edu.au/programme/fullpapers/1919/com1919.htm.
5. B. Friedman and H. Nissenbaum, "Bias in Computer Systems," *ACM Trans. Information Systems*, Vol. 14, No. 3, July 1996, pp. 330-347.
6. G. Pringle, L. Allison, and D.L. Dowe, "What Is a Tall Poppy among Web Pages?" *Proc. 7th Int'l World Wide Web Conf.*, 1998, http://www7.scu.edu.au/programme/fullpapers/1872/com1872.htm.
7. A. Sen, "The Moral Standing of the Market," *Social Philosophy & Policy*, Vol. 2, No. 2, Spring 1985, pp. 1-19.
8. E. Anderson, *Value in Ethics and Economics*, Harvard University Press, *Cambridge, Mass. and London*, 1993.
9. D.R. Johnson and D. Post, "Law and Borders—The Rise of Law in Cyberspace," *Stanford Law Review*, Vol. 48, No. 5, 1996.
10. D. Kellner, "Intellectuals, the New Public Spheres, and Techno-Politics," http://www.gseis.ucla.edu/courses/ed253a/newDK/intell.htm.
11. J. van den Hoven, "Towards Ethical Principles for Designing Politico-Administrative Information Systems," *Informatization and the Public Sector* (now *Information Infrastructure and Policy*), Vol. 3, 1994, pp. 353-373.
12. R. Civille, "The Internet and the Poor," *Public Access to the Internet*, B. Kahin and J. Keller, eds., MIT Press, Cambridge, Mass., 1996.
13. D.L. Hoffman and T.P. Novak, "Bridging the Racial Divide on the Internet," *Science*, Vol. 280, 17 Apr. 1998, pp. 390-391.
14. L. Lessig, "The Law of the Horse: What Cyberlaw Might Teach," *Harvard Law Review*, to appear.
15. A. Sen, *On Ethics and Economics*, Blackwell, Oxford, UK, 1987.

***Lucas Introna** lectures on information systems at the London School of Economics and Political Science and is Visiting Professor of Information Systems at the University of Pretoria. His research interest is the social dimensions of information technology and its consequences for society. He is associate editor of* Information Technology & People *and co-editor of* Ethics and Information Technology. *His practical and consulting experience includes design and implementation of information systems for various large corporations as well as strategic management consulting. He is an active member of IFIP WG 8.2, The Society for Philosophy in Contemporary World (SPCW), International Sociological Association WG01 on Sociocybernetics, and several other academic and professional societies. Introna received degrees in management, information systems, and philosophy. Contact him at l.introna@lse.ac.uk.*

***Helen Nissenbaum** is a research associate at the Center for Human Values, Princeton University, where she teaches and writes in the area of social, political, and ethical dimensions of information and communications technology. Her publications have addressed issues such as privacy, accountability, trust, and values in system design. She is one of four founding editors of the new journal* Ethics and Information Technology. *Nissenbaum received a PhD in philosophy from Stanford University. Contact her at helen@princeton.edu.*