University of Dayton Law Review

Volume 26 | Number 2 Article 3

1-1-2001

# Let the Crawlers Crawl: On Virtual Gatekeepers and the Right to Exclude Indexing

Niva Elkin-Koren
*University of Haifa School of Law*

# Let the Crawlers Crawl: On Virtual Gatekeepers and the Right to Exclude Indexing

## Cover Page Footnote

I thank Michael Birnhack, Michael A. Einhom, Peter Hirtle, Helen Nissenbaum, Eli Salzberger, Haim Ravia, and Alfred C. Yen for their valuable comments. I also wish to thank the participants of the faculty colloquium at IDC and the participants of the Interdisciplinary Conference on the Impact of Technological Change on Intellectual Property at Ohio State University in which portions of this paper were presented. I am grateful to Matan Goldblat and Yael Bregman for their research assistance.

# SYMPOSIUM ARTICLES

# LET THE CRAWLERS CRAWL: ON VIRTUAL GATEKEEPERS AND THE RIGHT TO EXCLUDE INDEXING

*Niva Elkin-Koren*

TABLE OF CONTENTS

# LET THE CRAWLERS CRAWL: ON VIRTUAL GATEKEEPERS AND THE RIGHT TO EXCLUDE INDEXING

*Niva Elkin-Koren*[*]

## I. INTRODUCTION

The high hopes associated with the Internet for enhancing freedom of speech and personal autonomy were tied to its decentralized nature. Technically, anyone connected to the net could post their content and make it available through various channels to a massive number of people. For "speakers" the net was a powerful platform for efficiently expressing their views or distributing content to the public at large. Yet, the proliferation of information introduced a new challenge of how to capture users' attention. While posting on the web may be inexpensive and easy, making your content noticeable and detectable by users becomes extremely competitive. Since it is no longer feasible to restrict the available content on the web, information providers increasingly seek to directly control users' attention.

The proliferation of information in cyberspace challenges not only speakers but also "users." While in the past individuals suffered from lack of information that was necessary to perform their transactions or exercise their political rights, cyberspace creates an information overflow. Consequently, users are increasingly dependent on technical means that allow retrieval and selection of relevant information from the large bulks of available data. Search engines were developed to serve this function.[1]

Search engines are becoming the new virtual gatekeepers of cyberspace. Information that is undetectable, or otherwise remains unlisted on the

---

[*] Senior Lecturer, University of Haifa School of Law. J.S.D. Stanford Law School, 1995; LL.M. Harvard Law School, 1991; LL.B. Tel-Aviv University, 1989. I thank Michael Birnhack, Michael A. Einhorn, Peter Hirtle, Helen Nissenbaum, Eli Salzberger, Haim Ravia, and Alfred C. Yen for their valuable comments. I also wish to thank the participants of the faculty colloquium at IDC and the participants of the Interdisciplinary Conference on the Impact of Technological Change on Intellectual Property at Ohio State University in which portions of this paper were presented. I am grateful to Matan Goldblat and Yael Bregman for their research assistance.

[1] The first search engine, Archie, was developed by Alan Emtage, a student at McGill University in 1990, before the World Wide Web was created. *See* WES SONNENREICH & TIM MACINTA, WEB DEVELOPER.COM GUIDE TO SEARCH ENGINES 1-2 (1998) [hereinafter WEB DEVELOPER.COM].

search results, is almost nonexistent on the web. In an age of information overload, control over such virtual gatekeepers may give one the power to filter information, that is, to locate some information while avoiding other information. Search engines thus become a focal point for controlling access to information, and consequently they turn out to play a key role in shaping the information environment.

Control battles in the information market are no longer confined to commercial competition, but have recently reached the courts. eBay, the largest online person-to-person auction site, objected to the unauthorized use of its database by metacrawlers and data aggregators. Bidder's Edge, an auction aggregation site, allowed its users to search for items simultaneously across numerous online auctions, by conducting a single search and without having to search each auction site individually.[2]

Like many search engines, eBay prohibited the automated use of its database in its users' agreement, and further employed technical means to block access to Bidder's Edge. Eventually eBay brought a suit against Bidder's Edge and was granted a preliminary injunction enjoining Bidder's Edge from accessing eBay's systems by use of any automated querying program without eBay's written authorization.[3]

The reasons behind eBay's objection to Bidder's Edge activity are intriguing. Like many search engines and data aggregators, Bidder's Edge was not directly competing with eBay. eBay is a person-to-person trading site that allows sellers to list items for sale, and potential buyers to search the listings and bid on items. eBay collects commissions for sales. In fact, eBay's subscribers could benefit from the use of metacrawlers such as Bidder's Edge since it would increase their listings' exposure. A greater number of potential bidders would tend to increase the likelihood of closing transactions at a higher price.[4] Such contingency could potentially increase eBay's revenues, since eBay's commission is based on the closing

---

[2] *See* eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1061 (N.D. Cal. 2000). Bidder's Edge created a database of information compiled from various auction sites. *See id.* at 1061-62. As of March 2000 the site had compiled information on items auctioned on more than 100 auction sites. *See id.* at 1061.

[3] *See eBay*, 100 F. Supp. 2d at 1063, 1073. Similarly, mySimon (<http://www.MySimon.com/>), a popular Internet shopping search engine, sued Priceman.com meta-search engine, which was searching shopping engines, for using its search results and infringing its copyright. mySimon claimed that its search results are reprinted on the site with no attribution to the source. *See* Brian Banks, *Builder Beware*, CANADIAN BUSINESS, Oct. 29, 1999, available at <http://www.canadianbusiness.com/magazine_items/1999/oct22_99_builderbeware.shtml>.

[4] Indeed, posting the items for sale on data aggregators may normally introduce competition with sellers from other sites and push down the price. Yet, due to the method of auction such competition would not necessarily lower the closing price of any particular deal.

price.[5] If eBay's goal was to maximize profits, why did eBay object to the use by Bidder's Edge? eBay sought to preserve its dominance in the online auction industry.[6] If metacrawlers like Bidder's Edge are crawling around then eBay may loose its key assets. These assets include its control over the community of users that occupies its site since they will no longer be captured and restricted to a single site. Another business advantage is the power to shape preferences and affect actual choices made by users, by determining what alternative transactions are made available to them. Finally, there were opportunities for branding and establishing a distinctive reputation. Owning a powerful online brand would usually allow sellers to charge a higher price for identical products. When users are searching via bargain finders, competition turns to focus on price alone.[7] Under such circumstances there are less opportunities for cashing the benefits of branding. To secure its brand, eBay objected to the display of its results next to those of its competitors.

The fact that eBay sought protection of its commercial expectations does not entail, however, that it is legally entitled to do so, nor does it follow that the law should protect such expectations. This paper examines the ramifications of granting a right to exclude the use of search results and otherwise restrict the operation of search engines.

The newly created right to exclude indexers, established by the *eBay* decision, is strengthening a trend of propertizing information in recent years. The eBay rule opens new opportunities for accumulating control and interfering with competition in the search engine market. Such legal rights could shape our information environment on the net. A noncompetitive market of virtual gatekeepers could compromise open access and cause inefficiencies in electronic commerce. The law should therefore facilitate competition in the search engine market.

---

[5] *See Fees* (visited Jan. 22, 2001) <http://www.pages.ebay.com/help/sellerguide/selling-fees.html>. In addition to an insertion fee, eBay.com charges a Final Value Fee—between 1.25% to 5% of the final sale price. *See id.*

[6] *See* Steven Bonsiteel, *EBay Hints Other Auction Aggregators Could Be Blocked - Update*, NEWSBYTES (Nov. 4, 1999) <http://www.newsbytes.com/news/99/138931.html> (citing eBay's spokesman Kevin Pursglove). eBay further argued that it sought to protect its interest in the auction data compiled through its relationship with its customers. *See id.*

[7] *See* J. Bradford DeLong & A. Michael Froomkin, *Speculative Microeconomics for Tomorrow's Economy* (last modified Nov. 22, 1999) <http://personal.law.miami.edu/~froomkin/articles/spec.htm>.

## II. SEARCH ENGINES AND THE CONSTRUCTION OF KNOWLEDGE: THE VIRTUAL GATEKEEPERS OF CYBERSPACE

The decentralized nature of the Internet was considered one of its most significant characteristics that promised to turn it into a concrete alternative to existing content markets. The network infrastructure, it was thought, would replace centralized distribution methods employed by broadcasters and publishers with a potentially more decentralized flow of information. Technically, anyone connected to the net can post their content and make it available through various channels. Entry barriers in the online information market are presumably low. Not only is the cost of distribution significantly low, but the cost of production is also going down. Indeed, the production of content still involves the high cost of human skills, but means of production are cheaper than they used to be.[8] Furthermore, unlike broadcasters, online providers are not required to get a license before posting on the net or setting up a distribution channel.

The potentially decentralized nature of the Internet increases pressure on the legal system to produce means of market control. In the absence of a central bottleneck in the infrastructure, market players increasingly rely on legal rights for exercising control over information markets and protecting market domination. But what could be a focal point of control in cyberspace?

In the past, control over distribution channels (such as television stations or motion picture studios) was central for acquiring control over content. The reason was that whoever controlled the distribution channels could precondition distribution in the assignment of intellectual property rights.[9] Nowadays, however, controlling the "pipes" would be insufficient for market domination. A wide range of competing technologies facilitate access to the Internet. There is a vigorous competition between different types of infrastructure and means of communication such as telephony, cables, cellular telephony, and satellite transmission.

Furthermore, while production and distribution means of information are relatively available and inexpensive, new barriers on entry are introduced. Even though anyone can make information available on the net, only few are noticed. The greatest challenge for companies competing in the content market is thus to capture users' attention. The bulk of

---

[8] It is no longer necessary to own a fancy studio, for instance, in order to record a song. The necessary equipment is now more widely available, and at a reasonable price.

[9] *See* RONALD V. BETTIG, COPYRIGHTING CULTURE - THE POLITICAL ECONOMY OF INTELLECTUAL PROPERTY 35 (1996) ("For to get 'published,' in the broad sense, actual creators must transfer their rights to ownership in their work to those who have the means of disseminating it.").

information available for users cause an attention deficit. As stated by Herbert Simon: "[A] wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."[10] All types of informational products are competing for the scarce attention of the audience,[11] and users' attention becomes a valuable asset. Selling viewers' attention had long been the business model of advertisers' supported radio and television broadcasts.[12] Television viewers are captured, exercising a limited choice among channels, the content of which is determined by the owners. Consequently, content providers who sought greater control over viewer's choices were induced to own more channels. Vendors in the entertainment industry managed viewers' attention by controlling the number of releases, the timing of shows, and the frequency in which access to old releases becomes available. The need to manage viewers' attention in those markets pushed for consolidation and concentration of ownership.

Managing users' attention in cyberspace poses a different challenge. When everyone can put content on the web, and users presumably exercise more control over their choices, competition over users' attention becomes more aggressive. Restricting the availability of content in cyberspace is impractical. Information originated by numerous sources is prohibitively expensive and difficult to manage.[13] Therefore, rather than controlling the posted content, commercial providers are increasingly seeking means that would directly control users' attention. Indeed, when the problem turns out to be information overload, rather than access, the real value rests in locating and filtering information that would be relevant to users.[14] Consequently, location tools that make content detectable, noticeable, and available for retrieval by users become a new focal point of control.

Search engines cannot keep out undesired materials, but they can nevertheless effectively control access to information. If you are not listed

---

[10] Herbert A. Simon, *Designing Organizations for an Information-Rich World*, *in* COMPUTERS, COMMUNICATIONS, AND THE PUBLIC INTEREST 37, 40-41 (Martin Greenberger ed., 1971).

[11] That is especially true for the entertainment industry. Every film competes with all other films released into the marketplace. *See* Eileen R. Meehan, *"Holy Commodity Fetish, Batman!": The Political Economy of a Commercial Intertext*, *in* THE MANY LIVES OF THE BATMAN 47 (Roberta E. Pearson & William Uricchio eds., 1991).

[12] For a critical review of the role played by mass advertising in the newspapers and broadcasting industries see BEN H. BAGDIKIAN, THE MEDIA MONOPOLY 105-192 (6th ed. 2000).

[13] At the same time, however, the Internet opens new opportunities for audience targeting. *See* Patrick Fair, *Annex: Advertising on the Web - The Business View*, *in* WEBVERTISING: UNFAIR COMPETITION AND TRADEMARKS ON THE INTERNET 249, 252-54 (Matthias W. Stecher, ed., 1999) (reviewing various methods of online targeting).

[14] *See* CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY 6 (1999) [hereinafter INFORMATION RULES].

in the search results you are almost nonexistent on the web. Indeed, we sometimes look for a particular vendor or a certain website, and in fact, the more specific the query gets (such as a particular URL), the less it is necessary to use a search engine. In the majority of cases, however, users navigate with no particular destination, and they are often unaware of the URL they seek.

Search engines are often described as devices that allow users to find information they value and to avoid the rest.[15] Yet, this description is somewhat simplistic and naive. It assumes that users already know what information they want and what information they value prior to the search. More often, however, users do not know what information they need. In fact, to a great extent, what users consider as valuable may be highly dependent on the search results.

Search engines thus turn out to be one of the most significant players in the information economy. They could play a key role in structuring the online market for goods and services by granting prominence to some brands while avoiding others. The search results would determine what is available, from which a user can make her choice. In this sense, search engines function as virtual gatekeepers and could considerably affect the available options for online consumers, and thus, their actual choices.[16] Decentralized competition among search engines is therefore essential for keeping a competitive market in electronic commerce.[17]

Furthermore, in addition to their extraordinary commercial power, search engines may also have a political dimension. By effectively constructing meaning and shaping choices, search engines could be politically significant to social relations and the shaping of public opinion.[18]

Search results do not simply locate materials. They construct meaning. They affect the organization and meaning of information. They structure

---

[15] *Id.* at 6.

[16] For instance, in commerce a search engine could locate only resources associated with a particular owner or affiliated companies. It can systematically avoid resources that belong to competitors or more likely push them to the bottom of the list. In a political context, while a newspaper may have an ethical code that requires it to provide an opportunity to express both sides, a search engine could bring only one.

[17] Introna and Nissenbaum argue that competition in the market is actually not enough to guarantee access to the Web. That is because search engines playing by the market rules would give prominence to popular and wealthy sites at the expense of others, either through their algorithms or by selling prominence for a fee. *See* Lucas D. Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 THE INFORMATION SOC'Y 169, 176-78 (2000) [hereinafter *Shaping the Web*].

[18] *See id.* at 178-81 (claiming that search engine design is not only technical, but also a political matter, and arguing, therefore, that regulation of search engines should not be left to the marketplace).

categories in response to users' queries, and thereby have the capacity of creating categories for grasping the world.[19] By defining which information becomes available for each query, search engines may shape positions, concepts and ideas.[20]

Indeed, the high hopes raised by cyberspace for changing the information landscape require open access. Yet, in a world of virtual gatekeepers, physical access would be insufficient.[21] The fact that it is cheaper and easier to get published by posting your opinion on the web does not guarantee that your opinions will ever be heard. Access to the web could be opened, but if information is not detected by a search engine then it could be easily lost forever in the endless bulk of information on the net. To allow individual users on the web not only to freely post their opinions, but also to be heard and read, it would be necessary to actively support decentralized search methods.[22] This could only be achieved in a well-functioning, competitive market for search engines that is free of proprietary rights. This requires exercising legal policies that would facilitate competition among search engines and minimize attempts to centralize control.

---

[19] *See* Elinor Rosch, *Cognitive Representations of Semantic Categories*, 104 J. OF EXPERIMENTAL PSYCHOL. 192 (1975) (pioneering research in cognitive psychology on the formation of real world categories).

[20] Concepts are fundamental units of thought which function to organize objects, events and relations in our physical and mental world. Concept formation may be affected by search results. Consider for instance an elementary school student who is searching the web for materials on "Jews." The three top results retrieved by Excite were: The Christian Jew Foundation (a messianic missionary organization which presents the gospel to the Jewish people), Jew Watch (a hate site allegedly reporting Jewish monopoly, banking, and media control worldwide), HAIKUS FOR JEWS, (humorous haikus combining ancient Zen wisdom with timeless Jewish noodging). A different input is provided by the search results retrieved by HotBot. The top three results were: Judaism 101 (an encyclopedia on information about Judaism and Jewish practices), Jewishnet (global Jewish information network), and, as in Excite, HAIKUS FOR JEWS. A student having no significant previous knowledge on the subject would end up forming different concepts based on the different search experiences.

Searching for the word "Piracy" may entail similar differences. The top three results on Excite were: BSA Anti-Piracy-Site, Worldwide Piracy Initiative, and SPA Anti-Piracy. The same search at HotBot retrieved: Alliance Against Counterfeiting & Piracy, Maritime Security and Counter-Piracy Research Site, and Piracy is Your Friend (a manifesto for musicians who want to make money in the new economy).

[21] *See Shaping the Web*, *supra* note 17, at 180-81 ("Access involves not merely a computer and a network hookup, as some have argued, nor, in addition, the skills and know-how that enable effective use. Access implies a comprehensive mechanism for finding and being found.").

[22] Alternatively, public interest in the search engine market could be served through publicly funded search engines. *See Shaping the Web*, *supra* note 17.

## III. WHAT IS WRONG WITH THE SEARCH ENGINE MARKET?: ON BIAS AND IMPEDIMENTS TO COMPETITION.

### *A. Biases in Search Results*

Search engines are computer programs that search the Internet for relevant websites based on various search algorithms.[23] These locating tools allow users to locate resources posted on the Internet that match their query. While some search engines are general,[24] others focus on selective subjects.[25] Each search engine works differently. Some search engines collect listings independently, while others rely on information submitted by website administrators during the registration process.[26] Search engines further differ in the way data is organized. Some search engines use directories and list website addresses under headings and subheadings.[27] Other search engines create a database by sending out spiders (also called "crawler" or "bot"), computer programs that continuously search the web by following links on existing pages, to fetch as many documents as possible. The output is then indexed, creating a database that is used when an actual query is submitted.[28]

The main difference between search engines is in their classification and editorial analysis, which determine what would count, and rank, as the most relevant. These indexing and retrieval schemes are treated as proprietary information and are usually not disclosed.[29] Weight may be

---

[23] There are various ways to classify search engines. Search engines in the narrow sense employ crawlers (spiders) to locate, retrieve, index, and update information regarding new websites (for example, Infoseek, Lycos, Webcrawler, Northern Light). Other engines are directories, such as Yahoo!. In super search engines the spider is programmed to search for keywords within the text of the webpage instead of just in the page title, description, and metatags (Excite, AltaVista, HotBot). Metacrawlers enable searching multiple search engines simultaneously. They route the requests to a variety of different engines, compile and return the results to the user (Metacrawler, Dogpile). *See also* WEB DEVELOPER.COM, *supra* note 1.

[24] *See, e.g.*, Yahoo! (<http://www.yahoo.com>); AltaVista (<http://www.altavista.com>); Google (<http://www.google.com>). *See also* WEB DEVELOPER.COM, *supra* note 1, at 17-21 (Yahoo!), 31-37 (AltaVista).

[25] For instance, FindLaw (<http://www.findlaw.com>) for legal materials, or <http://www.searchnt.com> for Windows NT.

[26] *See* MICHAEL K. BERGMAN, THE DEEP WEB: SURFACING HIDDEN VALUE (2000), (visited Jan. 4, 2001) <http://128.121.227.57/download/deepwebwhitepaper.pdf> [hereinafter THE DEEP WEB].

[27] The most popular directory of that sort is Yahoo!.

[28] *See* (Webopedia)™ (visited Feb. 17, 2001) <http://www.webopedia.com/> (visited Feb. 17, 2001).

[29] Disclosing the search algorithm could help competitors gain a competitive edge. It could further allow manipulation by website owners who are seeking to promote their sites.

given to words in the title, the frequency in which a keyword appears on the page or is listed in the HTML Metatags.[30] The number of links is often used as an authority for quality, thus giving a higher rank to webpages that are connected by many links.[31] Other algorithms (such as Google's algorithms) analyze the links and consider where they come from (once again, giving more weight to links from sites that are more heavily linked).[32] Most spiders would skip a webpage that lacks any links from any other webpages.[33]

Indexing and retrieval algorithms could bias the search results. The high dependency on links gives prominence to pages on topics of mass interest, such as sports and sex rather than to less popular sites that sometimes might be more relevant to a particular query (such as sites on molecular biology). This type of ranking is further biased toward commercial websites that would often be pointed at from many sites due to business alliances. Commercial websites further enjoy an obvious advantage in promoting their own sites by advertising on television and radio as well as in newspapers.

Bias may also stem from a particular business model. Search engines could typically use three sources of income: fees collected from users, fees collected from advertisers or other third parties, and, increasingly, commission collected from websites. Search engines would either apply one of these business models or any combination thereof. While few search engines charge for their services either by selling copies or charging for access,[34] most search engines provide their services free of charge, thereby enhancing the number of users who visit the service and increasing the exposure for advertiser-supported links.[35] Revenues often come from

---

[30] Web page owners are seeking to manipulate search engines and gain a higher ranking on the search results, may use various tricks such as inserting keywords that are invisible to the viewers, or embedding keywords in the HTML (HyperText Markup Language) code that underlies a page (metatags). *See* F. Gregory Lastowka, *Search Engines, HTML, and Trademarks: What's the Meta for?*, 86 VA. L. REV. 835 (2000) (explaining at pages 843-46 what metatags are, and discussing throughout the whole paper the problems that arise from them in the field of trademark law).

[31] *See* Declan Butler, *Souped-Up Search Engines*, 405 NATURE 112 (2000).

[32] *See* Steve Lawrence & C. Lee Giles, *Accessibility of Information on the Web*, 400 NATURE 107 (1999) [hereinafter *Accessibility of Information*].

[33] *See* THE DEEP WEB, *supra* note 26, at 2.

[34] *See, e.g.*, LexiBot (<http://www.lexibot.com>) by BrightPlanet. Few search engines are owned by service providers who charge an access fee. Some search engine companies license their product for corporations who wish to provide web-like search capabilities on their intranet. *See* WEB DEVELOPER.COM, *supra* note 1, at 12.

[35] mySimon.com's www.mysimon.com offers a program which enables merchants to promote themselves through marketing done by mySimon.com (<http://www.mysimon.com/corporate/index.jhtml?pgid=merchantprogram&cid=bold>) (visited Feb. 17, 2001) and a Free Listing Service in which joining merchants can list their products in the search

selling banners to advertisers or selling information collected on users' behavior to third parties. Advertising space sold to advertisers is often associated with search queries, allowing advertisers to discerningly tailor the ads to the individual user.[36]

A third source of income is commission collected from site owners. Some spiders collect payments from sites for a larger, boldface look in the search results. Others collect fees from websites for submission of the site to be included in its listings.[37] Another model is a paid-inclusion system. Under such a payment scheme websites are charged for a more extensive indexing, thus increasing their chances of being included in the search results.[38] This would not necessarily guarantee the highest ranking at the search results page, but it would increase the likelihood that the website would be cited in response to a wider range of queries. Several search engines have developed a pay-for-placement model, under which websites pay to acquire a premier placement in the search results.[39] It is assumed that most users would only review the top results, and would seldom check bottom links. Consequently, even when a site is listed, it could be buried at the bottom of the list. Commission could be based on the amount of traffic the search engine sends to the site.[40] Under another scheme, common among bargain finders, an engine would charge sellers a certain percentage of any sales that resulted from its reference.

---

results free of charge (<http://www.mysimon.com/corporate/index.jhtml?pgid=merchantprogram&cid=freelist>) (visited Feb. 17, 2001). BiddersEdge.com offers users the opportunity to advertise their business. *See About Us: Advertising* (visited Feb. 17, 2001) <http://www.biddersedge.com/advertising.jsp>.

[36] This may raise serious privacy concerns. *See* A. Michael Froomkin, *The Death of Privacy?*, 52 STAN. L. REV. 1461, 1486-1489 (2000).

[37] *See, e.g.*, LookSmart submission program (last visited Nov. 2000) <http://submit.looksmart.com/> (charging commercial websites a submission fee for an expedited indexing process).

[38] For instance, LookSmart (<http://www.looksmart.com/>) offers websites a paid service of deep indexing. A site that is deeply reviewed would be included under a larger number of categories, resulting in its appearance in response to a greater range of queries than only if the homepage was listed. LookSmart collects a fee for each user it referred to the site. LookSmart further employs a pay-per-submission program charging a fee from commercial websites to be listed in its directory. *See* Danny Sullivan, *Paid Inclusion at Search Engines Gains Ground*, THE SEARCH ENGINE REPORT (Nov. 3, 2000) (visited Nov. 2000) <http://www.searchenginewatch.com/sereport/00/11-inclusion.html>.

[39] GoTo.com offers a program ("Pay for Performance™") for advertising auction sites: "An auction site selects the categories that are most relevant to their auction products and determines what to pay on a per-click basis. The higher a company bids, the higher in the search results the company's auction items will appear." *About GoTo Auctions* (visited Feb. 6, 2001) <http://www.auctions.goto.com/About/default.asp?LGSI=2860019621>. *See also Advertise With Us* (visited Feb. 6, 2001) <http://www.auctions.goto.com/Advertise/default.asp>; Danny Sullivan, *Pay For Placement?*, SEARCHENGINEWATCH.COM, (Nov. 8, 2000) (visited Nov. 2000) <http://searchenginewatch.internet.com/resources/paid-listings.html> (listing different search engines and reviewing their business models).

[40] *See, e.g.*, Ask Jeeves (visited Nov. 2000) <http://www.askjeeves.com/docs/about/policy.html>.

These business models are particularly worrisome. The strong link between editorial considerations and commercial interests raise serious concerns regarding the reliability of search results. Search engines would tend to structure search results to maximize their profits. To the extent that engines' revenues depend on websites that are included in its results, they would be inclined to assign a higher ranking to those sites that pay more.

The limits of search engines in locating resources on the net are increasingly acknowledged.[41] Biases, either those inherent to the algorithm or those related to business models, are raising concerns regarding the reliability of search results and the feasibility of locating information posted on the net. Search engines currently cover a small portion of the information published on the web.[42] Researchers believe that the largest search engines index no more then 16% of the content posted on the web.[43] Consequently, search engines provide only limited and selective access to information posted on the web.[44]

Furthermore, it increasingly becomes apparent that data mining on the web can no longer be thought of as collecting information on static webpages and creating a static directory structure. Locating information on the web requires an active search, which involves direct inquiries of internal searchable databases. Recent research distinguishes between the "Surface Web" and the "Deep Web."[45] Information in the Deep Web, researchers argue, resides in searchable databases and does not become apparent until compiled as a response to a direct query.[46] Such searchable databases include internal pages of large sites that are dynamically created, topic databases (such as patent records, SEC corporate filing, medical databases), auctions, classified, portals, or internal holdings of academic libraries.[47] Researchers concluded that information on the Deep Web is not

---

[41] *See Souped-Up Search Engines*, *supra* note 31 (claiming that even the most wide-reaching search engines cover barely half of the webpages on the World Wide Web).

[42] According to a research published by BrightPlanet, popular search engines are capable of searching only static and linked pages on the Web. Traditional search engines, it is argued, can only search the surface of the Web. They cannot retrieve content from the Deep Web, namely, websites that only produce results dynamically in response to direct requests. *See* THE DEEP WEB, *supra* note 26.

[43] For a discussion of the limitations of search engines see *Accessibility of Information*, *supra* note 32.

[44] Recent research suggests that information on the Web is 500 times greater than that retrievable by conventional search engines. *See* THE DEEP WEB, *supra* note 26, at 4.

[45] *See id.* at 3-4.

[46] The use of database technologies on the Internet (beginning at about 1996 through vendors such as Oracle), and a technology that facilitates dynamic serving of Web pages, introduced a whole new practice of database-driven designs.

[47] Covered by the research were deep websites such as Alexa, Amazon.com, Informedia (Carnegie Mellon University), MP3.com, IBM Patent Center, and eBay.com.

only greater than that on the Surface Web, but also tends to be of higher quality.

While conventional search engines collect information and index static pages, a query submitted to multiple search sites, directly and simultaneously, could retrieve content from the Deep Web.[48] Researchers conclude:

> Searching has got to evolve to encompass the complete Web. Directed query technology is the only means to integrate deep and surface Web information. But because deep information is only discoverable via directed query and can not be comprehensively indexed, it is unclear today whether a "global" solution can be found. . . . The information retrieval answer has to involve both "mega" searching of appropriate deep Web sites and "meta" searching of surface Web search engines to overcome their coverage problem.[49]

Searching the net in a reliable and comprehensive way is crucial for its functioning both as an electronic marketplace, as well as a democratic forum. An open environment that enables automated querying of publicly opened searchable sites is a precondition for covering the net by searching tools. As further suggested by the next section, concerns regarding the reliability of search results could also be addressed by an open and competitive search engine market.

### *B. Can the Market Provide a Solution?*

It is arguable that competition among search engines could fix some of the distortions described above. After all, in a competitive market search engines will be forced to provide useful and reliable results in order to keep their market share. If users could choose, they would arguably abandon any search engine that provides partial or inaccurate results and would prefer those search engines that provide the most comprehensive, relevant results. Thus, competition would presumably refine existing search methods.

A possible flaw in this market solution is information deficiencies. Services provided by search engines are not transparent and cannot be easily compared by users. How can a user determine whether the search results are relevant? Users may appreciate the ease and friendliness of use, the speed, the look and feel of the results page, or the fees charged for the

---

[48] *See* THE DEEP WEB, *supra* note 26, at 4.

[49] *Id.* at 28.

use (if any). The search results themselves, however, are not transparent. Introna and Nissenbaum argue that competition cannot develop in the search engine marketplace since the prerequisites of an efficient market are simply not met. Thus, they argue, users' choice to use a certain search engine cannot be said to reflect users' preferences since users lack critical information about the alternatives.[50]

Relevancy and accuracy of search results cannot be easily determined, unless users are aware of an alternative. Unless a user is looking for a particular document she knows was posted on the web, she would never know whether some sites exist at all. Users may appreciate only the results they see, and would not be aware of other results not retrieved by the search engine. Furthermore, while some results could be more easily comprehended (such as a list of sites offering products at different prices), the usefulness and trustworthiness of other types of results would not be as easily appreciated.

A well-functioning market could, however, address some information deficiencies. Competitors are likely to discover information and inform users of the limits suffered by search engines. Newcomers who wish to enter the search engine market would have sufficient incentives to reveal inadequacies in existing technologies and to demonstrate the shortcomings of established search engines in order to emphasize the advantages provided by their own technology.[51] In other words, competition itself can increase transparency and stimulate users' demand for better search engines.

### C. *Technologically Enabled Competition: On Metacrawlers and Why They Are Important*

One of the impediments for developing a competitive market in search engines is the time it takes to compare search results. The average user is unlikely to invest in comparing search results of different search engines since that would be a time-consuming task. Here technology may drive competition. If there were different searches that any given user could conduct simultaneously, in a cost-effective way, chances are that users

---

[50] *See Shaping the Web*, *supra* note 17, at 177 ("Given the vastness of the Web, the close guarding of algorithms, and the abstruseness of the technology to most users, it should come as no surprise that seekers are unfamiliar, even unaware, of the systematic mechanisms that drive search engines.").

[51] *See, e.g.*, THE DEEP WEB, *supra* note 26 (revealing flaws of existing search engines to demonstrate how new technology developed by the research sponsors, BrightPlanet, may address them). *See also* LexiBot (visited Jan. 27, 2001) <http://www.lexibot.com/>.

would be able to compare search results provided by different engines. Search engines are themselves searchable databases, and conducting a search by sending a query simultaneously to several search engines could raise users' awareness of the limits and biases of search engines.[52]

Metacrawlers allow a more efficient search by passing the user's query to several search engines simultaneously. Unlike search engines, metacrawlers do not crawl the web themselves to build listings. These computer programs are designed to query and aggregate results from various search engines and online databases (such as Excite, Infoseek, and Yahoo!). Such search results provide the raw material for the metacrawler's editorial processes. Metasearchers compile, modify, and edit search results in various degrees, depending on the crawler's design and user's preferences. The processed results are then blended together onto one page. Some metacrawlers offer value-added services, such as automated searches from the text itself, or contextual searches.[53]

Metacrawlers may facilitate competition in the search engine market. This type of application allows automated and efficient comparison of search results and assists users in proficiently comparing the different choices available to them.[54] They further allow easy switching from one engine to another, thus lowering the entry barriers for new engines.

Consequently, metasearch engines may threaten the commercial interests of spiders.[55] While search engines do not compete with the content provider (and may actually increase traffic to the indexed website),

---

[52] Effective searching requires a mixture of techniques. If you want to trawl for background information before beginning a research project, you might use a search engine to identify key sites, in combination with a 'metasearch' engine which allow users to query multiple search engines simultaneously. *See Accessibility of Information*, *supra* note 32.

[53] *See, e.g.*, Atomica.com, (visited Feb. 10, 2001) <http://www.atomica.com/>; Zapper.com (visited Feb. 10, 2001) <http://www.zapper.com/>. Other metasearch engines analyze the results to provide a different ranking, check for broken links, and weed out duplicate entries. *See Accessibility of Information*, *supra* note 32.

[54] One danger is that biases suffered by search engines would be suffered by Metacrawlers as well, in that they compile search results of various search engines. Introna and Nissenbaum argue that "We are unlikely to find much relief from these robust irregularities from meta search engines like Metacrawler, Ask Jeeves, and Debriefing, 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." *Shaping the Web*, *supra* note 17, at 176. This, they argue, is due to a fair degree of convergence in the results yielded by various search engine algorithms and decision criteria. *See id.* For various reasons discussed above, it is not justified to assume that the search engine market would be static and that new technologies and new competitors would not seek to enter this market. The high level of control potentially exercised over access to information is likely to push for a more aggressive competition.

[55] Note that search engines will not always object to the use of Metasearchers. Indeed, some search engines pay to acquire a placement in the metasearch results (pay-for-placement). This is particularly important for small stand-alone search engines that wish to enter the market and acquire a wider exposure to users.

they could adversely affect the commercial interest of the search engine. Search engines striving for revenues and market share, are likely to resist competition of this kind. Metacrawlers will be particularly objectionable to those search engines that derive their income from selling advertisements. The reason is that metacrawlers enable users to "deep link" directly to the requested pages, and thereby bypass the homepage search results page of the originating engine. Consequently, users will skip the advertisements posted on the search engine's webpage, reducing its attractiveness to advertisers. For each individual search engine, competition is risky. Not only does it threaten to reduce its market share, it could also require some search engines to change their business models and search algorithms.

Search engines may therefore seek to avoid competition as much as possible. Indeed, the search engine market is consolidating and is already going through a significant wave of mergers and acquisitions.[56] A good number of search engines are either owned by, or affiliated with, Metacrawlers.

### *D. Legal Impediments on Competition*

It is apparent that the search engine market suffers some distortions. How should the law address such distortions? One approach is regulation, which seeks to address information deficiencies, requiring search engines to disclose their search methods or business models. Consumer protection policies typically assume that consumers are suffering from systematic information disparities vis-à-vis providers, which lead to market inefficiencies. Such a *consumer protection approach* could be inadequate in an environment that already suffers from information overflow. Even information that is available cannot be practically processed by consumers and is rarely internalized in their decision making process.[57]

Regulation is insufficient, however, for preserving the decentralized nature of the net. Since search engines are functioning as virtual gatekeepers, a competitive market is essential for reserving the

---

[56] Go2Net owns the metasearch sites Dogpile and MetaCrawler. Intelliseek, which owns BullsEye meta search software, acquired the metasearch engine ProFusion in April 2000. Cnet acquired SavvySearch in October 1999. *See Movement in Meta Search*, THE SEARCH ENGINE REPORT (May 3, 2000) <http://searchenginewatch.internet.com/sereport/00/05-metasearch.html>.

[57] *See* Niva Elkin-Koren and Eli M. Salzberger, *Law and Economics in Cyberspace*, 19 INT'L REV. L. & ECON. 553 (1999). Such inadequacy is already evident in the management of privacy policies on the Internet. Users are bombarded with privacy statements, many of which are extremely long, technical, and incomprehensible. Simply requiring online providers to disclose their information gathering practices could not sufficiently guarantee users privacy.

decentralized nature of the information landscape in cyberspace, and for supporting competitiveness in electronic commerce. This suggests that it is highly important to keep the search engine market as competitive as possible.

Open competition in the search engine market could remedy some of the distortions discussed above. Competition may increase accuracy of search results and maximize coverage of posted information. While both market mechanisms and a technological race could drive competition, legal rights of exclusion may hinder it. If search engines can legally exclude metacrawlers from using their output, then metacrawlers can no longer function to facilitate competition. The same is true for a right to prevent indexing by a search engine. Website owners may refuse a license to index altogether. Indeed, they could have an interest in reaching as many new users as possible. Yet, they might wish to control the context in which their site is presented. A legal policy that seeks to facilitate competition in the search engine market should avoid any claims to legally prevent indexing and listings. Such a policy should facilitate free access to search results so that indexing and retrieval processes remain as open as possible to competition.

Expansive limits on the use of search engines may reduce competition and result in an exclusionary information environment. Such limits were recently introduced in the *eBay* case, in the form of exclusion rights of sites against the mining of data and indexing. Under the eBay rule every site can legally stop indexing by any search engine. This rule further allows search engines to legally prevent the use of their search results by metacrawlers, data aggregators, and directories.

The next section takes a look at the legal rules that affect the search engine market.

## IV. PROTECTING THE DOMINANCE OF SEARCH ENGINES: LEGAL RIGHTS IN SEARCH RESULTS

The law may allow search engines to legally prevent the use of their search results, thus requiring anyone who wishes to use such results to acquire a license. Search engines seeking to prevent others from using their search results may claim passing off, unfair competition, copyright infringement, and recently under the eBay rule, trespass to chattel. This section reviews the legal rights afforded to search engines over the use of their engines by metacrawlers. I then turn to discuss the *eBay* decision and its ramifications to competition in the search engine market.

### *A. The Pre-eBay Legal Regime*

The pre-*eBay* regime offered a relatively limited protection to search results. Legal rights afforded by copyright law to search results are narrow. The results returned by a search engine in response to a user's query are facts indicating the URLs of webpages that match the query defined by the user. Consequently, when any search engine uses the search results produced by another, it is making use of facts. Facts, as such, are not protected by copyright law and are purposely left in the public domain.[58] As long as a search engine extracts only the information, and does not copy its organization or presentation, which might be a protected *expression*, there is no copyright liability.

Search results could themselves constitute a copyrightable subject matter. Thus, compilations are protected for their originality in the selection and organization of data. Such selection could be copyrighted if it reflects the author's choice of the most valuable or relevant items.[59] The interactive nature of a search suggests, however, that the selection of items is often a by-product of the query defined by the user, so that the end-user could be considered a co-author of the compilation. Therefore, to the extent such compilations reflect authorship, it is arguable that such authorship is attributable not only to the search designer but also to the user. Moreover, metasearch engines often do not copy the entire list of search results (or a significant part thereof) in the same order or organization.[60] In fact, most metacrawlers edit the results and offer a new selection and ranking. If the original selection or organization are not copied, then there could be no copyright infringement of the compilation.[61]

Search results may further contain titles, abstracts, or thumbnail sketches of the original site that could be protected as a work of authorship. Those, however, are rarely owned by the search engine itself, and are often created and owned by the author of the original webpage.[62] Copying the abstract without authorization could infringe the copyright owner's

---

[58] *See* Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991).

[59] It is questionable whether authorship could be manifested by a computer program which embodies the selection and ranking criteria, and to what extent such authorship could be attributed to the designer of the computer program.

[60] Unauthorized posting of an entire page-results, including advertisements, may constitute copyright infringement. Paradoxically, this practice could advance the commercial interests of the original search engine since it does not threaten the commercial value of its advertising banners.

[61] *See* Kregos v. Associated Press, 3 F.3d 656 (2d Cir. 1993).

[62] A search engine that independently creates an abstract of an original work posted on a website might infringe the exclusive right of the website copyright owner to create a derivative work. *See* 17 U.S.C. § 106(2).

exclusive right to reproduce. A recent case suggests, however, that this could be considered fair use. Fair use analysis would balance the nature of the use and its impact on the original author.[63]

In *Kelly v. Arriba Soft Corp.*,[64] the court addressed the use of a visual search engine operated by the defendant ditto.com. The search engine allowed users to search for images posted on the web, and posted a list of reduced thumbnail pictures in response to a search query entered by the user. The search engine maintained an index database of thumbnail images.[65] When a user clicks on the thumbnail, two windows open simultaneously:[66] one contains the full size image (by opening a link to its originating webpage) and the other contains the original webpage in full. Photographs owned by the plaintiff, which were posted on the net, were indexed by the ditto.com search engine and were made available in a thumbnail form.

The court held that the reproduction and display of copyrighted images in thumbnail form constituted copyright infringement unless authorized by the copyright owner. Yet, in view of the established importance of search engines and the transformative nature of using reduced versions of the images, they exempted such use of the copyrighted materials, holding it was fair use. The court analyzed four factors to determine whether the defendant's use was "fair use": (1) the purpose and character of the use, (2) the nature of the copyrighted work, (3) the amount and substantiality of the portion used, and (4) the effect of the use on the potential market or value. The court found the use to be fair despite the creative nature of the copied photographs and the substantiality of copying, which weighed against fair use. It emphasized the transformative nature of the use of the image, noting that while the photographs are artistic works, which have an

---

[63] *See* RAYMOND T. NIMMER, INFORMATION LAW, ¶ 4.09, at 4-34 (1996 & Supp. 2000).

[64] 77 F. Supp. 2d 1116 (C.D. Cal. 1999).

[65] The images are obtained by a computer program that travels the web in search of images (crawler), then they are screened and ranked by the staff, and converted into thumbnail images. Full size images are temporarily stored on the search engine's server until the thumbnail is made.

[66] An older version of the search engine (Arriba Vista Image Searcher) allowed users to display the full size version of the image by clicking on the desired thumbnail, by opening a link to its originating webpage. Only the image itself was displayed and not any other part of the originating page. The search engine also displayed the image attributes (the image's dimension, and the URL where it originated). On this application the court noted:

> It allowed users to view (and potentially download) full-size images without necessarily viewing the rest of the originating Web page. At the same time, it was less clearly connected to the search engine's purpose of finding and organizing Internet content for users. The presence of the image attributes page in the old version of the search engine somewhat detracts from the transformative effect of the search engine.

*Kelly*, 77 F. Supp. 2d at 1119.

aesthetic character, the visual search engine has a functional purpose and is designed to catalog and improve access. Thus, the court found no evidence of market harm to the plaintiff.

Search engines striving to stop unauthorized use of their search results may also seek protection under contract law. Many terms of use or license agreements[67] employed by search engines include various restrictions concerning the use of the search engines' results. For instance: the license is limited to personal use; the user agrees not to forward the results in whole or in part, and not to mirror the results; the user agrees not to modify the results (reformat and display them or mirror them on their website); the user may not send automated queries to the search engine.[68] Recently, end user license agreements explicitly prohibit the use of robots.[69] *Users* are often broadly defined by such form contracts as anyone who entered, accessed, or used the search engine. Such broad language intends to cover not only individual users, but also computer programs, which conduct automated searches. These contractual restrictions seem like an attempt to protect existing business models, more than an attempt to protect the system's functionality.

Many of the functions performed by metacrawlers could be in breach of such license provisions, such as sending automated queries or forwarding the search results. If such restrictions are enforceable against metacrawlers, its operation may constitute a breach of the license agreement. Such a breach may result not only in contract liability but also in liability for copyright infringement for using a copyrighted work (the computer program) without a license.[70]

Search engines raising a contractual claim against metacrawlers must first establish that a contract was formed. Most search engines do not require a positive acceptance (i.e. clicking "I accept") of the terms and conditions of the license agreement. Assent is rather fictitious and is

---

[67] Search engines are computer programs that search publicly opened or restricted databases. Computer programs are copyrighted to their owners and their use is subject to a license agreement.

[68] *See, e.g., Google Terms of Service* for Google.com (visited Jan. 23, 2001) <http://www.google.com/terms_of_service.html>.

[69] *See, e.g.*, Scour End User License Agreement (visited Jan. 22, 2001) <http://www.scour.com/Software/Scour_Exchange/EULA.phtml> ("You may not use 'robots' or other computer programs or devices which automatically enter search queries or retrieves links from SX. The use of such programs places an undue load on the SX system which, at a minimum, slows down the service for live users waiting for results and may even jeopardize the integrity and operability of the entire system. Accordingly, SX users who use robots and the like may have their SX registration permanently cancelled.").

[70] *See* Telerate Systems, Inc. v. Caro, 689 F. Supp. 221 (S.D.N.Y. 1988) (holding that access by unauthorized means, e.g., a different terminal type, was not only a breach of contract but also violation of copyright).

attributed to the mere use of the search engine. The homepage would often include a link to the Terms of Use or License Agreement. A statement at the top of the license agreement would commonly state that: "the use of this search engine constitutes your acceptance of the following terms and conditions." The enforceability of automated contracts was challenged in court. Recent court decisions[71] and state legislation adopting the Uniform Computer Information Transaction Act (UCITA) facilitate the enforcement of such contracts. By doing so, owners of search engines are given a monopoly over data, and, consequently, users' choices could be significantly narrowed.[72]

It seems, however, that the applicability of contract law to terms of use that are designed to cover devices that automatically surf the web, raise somewhat different considerations. The enforceability of automated contracts was recently tested in court in *Ticketmaster Corp. v. Tickets.com, Inc.*[73] Plaintiff Ticketmaster.com operates a website that includes a directory of entertainment events and sells tickets to such events. Tickets.com provides a similar service. Where the exclusive tickets broker is Ticketmaster, Tickets.com creates a link to the interior Ticketmaster events page where the user can purchase tickets for the particular event she seeks. Ticketmaster filed a suit against Tickets, claiming among other things that Tickets' operation amounts to a breach of contract. A statement posted on the homepage provided that anyone going beyond the homepage agrees to the terms and conditions set forth. The terms of use provided that the information is for personal use only and may not be used for commercial purposes.[74] Emphasizing the lack of knowledge of the terms, and the absence of showing any implied agreement to them, the court distinguished between merely posting terms on a homepage, and shrinkwrap licenses or "click-on agreements" that were held by courts to

---

[71] *See* ProCD, Inc. v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996); Hill v. Gateway 2000, Inc., 105 F.3d 1147 (7th Cir. 1997).

[72] For a critical analysis of these provisions see J.H. Reichman & Jonathan A. Franklin, *Privately Legislated Intellectual Property Rights: Reconciling Freedom of Contract with Public Good Uses of Information*, 147 U. PA. L. REV. 875 (1999). *See also* Niva Elkin-Koren, *Copyright Policy and the Limits of Freedom of Contract*, 12 BERKELEY TECH. L.J. 93 (1997).

[73] No. CV 99-7654 HLH, 2000 U.S. Dist. LEXIS 4553 (C.D. Cal. 2000).

[74] *See Permitted Use* (visited Feb. 11, 2001) <http://www.ticketmaster.com/h/terms.html> ("You agree that you are only authorized to visit, view and to retain a copy of pages of this Site for your own personal use, and that you shall not duplicate, download, publish, modify or otherwise distribute the material on this Site for any purpose other than to review event and promotional information, for personal use, or to purchase tickets or merchandise for your personal use, unless otherwise specifically authorized by Ticketmaster to do so. You also agree not to deep-link to the site for any purpose, unless specifically authorized by Ticketmaster to do so.").

be enforceable.[75] "It cannot be said," the court held, "that merely putting the terms and conditions in this fashion necessarily creates a contract with anyone using the web site."[76] Other courts, however, found such posted terms to constitute an enforceable contract. In *Register.com, Inc. v. Verio, Inc.*,[77] the court addressed a breach of contract claim by an authorized domain name registrar (Register.com) for a breach of contract governing the use of data contained in its WHOIS database. The terms of use were published on the home page of the plaintiff's website, stating that by submitting a query the user agrees to abide by these terms. The court rejected the defendant's claim that it had not assented to terms of use, holding that by submitting a WHOIS query, the defendant manifested its assent to be bound by the contract.[78]

Furthermore, even if such contracts are formed, it is arguable that some restrictions on the use of search results may not be enforceable.[79] A license that restricts the use of (otherwise unprotected) information could be preempted under copyright law.[80]

The applicability of other legal claims, such as passing off, unfair competition, misappropriation, or unjust enrichment to the use of search results may be limited. Liability for passing off[81] may arise if the meta-search engine does not attribute results to the original engines, and misleads users into believing that the search services were provided by the

---

[75] *See* Caspi v. Microsoft Network, L.L.C., 732 A.2d 528 (N.J. Super. Ct. App. Div. 1999) (holding that a choice of forum clause in a click-on agreement was enforceable).

[76] *Ticketmaster*, 2000 U.S. Dist. LEXIS 4553, at *8.

[77] 126 F. Supp. 2d 238 (S.D.N.Y. 2000).

[78] *Id.* at 248 ("Verio does not argue that it was unaware of these terms, only that it was not asked to click on an icon indicating that it accepted the terms. However, in light of this sentence at the end of Register.com's terms of use, there can be no question that by proceeding to submit a WHOIS query, Verio manifested its assent to be bound by Register.com's terms of use, and a contract was formed and subsequently breached.").

[79] For instance, one could claim that such restrictions should be unenforceable in that their enforcement would be contrary to public policy for reasons discussed below. For various reasons contract doctrines are likely to offer only limited help in policing restrictions on the free flow of information. *See* Niva Elkin-Koren, *A Public-Regarding Approach to Contracting Over Copyrights*, *in* EXPANDING THE BOUNDS OF INTELLECTUAL PROPERTY: INNOVATION POLICY FOR THE KNOWLEDGE SOCIETY, 191. (Rochelle Dreyfuss, Dianne L. Zimmerman, & Harry First eds., Oxford University Press 2001).

[80] *But see* ProCD, Inc. v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996) (holding that contracts, as opposed to state laws, could never be preempted under copyright law). This claim was raised in the *eBay* case, but was never addressed by the court since the preliminary injunction was granted on the basis of trespass.

[81] Passing off is the false designation of origin, or misleading representation of other related facts. The general purpose of this rule is to prevent one person from passing off his goods or his business as the goods or business of another. *See* American Steel Foundries v. Robertson, 269 U.S. 372, 380 (1926).

metacrawler rather than by the original engine. Metacrawlers, by their very nature, however, would usually explicitly declare that they are using other search engines. In fact, that is the added value service that they offer to provide.

The misappropriation claim was developed by the Supreme Court in 1918 in the landmark case of *International News Service v. Associated Press*,[82] which provided a legal remedy against a news agency that copied and used time-sensitive news reports gathered by the plaintiff. The Court held that the compiler is entitled to prevent a competitor from the wholesale appropriation of "hot news" in direct competition with the compiler.[83] Recent case law suggests that applying the misappropriation claim to instances of simply copying data would be preempted under copyright law.[84] Preemption provisions in the 1976 Copyright Act preempt any right that is *equivalent* to any of the exclusive rights of a copyright owner in works of authorship fixed in a tangible medium, that come *within the subject matter* of copyright.[85] Even though copyright law does not afford protection to data as such, a state claim that is "equivalent" to a copyright claim in data would nevertheless be preempted.[86] A misappropriation claim merely on the basis of copying search results is therefore likely to be preempted.

To avoid preemption the *International News Service*-like misappropriation claim should meet the following requirements:

> (i) a plaintiff generates or gathers information at a cost; (ii) the information is time-sensitive; (iii) a defendant's use of the information constitutes free riding on the plaintiff's efforts; (iv) the defendant is in direct competition with a product or service offered by the plaintiffs; and (v) the ability of other parties to free-ride on the efforts of the plaintiff or others would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened.[87]

---

[82] International News Serv. v. Associated Press, 248 U.S. 215 (1918).

[83] *Id.* at 240 (calling the process "unauthorized interference with the normal operation of complainant's legitimate business precisely at the point where the profit is to be reaped, in order to divert a material portion of the profit from those who have earned it to those who have not; . . .").

[84] *See* N.B.A. v. Motorola, Inc., 105 F.3d 841 (2d Cir. 1997) (narrowly defining the International News Service misappropriation cause of action, holding that a broad misappropriation doctrine is preempted as virtually synonymous with wrongful copying and indistinguishable from copyright infringement).

[85] *See* 17 U.S.C. § 301(a) (1994).

[86] *See, e.g., N.B.A.*, 105 F.3d 841 (holding that a state law misappropriation claim regarding the use of basketball game results is preempted under § 301); ProCD, Inc. v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996) (holding that preemption doctrine applies to all works covered by § 102 and § 103, but finding no preemption of a contract claim).

[87] *N.B.A.*, 105 F.3d at 845.

Even if search results could resemble news, they would not easily be perceived as time-sensitive, and their value would not usually be high for a brief period of time.[88] For "hot news" to be valuable it must be current and report the latest on an ever-changing list of items. Even though users are likely to prefer a prompt response to their search queries, the value of search results would lay in their comprehensiveness, breadth, relevancy, efficiency, or reliability, and would normally last for longer than a "brief period of time".

Furthermore, metacrawlers do not necessarily interfere with the normal operation of search engines. Metacrawlers do not substitute search engines, but instead provide an added value to search results originated by various engines. Indeed, a listing with a metacrawler may enhance exposure to the search engine's search results. The impact of metacrawlers on search engines businesses would depend on the search engine's particular business model. For instance, search engines that collect fees from listed sites, either for submission or for extensive indexing, would not suffer any loss from the use of their results by metacrawlers. Search engines which receive commissions based on the number of hits originated by the engine, may indeed benefit from the larger exposure of its search results when those are listed by a metacrawler. On the other hand, search engines that collect fees for banner advertisements on the search-results page might suffer from a reduction in advertising fees due to an increase in automated searches on their site. Such search engines, however, would not be driven out of business, although the presence of metacrawlers might pressure them to adopt an alternative business model, suitable for the changing technological environment.[89]

Since the operation of metacrawlers doesn't replace search engines and thus, does not prevent them from running their "business," the claim of misappropriation against metacrawlers should not succeed.

---

[88] For a different view see Michael J. Schmelzer, *Protecting the Sweat of the Spider's Brow: Current Vulnerabilities of Internet Search Engines*, 3 B.U. J. SCI. & TECH. L. 12 (1997) (arguing that search results resemble news, and thus the taking of search results could constitute a misappropriation claim a la *International News Service*, 248 U.S. 215 (1918)).

[89] In *Fred Wehrenberg Circuit of Theatres, Inc. v. Moviefone, Inc.*, the court held: "For a claim of misappropriation of 'hot news' to succeed, defendant's actions must make plaintiff virtually cease to participate in the business in question." Fred Wehrenberg Circuit of Theatres, Inc. v. Moviefone, Inc., 73 F. Supp. 2d 1044, 1050 (E.D. Mo. 1999).

### B. *The eBay Decision and Its Significance*

Bidder's Edge, an auction aggregation site, allows its users to search for items across numerous online auctions simultaneously by conducting a single search, without having to search each auction site individually.

eBay, the largest online person-to-person auction site, objected to the unauthorized use of its database by metacrawlers and data aggregators. eBay's user agreement prohibited the use of "any robot, spider, other automatic device, or manual process to monitor or copy our web pages or the content contained herein without our prior expressed written permission."[90] Thus, any automated search becomes an alleged breach of a license agreement. eBay further employed technical means to block access to Bidder's Edge, using a robot exclusion header. Once Bidder's Edge began to make requests through proxy servers, it turned out to be more difficult to trace back to the originating IP address and block queries from those addresses that were suspected of using a robot. Eventually eBay applied for a preliminary injunction against Bidder's Edge and won an injunction enjoining Bidder's Edge from accessing eBay's systems by use of any automated querying program without eBay's written authorization.

Rejecting Bidder's Edge's claim that it cannot trespass eBay's website because it is publicly accessible, the court held that eBay's servers and their capacity are private property and therefore access is conditioned upon the owner's permission. The court held that Bidder's Edge searches constituted an unauthorized use of this property, depriving eBay of the ability to use the occupied portions of its personal property for its own purpose.[91] The court concluded that automated access was explicitly prohibited by eBay either in its terms of use, or by its explicit notifications to Bidder's Edge and various attempts to technically block automated searches originated by Bidder's Edge. Even though Bidder's Edge's actual use of eBay's site did not overburden the system, the court found that such unauthorized use, if done by many spiders, could cause eBay irreparable

---

[90] eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1060 (N.D. Cal. 2000) (citation omitted). eBay used a Robot Exclusion Standard, namely a message detectable by robots which are designed to read a particular data file, robots.txt and comply with the control directives it contains. This robot was described by the eBay court as "a computer program which operates across the Internet to perform searching, copying and retrieving functions on the web sites of others." *Id.* (citation omitted). The court described a robot as capable of executing thousands of instructions per minute, in excess of what a human can accomplish, and therefore also consumes the processing and storage resources of the system.

[91] The court held that "[t]he law recognizes no such right to use another's personal property." *Id.* at 1071.

harm.[92] It concluded that "if [Bidder's Edge's] activity is allowed to continue unchecked, it would encourage other auction aggregators to engage in similar recursive searching of eBay systems such that eBay would suffer irreparable harm from reduced system performance, system unavailability, or data losses."[93]

As several commentators have argued, the extraction of data from a site could not be thought of as accessing the physical medium on which that data is encoded.[94] Such a rule would be too broad for regulating the use of digitally encoded information since digital information is not comprehensible, merely reading such information would necessarily involve the "use" of the medium in that broad sense.[95]

Furthermore, the court emphasized eBay's objection to the use of information resided on its servers. The objection of a website owner to the use of information could not, in itself, make such use "unauthorized." The fact that eBay was technically trying to prohibit the use of its database by Bidder's Edge[96] is insufficient for constituting a legal claim. To turn the use of data into an "unauthorized" use, one must first establish that the site owner has an exclusive right to authorize it. Information could be excluded in Intranet systems behind firewalls. If an owner of a database seeks to keep it proprietary and make it available exclusively for the internal use of her organization, access to such a database would be

---

[92] *Id.* at 1071-72 ("If the court were to hold otherwise, it would likely encourage other auction aggregators to crawl the eBay site, potentially to the point of denying effective access to eBay's customers. If preliminary injunctive relief were denied, and other aggregators began to crawl the eBay site, there appears to be little doubt that the load on eBay's computer system would qualify as a substantial impairment of condition or value.").

[93] *Id.* at 1066. Maureen A. O'Rourke criticizes the awarding of a remedy based on such speculative harm since she predicts that no proliferation of spiders is anticipated. *See* Maureen A. O'Rourke, *Shaping Competition on The Internet: Who Owns Product and Pricing Information?*, 53 VAND. L. REV. 1965 (2000) [hereinafter *Shaping Competition*]. Assuming that the business model of indexing sites is selling advertising, she predicts: "Chances are high that the number of indexing sites that could attract enough money to remain in business is less than the number that would materially adversely affect system performance." *Id.* at 1981.

[94] *See, e.g.*, Dan L. Burk, *The Trouble with Trespass*, 4 J. SMALL & EMERGING BUS. L. *27 (2000) [hereinafter: *The Trouble with Trespass*]; *Shaping Competition, supra* note 93.

[95] *See* Jessica Litman, *The Exclusive Right to Read*, 13 CARDOZO ARTS & ENT. L.J. 29 (1994).

[96] eBay employed technical means which relied on a convention (often respected by search engines), that signaled to search engines that access to the servers is blocked (robot.txt). Bidder's Edge did not comply with that convention. Does it make the use unauthorized? Consider the context of privacy. If a user disables cookies or any other technical means by which a website collects her personal data, would such an act be unauthorized? Would it become unauthorized simply because the website owner stated in the terms of use that the use of the site is authorized only subject to the exercise of intrusive technical means? If a site owner has no legal right to collect personal data without the user's consent, the mere use of data collection means would not make their disablement unauthorized, and certainly not illegal. The individual's right to privacy would outweigh data collection interests.

technically restricted. Other databases may be restricted to subscribers only and protected by passwords or encryption. The novelty of the eBay rule is in granting site owners the legal right to determine the terms of access to publicly available information.

The eBay rule allows the owners of search engines and searchable sites to stop any undesirable (potentially competitive) use of data. A search engine could seek an injunction against a metacrawler or any automated use of the site. This means that any such use would require a license. Thus, even though the eBay rule does not accord any property rights in search results or data retrieved from a website, it does create a de facto right to exclude the use of search results.

The eBay rule established a broad right to exclude. Prior to the *eBay* decision, Dan Burk predicted that trespass in chattel as broadly interpreted in the context of cyberspace, would become the "all-purpose cause of action for the Internet," covering almost every application from software agent to indexing spiders.[97]

In the context of the search engine market the eBay rule expansively prohibits any unauthorized data mining. By upholding eBay's claim that automated search on its site was unauthorized, the court in fact endorsed a novel right of site owners to govern the indexing of their site and of any information posted on it. Indeed a website owner could technically affect whether they will be located by a search engine, and how they would be listed.[98] Prior to the eBay rule, however, site owners were not legally entitled to determine how and to what extent their site would be indexed. Such a right would allow every site owner to control the way its site is referred to and indexed. Most sites, however, are unlikely to exercise such a right since they are highly dependent on search engines for reference. The eBay rule would certainly increase biases toward large commercial sites that could use it to control the indexing process. The harsh consequences of the eBay rule will directly affect the search engine market and the way it will develop in the near future.

---

[97] *The Trouble with Trespass*, *supra* note 94, at *46. As phrased by Burk while analyzing the trespass claim "the cause of action might better be named 'using a networked computer.'" *Id.* at *47.

[98] A site owner can technically signal that automated indexing is not welcome by employing a robot exclusion header in its robots.txt file. *See* <http://www.searchenginewatch.com/> (visited Nov. 2000). Furthermore, site may also manipulate the indexing by using metatags. Metatags are used to determine the keywords that would match with their site, and sometimes to manipulate the indexing algorithm so that the webpage will be listed in a certain category.

### C. How Does the Law Shape Competition Among Search Engines?

What are the ramifications of the eBay rule? A right to exclude information mining and indexing is likely to hinder, rather than encourage, competition in the search engine market. Such a legal right may affect both vertical relationships between sites and search engines in which they are listed, and competition among search engines, metacrawlers, and data aggregators.

At the site-search engine level, when search engines must acquire a license to locate and refer to post information, the cost of the search is higher. That is due to higher transaction costs involved in negotiating and acquiring the necessary licenses and paying the license fees. The commercialization of the reference process would increase barriers on entry and allow a considerable advantage to commercial engines.

The eBay rule further creates dependency of the indexer in the subject of indexing. Sites could condition licensing in exchange for receiving a higher ranking, specific presentation on the search results, or the exclusion of others on search results. Such dependency of indexers in their subjects, created by law, would hardly contribute to the reliability of the search engine market. Once again, this is likely to give an advantage to commercial sites.

Search engines themselves could rely on the eBay rule to object to the use of their output search results. They could use licensing search schemes to acquire market domination, force business alliances when they are not technically necessary, and limit the operation of competitors. Those business interests, however, are not necessarily compatible with the public interest. It could undermine the feasibility of introducing new, independent metacrawlers in the search engine market.

Finally, a legal right to exclude, as opposed to the technical ability to do so, is propertizing the search results. It is turning search results into a corporate asset. An exclusive right to use search results, such as the eBay rule, protects a market share. Search results as a corporate asset could be accumulated. It could be used strategically to prevent indexing by competitors, or restrict indexing to advance commercial interests. A right to exclude indexing could therefore serve to hinder competition among search engines and would facilitate centralization of the search engines marketplace.

How is this likely to shape the information environment? Several commentators have focused on the problem of segmentation.[99] A broad

[99] O'Rourke predicts that "the Internet is likely to evolve into a 'place' characterized by both open and closed areas." Maureen A. O'Rourke, *Fencing Cyberspace: Drawing Borders in a Virtual World*,

legal right to exclude publicly available information would compartmentalize the networks into guarded zones.[100] Such processes of zoning, segregating, and dividing cyberspace could detrimentally affect free flow of information and undermine the public benefits associated with a network.[101]

The eBay rule entails, however, another outcome. The newly introduced legal right to exclude facilitates control. It is likely to support centralization of the information environment by enabling control over access to information by legal means.

The eBay rule introduces a broad right to control access. Information is abstract and normally detached from any physical presence. Information in digital form, however, requires some sort of electronic access to make it intelligible and instrumental. Consequently, paradoxically, the virtual space accompanied by the newly created rights to exclude access reconstruct the physicality of information.

Apparently the preliminary injunction issued in the *eBay* case allowed eBay to deny only a particular type of access. It is arguable that access is technically available to all information that is posted on the web, and that it remains so even under the eBay rule. The court explicitly restricted the scope of the injunction to automated search,[102] excluding other forms of search from its scope: "[n]othing in this order precludes [Bidder's Edge] from utilizing information obtained from eBay's site other than by

---

82 MINN. L. REV. 609, 703 (1998) [hereinafter *Fencing Cyberspace*]. She further argues that "the market will determine which sites are open and which are closed." *Id.* Regulators, she argued, should avoid forcing sites to be more or less open, and ought to remain passive. This view overlooks the role of law in shaping the information landscape. Technological means of exclusion are not developed in a vacuum. They are not only a function of technological necessity and availability. Technological development is affected by commercial needs, business models, and legal rules. The same is true for market processes. These are highly shaped by the availability of legal rights and the presence or absence of regulation.

[100] Dan L. Burk argues that recognizing such claims would require any content provider to acquire a license and would lead to a type of zoning. *See The Trouble with Trespass*, *supra* note 94. O'Rourke predicts that the open areas on the Internet would contain less valuable information, while the closed areas would contain the most valuable information subject to technological restrictions. *See Fencing Cyberspace*, *supra* note 99, at 703. Such predictions seem, however, inconsistent with the finding of recent research indicating that open databases are providing high quality information. *See, e.g.*, THE DEEP WEB, *supra* note 26. It also contradicts the common experience of average surfers comparing content on non-profit academic sites with that posted on commercial sites.

[101] *See* Burk, *The Trouble with Trespass*, *supra* note 94, at *48, arguing that "there are public benefits to be had in a cyberspace network that are not captured, and indeed may be destroyed by over-propertization." *But see Fencing Cyberspace*, *supra* note 99, at 703-04, arguing that the market will cause the Internet to consist of open and closed areas, and the legal system should interfere as little as possible with this outcome.

[102] The court issued a preliminary injunction enjoining Bidder's Edge "from using any automated query program, robot, web crawler or other similar device, without written authorization, to access eBay's computer systems or networks, for the purpose of copying any part of eBay's auction database." eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1073 (N.D. Cal. 2000).

automated query program, robot, web crawler or similar device."[103] Nevertheless, as I argued above, physical access is not sufficient to guarantee actual access to information. Access to information in cyberspace is enabled by search engines, and controlling such enablers and their use determines what information is actually made available.

The fact that in *eBay*, access was denied to otherwise publicly available search results suggests that *control over the terms of access*, rather than *access per se*, is at stake. The legal right to exclude granted to eBay facilitates control over the terms of access. The eBay rule allows far-reaching, expansive restrictions on the use of information. Indeed, we sometimes set limits on the use of information that is publicly available, to protect certain entitlements, such as proprietary rights, individual privacy, individuals' right to their good name, or the right of some individuals to control their public image. All these entitlements are subject to restrictions such as free speech. The eBay rule facilitates control over information that incorporates no such balances. We may wish to prevent information mining when it is necessary to protect the individuals to which information is pertained. Yet, when rather than protecting entitlements (such as users' privacy) the court instead creates a broad right to control access to information, it also allows private parties to legally control other aspects of information that should be left uncontrolled, such as its informative value or its meaning.

We live in a world of representations, and the eBay rule allows one to exercise control over representations and meanings. The overall affect of the eBay rule is to induce centralization and concentration trends in the search engine market, which is in great need of facilitating competition. Centralization in the search engine market means centralization in the information environment.

## V. WHAT ARE THE CHALLENGES FOR INFORMATION POLICY IN THE INFORMATION AGE?

While a right to exclude publicly available data is explicitly exempted under copyright law, other legal doctrines are employed for securing control over search results, such as contract, unfair competition, misappropriation, and recently, trespass to chattel. Copyright law was designed to regulate use and information, and therefore it includes some checks and balances informed by the unique character of informational works and their social significance. The displacement of copyright law by

---

[103] *Id.*

other common law doctrines for establishing a right to control access to information is worrisome. Information policy should be informed by the unique characteristics of the information market. Such considerations could be easily obscured, however, when general legal doctrines are strictly applied to create a new type of control over access to information.

The information economy in which we live transformed the definition of market power.[104] Economic power is no longer based on possessing and controlling tangible assets, but is increasingly defined in terms of control over the production and distribution of information. Consequently, it is only natural that competition in this market would focus on gaining control over the new virtual gatekeepers of information. Information policy for the information age should minimize such control.

While modern legal systems recognized the risks involved in centralized communication markets, awareness of the centralization of information markets has not yet matured. Indeed, search engines and site owners could be under no duty to provide their data at no charge. Yet, data regarding the site and its publicly opened content should be freely described, reported, and indexed by others.

To allow individual users on the web not only to freely post their opinions, but also to be heard and read, it would be necessary to actively support decentralized search methods.[105] This could be achieved in a well-functioning, competitive market for search engines, which is free of proprietary rights. Competition is not a magic cure, however, and some regulatory measures might also be required. Yet, once we acknowledge the true meaning of *open access* in the online information environment, it becomes obvious that propertizing raw data and virtual spaces are inappropriate measures. A truly free information environment requires legal policies that would facilitate competition among search engines and minimize attempts to centralize control over the virtual gatekeepers of cyberspace.

---

[104] For further discussion of the term "Information Economy" see Ian Miles, MAPPING AND MEASURING THE INFORMATION ECONOMY, A Report Produced for the Economic and Social Research Council's Programme on Information and Communication Technologies (1990).

[105] Alternatively, this could be done by publicly funded search engines. *See Shaping the Web*, *supra* note 17, at 180-81.