

THIS IS THE ORIGINAL DRAFT SUBMISSION MADE TO THE JOURNAL. IT IS NOT THE DEFINITIVE VERSION OF THE ARTICLE AND IS BEING RELEASED AS BACKGROUND ONLY.

Tyler Handley
MA Digital Society and Culture
King's College,
University of London

P2P Search as an Alternative to Google:

Can nascent forms of P2P search re-capture network value and search practices from monetized search models?

Abstract

This paper examines the contradiction between Google's desire to "database the world's knowledge" and the many ways in which Google's approach affects both the nature of the information users find and how they find it. The paper will argue that Google has monopolized the socially constructed nature of the World Wide Web; Benkler's concept of social production will be used as an example of this process. Google capitalizes on the attention economy, using a combination of PageRank and personalization to dominate the search market. To do so, it must store and retain vast amounts of user data, this data being a representation of the cultural and social relations of Google users. By storing user data in "centralized" logs, Google's approach to search opens up questions about how such sensitive data should be stored, and what the ownership of such a social 'map' by a private corporation means. To further establish the meaning of Google's position this paper outlines the potential for new contrasting forms of search, that allocate more control to the user. In particular, this paper will analyze the Peer-to-Peer distributed search engine YaCy to see how it can alleviate the specific problems of various censorship and filtering that affects Google search results, and how it can address the wider issue of the private appropriation of social and cultural networks. This comparison of Google and Peer-to-Peer search will allow a clear view of the issues at stake as search is developed over the next decade, issues which will have resonating consequences on what information we receive.

Table of Contents

1.0 Search

- 1.1 History of the search engine
- 1.2 How Google won the search environment
- 1.3 Wealth of networks
- 1.4 How Google monopolized the Wealth of Networks

2.0 Why a monopoly on the wealth of networks is problematic

- 2.1 Prone to censorship
 - 2.1.1 Technological filtering
 - 2.1.2 Government filtering
 - 2.1.3 Economic filtering
- 2.2 Advertising and personalization

3.0 Alternatives to Google search

- 3.1 Social search
- 3.2 P2P search
 - 3.2.1 YaCy
 - 3.2.2 How YaCy solves Google's search problems
 - 3.2.3 YaCy's limitations

Introduction

Defining Google as the most visited website in the world fails to give its prevalence the recognition it deserves. As a search engine, its role in our digital lives is of paramount importance. It is the first place many of us look for information on everything from healthcare to shopping. Following from this, how Google presents us with information is of great importance to the fundamental structure of information online. Along with its notable successes, it also presents its users with well defined problems of information diversity, autonomy, and privacy, all of which stem from various censorship and filtering practices.

In this paper I will approach these problems and propose P2P search as a conceivable alternative. Section 1.0 will focus on the history of the search engine, its importance in the Networked Information Economy, and how Google has capitalized upon the internet's wealth for monetary gain. Section 2.0 will focus on the various practices of censorship and filtering that are technologically endemic to Google's ranking mechanisms and central server approach to search. Section 3.0 will analyze alternatives to Google search, lay out five criteria for a search engine to solve Google's stated problems, and then apply the P2P search engine YaCy to these five criteria to discern its efficacy in alleviating them.

1.0 Search

1.1 History of the search engine

The original intent of the various packet-switching networks that came to form the Internet was to share information. The more people used the Internet to share, the more information became available. Unfortunately for users, the Internet lacked (and still lacks) a built in Information Retrieval System (IRT), making information location difficult. Archie was the first attempt at making information online accessible. It visited existing FTP servers and indexed the titles of all of the files, allowing people to find particular documents easier (Halavais, 2008, pp. 21). It did so by utilizing a basic keyword index. Gopher was a further advanced attempt to make FTP servers more accessible; it organized files into hierarchical categories (Halavais, 2008, pp. 22).

The idea of organizing information online hierarchically in categories carried over into the browser-centric environment of the World Wide Web (WWW) through Yahoo!. It relied on human agents to scour the web and then organize its information into themed categories. Such a

system may have dealt with the amount of pre-WWW information on the internet in a useful way, but the commercialization of the WWW, especially after the release of the popular Mosaic browser in 1993, brought plenty more users online to share information. In 1994, Yahoo! "... quickly ran into deep problems, both in terms of scale (impossibility to keep up with the growth of the Web) and ontology (the categorical system could not contain the complexity and dynamism of the information space it claimed to organize) (Feuze, Fuller and Stalder, 2011). Since the architecture of the WWW allowed anyone with a connection to post information, Yahoo! became inundated with new information on a scale not amenable to its search system. 1995 saw the release of AltaVista, a "faster and more comprehensive" search engine that crawled the web with a spider and automatically indexed the information (Feuze, Fuller and Stalder 2011). With AltaVista, users only needed to type in their search request, rather than look through categories for it. It made search simpler.

Although AltaVista and Yahoo! differed in how they organized and presented results, they did share one concept in common: the idea that search should be intertwined with corporate interests. Both sites, and many others that came after, presented themselves as 'portals' through which users could access not only their search results, but also a myriad of other media and services (Feuze, Fuller and Stalder 2011), plenty of which were provided through third-party corporate entities. Acting as 'portals,' search sites had failed to take advantage of the a key concept in the distributed nature of the WWW: sharing. Distributed sharing was the catalyst of the internet/WWW and, as I will argue later, a search site can work much better if it takes a more direct and social approach to defining what content on the WWW users would like to see. Instead, they pushed human edited content and search results that were tied to commercial interests.

1.2 How Google won the search environment

Launched in 1998, Google took a different approach to search than any major engine before it. Instead of acting as a web portal, it offered users a clean interface, devoid of anything but the logo, a search box, and two buttons (search, and I'm feeling lucky) (Weinberger, 2012).

Google's ranking algorithm returned much more accurate results than any other search engine. This accuracy came from Brin and Page's PageRank algorithm, which carried out an "...objective rating of the importance of websites, considering more than 500 million variables and 2 billion terms" by interpreting hyperlinks between websites as votes cast for one another (Google, 2008).

1.3 Wealth of Networks

The democratic underpinnings of PageRank are unquestionable. Each node in a network casts votes as hyperlinks, the accumulation of which nominates an “elected” authority over information. Concepts of democracy have always been closely tied to theories of the Internet, stemming from the inherent distributed nature of the WWW that allows any individual with a connection to participate.

Since PageRank uses hyperlinks as votes, and since these links are placed by individual users of the WWW, PageRank embodies what Benkler (2006) calls The Wealth of Networks.

In The Networked Information Economy, average citizens have the means of production and distribution that were once only available to people working within the confines of the Industrial Information Economy institutions/organizations. As a result, the spread of information and culture is no longer primarily shaped by market-based or government based actors.

The rise in the non-proprietary use of networks to share culture and information, and the resulting distributed ownership of “material requirements” for producing and sharing takes advantage of the distributed nature of the WWW. The Networked Information Economy has led to “effective, large-scale cooperative efforts” like Wikipedia, open source software, and peer-to-peer networks. It is these distributed efforts, along with the ability to do so - and their output - that represents the Wealth of Networks.

1.4 How Google monopolized the Wealth of Networks

The Wealth of Networks is a public space made up of the sum of human knowledge that exists online. As a search engine, Google doesn’t produce any of its own wealth. Instead, it extracts wealth from the WWW, indexes it, ranks it, and then presents it back to the users who initially uploaded the content (Pasquinelli, 2009; Jakobssen and Stiernstedt, 2010). More precisely, Google identifies the network value produced by the wealth accumulated through the social interactions of nodes in the WWW and then uses it as its own source of wealth. It is part of a wider trend in Social Media that “mark[s] a shift to a new economy in which value is not embedded in social relations but in which social relations are a primary source of value” (Stark, 2009 pp. 173).

Where mass-media broadcasters in the Industrial Information Economy produced their own content and then used the attention they gained from it as a commodity to be sold to advertisers, Google pools content from the WWW and uses the traffic from people seeking that content as a

commodity to sell to advertisers. It re-captures information that comes from the distributed structure of the internet to sell as a commodity.

It does so with the help of the Google AdWords and AdSense services. Google, having dominated the search market and consequently having millions of people view its homepage per day, was in the perfect position to revamp the advertising industry for the online environment. Being such a large site with an estimated 900 thousand servers (Kooimey, 2011) also meant that it needed to monetize[1].

Google's idea with AdWords was to make advertising less intrusive and more accurate. They did so by making the sponsored results look very similar to the organic results, only delineated by appearing to the right and top of organic results, and by being discretely labeled "sponsored ad" (Levy, 2011, pp. 91). The system made sure that only advertisements relevant to the particular search query appeared.

AdWords harnesses the attention given to search results pages. It has immense power, considering that Google is the most visited website in the world (Alexa, 2012). AdSense differs in that it allows anyone with a website to display behavioral advertising. Compared to traditional "banner advertising," where the contextuality of an ad was mostly guesswork, AdSense ads nearly always appeal to the visitor, who is then more likely to actually click through the ad.

AdSense works by dropping a cookie onto a user's hard-drive when they visit a page that contains an AdSense ad. This cookie then provides Google with a wealth of valuable information. This information is combined with a user's search results (from the Google search page) to form a comprehensive log of information on said user (Levy, 2011, pp. 335). This information is not only used to display even more specific behavioural advertising to the user, but is also used in a recursive manner to constantly improve Google's search results and behavioural advertising[2].

2.0 Why a monopoly of the Wealth of Networks is problematic

It's evident that Google search is no longer so much about giving the world access to any knowledge at the click of a button, but rather about how they can continue to maximize profits by building ever more extensive logs about individual users.

These logs are a form of "dataveillance," which gives Google a wealth of information that "allows analysis to inductively construct the audience for sale" (Shaker, 2006). Consequently, Google is able to take a confusing mess of information and transform it into accessible categories for advertisers. This act is representative of a shift away from the "monolithic structures of state-sur-

veillance” towards a more dispersed organization of surveillance, typified by corporate entities (Haggerty and Ericson, 2000). Where as methods and tools of surveillance were once only available to the state, global corporate entities now have them at their disposal as well, and are using them to monitor citizens for monetary purposes (Röhle, 2007).

Google states that the main reason they collect and hold user data is to enhance ranking algorithms (Varian, 2008). Indeed, as Hoofnagle (2009) points out, they have strong incentives to collect a wealth of data to expand their advertising based business model. In this regard, “Innovation is raised as a privacy tradeoff in the context of data retention” (Hoofnagle, 2009).

It is this contestation between innovation and privacy that exemplifies Google’s ethos in privacy matters; they must not only provide the user with the best results possible, but also make sure not to violate their privacy by doing so[3].

It is undeniable that wherever Google stores its user logs is of great interest to third-party sources, perhaps even ones with criminal intent. Such was the case when hackers, later traced back to locations within China, gained access to sensitive information in Google’s servers. The hackers stole both valuable source code and access to the Gmail accounts of Chinese political dissidents and human rights activists (Levy, 2011, pp. 269)[4].

2.1 Prone to censorship

User logs located at “central[5]” locations are also prone to censorship and filtering practices, which consequently affects the quality of information users receive from the search engines themselves. These practices come from four sources: the technology that surrounds search, the Governments within which search servers reside, the monetary intent of search engines, and the mass-media economy that fights back against modern search. In this next section I will outline these four censorship and filtering practices, explain why they affect the sort of information users receive, and also delve into how advertising and increasing personalization further the problem.

2.1.1 Technological filtering

Halavais (2008, pp. 87) points to a variety of studies suggesting that searchers “satisfice” when looking for information. They won’t seek out the best answer, rather one that is simply good enough given the small amount of time they are willing to spend. Guan and Cutrell (2007) discovered - through eye-tracking studies - that users focus much more on the top results of a

search page than the ones further down. Taking these points into account we can say that a large portion of people who use Google are looking for “basic information needs” and are willing to settle for the easiest answers, which - given Google’s accuracy - are most likely at the top of the search results. It follows that whatever information is first presented to users at the top of a Google result page is of great importance, for that is the information they are most likely to consume.

Google’s results pages follow a Power Law pattern where a select few websites dominate the top results. Hindman et al. (2003) call this phenomenon, a “Googlearnarchy.” A Power Law structure is typified by a network in which “...most nodes will be relatively poorly connected, while a select minority of hubs will be highly connected (Watts, 2003 pp. 107). This network structure is further re-enforced by the aforementioned “satisficing” where searchers will be content with only looking at the first few Google search results.

Rogers (2009, pp. 176) showed, drawing on a search query for terrorism, how the top Google results were often self-referential. The top ranked site is often Wikipedia, which has cited the top news sites, as they already had journalistic authority. Thus, in a recursive manner, both Wikipedia and the news-sites act as link-farms, promoting each-other to the top through PageRank link value. Metahaven (2009, pp. 189) also shows how Google’s list of results, specifically the top 10, “...harnesses a preference for sources, many of which have become authoritative for their social structure.”

In another study, Hindman et al. (2003) crawled three million pages, organized the indexes in a manner similar to PageRank (links as votes), and then analyzed the link structure around controversial topics like abortion and gun control. They found that only a select few sites accrued most of the incoming links. These would be the top-viewed sites on search engine landing pages.

The scientific reference based voting structure of the PageRank algorithm constitutes a technological filtering that leads to the suppression of an estimated 80% of the information on the WWW (Ratzan, 2006). It is a “...form of power both more sneaky and more structural than old-fashioned coercion”, which “...suppresses alternatives without coercion being needed (186 Metahaven). Indeed, vote-based hyperlinking is a system of control without a face, one which is difficult to define unless one is technologically literate, but one that humans have themselves created, and must therefore adhere to. PageRank confirms Lawrence Lessig’s (2006) famous quote “code is law.”

2.1.2 Government filtering

Government pressure on Google in Europe and North America - aside from corporate pressure which I will address later - has minimal impact on the daily lives of most citizens. At present, these are the countries that, for the most part, aren't overtly affected by having a centralized server search engine like Google. The citizens living under authoritarian regimes are the ones who are affected by Government search censorship. Since China is the largest search market in the world (Internet World Stats, 2012) and because of Google's well documented tension with their government, I will use Chinese Government censorship as a case study to exemplify the problems citizens face when attempting search in a hostile environment.

China first uses a "porous network of internet routers" that filter blacklisted keywords. This is more widely known as the "Great Firewall of China." Like many other governments around the world, they use basic filtering techniques such as Domain Name System (DNS) tampering and Internet Protocol (IP) blocking in their firewall servers (OpenNet Initiative, 2009a). Unique to Chinese censorship is the practice of TCP reset filtering. Routers in the firewall identify black-listed keywords that were typed by internet users then break the connection from the user's intended destination back to the user (OpenNet Initiative, 2009a).

For websites residing outside of the firewall, this means that any term on their site deemed controversial by Chinese authorities will not be let into the country. This greatly affected Chinese search results on Google and is why, in 2006, Google obtained a Chinese business license to launch a Google.cn domain (Human Rights Watch, 2006). Residing within the firewall meant that search queries didn't have to pass through the entirety of the firewall. Unfortunately, it also meant that Google.cn would have to "...police their own content under the penalty of fines, shut-down and criminal liability" (OpenNet Initiative, 2009b).

China forces all Internet Content Providers (ICPs) to obtain a license before they can legally provide access to online content. A condition of the license is that the ICP must "...prevent the appearance of politically objectionable content through automated means, or to police content being uploaded by users for unacceptable material" (Human Rights Watch, 2006). To do so, companies employ people to create "block-lists" of what they expect the government to find objectionable. This type of "intermediary liability" is referred to by MacKinnon (2012, pp. 241) as "networked authoritarianism."

More ominous than filtering through the "Great Firewall" and ICP liability is the widespread implementation of the Green Dam software in China. It blocks "access to a wide range of web sites based on keywords and image processing, including porn, gaming, gay content, religious sites and political themes" (OpenNet Initiative, 2009a). It also has the ability to monitor computer behaviour on a personal PC level; it can view keystrokes done in software like Microsoft

Word and then terminate the application if it detects blacklisted words (OpenNet Initiative, 2009a).

In essence, the Green Dam software acts in a similar manner to P2P distributed networks in that it functions at a computer's local level to filter sensitive material. Where distributed computing is often touted as a democratizing technology, China realized its potential through the lens of censorship.

Fortunately, the use of the Green Dam software in personal PC's was never fully realized. The Minister of Industry and Information Technology announced that the aforementioned mandatory installation of the software would not be put into legal effect (OpenNet Initiative, 2009a). However, public computers such as those in Schools and Internet Cafes would still be required to have the software installed (OpenNet Initiative, 2009a). This is troubling considering that 42 percent of Chinese computer users access the internet from cafes (China Internet Network Information Center. 2009).

2.1.3 Economic filtering

As exemplified in the previous section, it is very difficult for average Chinese citizens to retrieve search results that aren't heavily influenced by Government control. The Chinese government has tight control over what information goes in and out of the country. Unlike Government censorship, which generally only affects what residents of a particular country can see, economic censorship has global reach - most notably through the Digital Millennium Copyright Act (DMCA) provisions aimed at combating illegal file-sharing, which have turned search engines into an "instrument of international power" (Halavais, 2008, pp. 129). DMCA takedown notices are the number one reason for content removal on Google, with 97% of the 3.3 million requests in 2011 being complied with (Rushe, 2012).

As mentioned previously, Google nearly always takes content down when a valid court order is issued. They've even gone so far as to "downgrade websites that persistently breach copyright laws" (Google, 2012b). Search results which link to files are taken down globally. When one search engine, such as Google, dominates the market by such a large margin, the fact that one country's legal requests can affect what citizens in other countries access is cause for concern. This is not to say that search engines are made for illegally sharing content, my point is to highlight the fact that corporate interests - particularly of the traditional Mass Media sort - can have plenty of power over information in the Networked Information Economy, even when they no longer own the means of production.

2.2 Advertising and personalization

One of Google's mantras is "to give you exactly the information you want right when you want it" (Google, 2008). They drastically improved searching online, but not everyone searches for the same results - people using the same query might be looking for different information. To tackle this problem, Google began to "...personalize search in order to deliver more relevant results to the users" (Feuze, Fuller and Stalder, 2011). Now, "...results are tailored to one's tastes, based on search history and results clicked" (Rogers, 2009, pp. 180).

An upside to personalization is that it diminishes the effect that the power-law structure of PageRank has on results. By factoring in past searches and user interests, the search results are less oriented towards the most popular results on the web and more oriented towards outlying sites that fit well with a user's taste profile. Personalization relieves Google of accountability for results returned. Users have partly themselves to blame for what their search query returns.

The downside to personalization is that it diminishes autonomy in search, constituting "...an obscure iron numeric cage that constrains users' freedom and their capacities of determination" (Lobet-Maris, 2009, pp. 81). Feuze et al. (2011) showed how, using three newly created Google accounts populated with a combined 195 812 individual search queries, personalized search results developed over a period of time. They found that over only a small amount of time the personalized results were glaringly different between the three accounts. They also found that:

"...Google is actively matching people to groups, which are produced statistically, thus giving people not only the results they want (based in what Google knows about them for a fact), but also generates results that Google thinks might be good to users (or advertisers) thus more or less subtly pushing users to see the world according to criteria pre-defined by Google."

As Röhle (2007) points out, such personalization "...implies an expansion of surveillance in the interests of commercial actors." In one of the first papers Sergei Brin and Larry Page wrote about Google they deplored mixing advertising with search, stating that the "...issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm" (Brin and Page, 1998). Contrary to this, the integration of personalization as a key ingredient in Google search is partly in place for commercial incentives, indicating that Google's search results are constrained by monetary persuasions.

3.0 Alternatives to Google search

As MacKinnon (2012, pp. 498-499) advocates, the best “counterweight” to corporate power on the internet is a strong digital commons. “A robust digital commons is vital to ensure that the power of citizens on the Internet is not ultimately overcome by the power of corporations and governments” (Mackinnon, 2012, pp. 508-513). Being gatekeepers of information, search engines form an integral part of the digital commons. They are an opportunity to realize the original democratic potential of the internet; a system that allows anyone to share information freely from node to node, without coercion in between nodes. As I will show in section 3.2.2 the greatest way of achieving this via search is to distribute it with a robust P2P network.

From the perspective of fostering a stronger digital commons, the most important components of a search engine are the ones that allocate users greater autonomy in choosing what they search, how they search, and what results they find, free of control from certain technological, government, and corporate constraints (Röhle, 2009, pp. 129). The five requirements for a digital commons based democratic search engine are that it be free from:

1. Technological filtering induced by PageRank-like algorithms.
2. Government censorship.
3. Excessive global economic censorship by copyright holders.
4. The filter bubble induced by personalization and behavioural advertising (monetary incentives cut across this).
5. Potential breaches of privacy by the storage of vast user logs in central servers.

3.1 Social search attempts. Other search engines

There are many other search engines trying to solve these problems in one way or another. Although I am focusing on Google, a brief outline of these other sites is an important segue into P2P search. Google dominates the search environment, so to compete with it most other search engines either try to copy it or capitalize on a niche market that Google has missed. These niches often involve attempts to resolve some of the issues I have outlined. For example, Yippy divides results into thematic clusters in the hope that users can find deeper associations between topics, thus relieving some of the deep web suppression that Google’s algorithms cause. Also, they don’t track users, thus safe-guarding their privacy.

Although not the main premise of Yippy, its function represents an opposition to Google’s way of suppressing the deep web. This fits well with the ideal that a more democratic search engine should make it so that users are “...able to grasp where the borders between...social currents

persist, and where they consent or diverge” (Metahaven, 2009, pp. 196). DuckDuckGo takes a different approach by focusing on popping the users Google induced “filter bubble” and ensuring complete privacy in the process - it doesn’t create user logs (DuckDuckGo, 2012). Unfortunately, there is advertising on both sites, and they are both still conducive to government and global economic censorship[6].

There are also “social” search engines that return queries based off of collaborative filtering. Swicki is a mix between a wiki and a search engine that relies on users to build search databases around certain subject areas (RÖhle, 2007). The Open Project is a web directory edited by users, somewhat similar to Yahoo!’s original premise. With more than 60 000 editors it is comprised of more than 4.5 million websites (RÖhle, 2007). Social search sites address the issue of technological filtering but fail to address any of the other issues of censorship. In fact, they appear to be more interested in further personalizing search, which is good for personal autonomy and democracy, but only within small clusters.

3.2 P2P Search

In this final section I will argue that Peer-to-Peer (P2P) search is currently the only method of search capable of resolving the five issues I have outlined[7]. I will then use the P2P search engine YaCy as a case study to test my claims. First, we must understand the fundamental characteristics of P2P.

A P2P network is a collectively produced structure of information that is accomplished through the formation of a “chain of interconnected applications” by individual users (known as peers or nodes), who share both resources and personal computing power (Rigi, 2012; Loban, 2004). A P2P network does not rely on centralized network servers like in a traditional web-client/server system (Svensson and Bannister, 2004). This “decentralization” means that nodes in a P2P network only share information with other nodes, the result of which is a network without a centre.

The wealth produced by a P2P network, through a distribution of labour (Rigi, 2012), cannot be capitalized on for monetary purposes. The immaterial labour essentially propagates itself, leading to a network more democratic in nature - and more closely related in structure to the early days of the Internet - than traditional central server networks such as the WWW. The potential power of distributed computing networks like P2P are vastly superior to other models, but they depend on how many people decide to participate. For example, Seti@home is a distributed NASA initiative that uses people’s spare computing power to compute signals from space, in search of extra-terrestrial communications (SETI@Home, 2012). It holds the Guinness world record for the largest data computation in history (Newport, 2005). If the distributed power of

Seti@home's magnitude could be used for web-search, not only would it alleviate the monetary constraints of maintaining large server farms, but it would also provide more computing power.

There are currently three different distributed P2P approaches to web search. FAROO is the largest of the three with more than 2.5 million peers (FAROO, 2008). Users install the software on their computer. Once active, every time they visit a website it is logged in an index. The index is then used to automatically compile and adjust - in real-time - search rankings across the FAROO network. Because it ranks pages automatically when a user visits a website, it is simple and democratic. Unfortunately, FAROO is relatively new so there is a lack of in-depth understanding of its inner-workings and their implications. According to their website, FAROO retains no search logs and is immune to censorship because of encrypted search queries and indexes. However, they also mention that they use "privacy protected behavioural advertising." Advertising in FAROO may not affect search results like it does in Google, but there is still surveillance in the pursuit of monetary gain, capitalizing once again on the wealth of networks, this time the wealth being a composition of its 2.5 million users. The lack of server costs associated with P2P networks negates the need for advertising revenue beyond what needs to be paid to the developers for coding and upkeep. There is no doubt of the democratic potential and innovative search mechanisms of FAROO, but its inclusion of advertising is a point of contention.

Seeks is another P2P web search application that can be installed on a personal computer. Like FAROO, it builds a social search network around users' online activity (Seeks, 2011). Based on search queries through other search engines, what results are clicked, and web search, Seeks builds a personal profile that is stored on the user's personal computer. This profile is used to filter search results. Personal profiles are placed into groups according to shared interests which further filters results. Within these groups users can filter even further in a collaborative manner and also comment on queries. Users can also link their own website or a company to any group or interest so that it is more likely for other users with shared interests to find that particular site (Seeks, 2011). Taking all of this into account, Seeks is more of a P2P social network built around searching than a distributed search engine. It offers plenty of autonomy for the individual user but scrapes its results from other search engines, meaning that Government and global economic censorship of Seeks' results would be the same as censorship of central-server search engines.

3.2.1 YaCy

Out of the three P2P search applications, YaCy most closely resembles a traditional search engine. Once installed on a user's computer, they gain the tools necessary to participate in every aspect of a traditional search engine. Running in private mode, a user can use the built in crawler to crawl the websites of their choice, the regularity of which can be set so that indexes

are always up-to-date. Once crawled, the indexes are stored on their personal computer. They can then search their local index, which is comprised of only the sites they themselves have crawled (YaCy, 2012a). Combined with the ability to adjust weighting to a wide-variety of different ranking parameters, YaCy can essentially act as a personal search engine, with nearly full customization.

In public mode, the user's crawl index is shared with every other peer acting in public mode, fittingly called the "freeworld." Unlike Google, which stores its search indexes in central servers, Yacy stores different small bits of its index in every single users' hard-drive. "There is no central body of control in YaCy" (YaCy, 2012b). Indexes are encrypted with a key and then placed into a Distributed Hash Table (DHT) which shares information with other nodes in the "freeworld." "This allows index data to reach the peer before a query for that information is even submitted" (YaCy, 2012b), meaning that every peer has a miniature version of the entire index in their hard-drive. When information is missing the peer will "call out" to other peers in the network for the particular information. These queries and requests traveling through the freeworld are all encrypted to safeguard user privacy (YaCy, 2012c).

3.2.2 How YaCy solves search problems

1. Technological filtering induced by the PageRank-like algorithms

Although like many other search engines, YaCy attempts to emulate the PageRank algorithm, it doesn't put as much weight on the PageRank score that it does for more traditional types of index ranking. The dearth of user settings gives users more autonomy to add their choice of weighting structure to their personal index[8]. "[E]veryone can assess the quality and importance of web pages by their own rules and adjust to their personal relevance as a ranking method..."(YaCy, 2012c).

Building on top of more traditional indexing methods, and since the global index in YaCy is a compilation of only what users have crawled, there is a wider variety of content than the organizational-web-dominated Google results. Because of this not only will the Power Law structure of PageRank results be less prominent, but there will also be an attenuation of deep web suppression[9].

2. Government censorship

Unfortunately, as much as YaCy claims to be a fully distributed P2P network, its multi-agent structure, using DHTs, means it must rely on four predefined servers to coordinate node lists,

much like Torrent files rely on trackers for downloading (Rudomilov and Jelenik, 2011). This is not to take anything away from YaCy, as no global network is fully distributed, but it does mean that it isn't fully immune to Government censorship. However, shutting down the server lists would be difficult, as they are likely stored in different locations to avoid such risks.

Also, since there is no central control and no storage of the index itself in central servers, there is no one place from which to censor particular results. To censor content a Government would need to push other countries to help them censor the four node lists or go after thousands of active peers. "YaCy results can not be censored as no single central authority is responsible for them and there are thousands of servers (personal computers) in multiple countries providing results (YaCy, 2012a).

The most comprehensive way for a Government to censor YaCy is to use distributed methods such as the Green Dam initiative. If the Chinese authorities had been successful in deploying it across the country it would have cut off access to the benefits of YaCy on a local level. For example, a user typing Falun Gong into the YaCy search box would be blocked by a keystroke censor, thus destroying the search.

3. Excessive economic censorship by copyright holders

Just like with Government censorship, corporations have no central servers to approach in regards to removing content. In fact, once content is placed in YaCy's DHT, it is out of a corporation's control indefinitely, as pieces of the freeworld index are stored in every node's computer. Where Google removes the links to copyrighted material in its search results if issued a DMCA take-down notice, YaCy - because there is no central server - is still able to make such links visible. There is no one node to target.

That being said, such files are usually located on locker-box type music sites, video streaming sites, or torrent search engines/trackers. YaCy is simply a search engine that would crawl these sites. In this respect, it can only bypass economic censorship so far as the sites of the original upload can. However, because it can scour deeper parts of the web, it is likely that it can find more obscure file uploads not targeted by copyright holders.

4. The filter bubble induced by personalization, behavioural advertising, and monetary incentives

Like Google, YaCy presents its users with a filter bubble. Yet, the abundance of adjustable ranking parameters and weighting in YaCy allows the user much more autonomy in how their filter bubble is defined, and conversely how the filter bubble defines them.

YaCy, once again because it doesn't rely on central servers, and has a small development team, doesn't need to make abundant amounts of money to keep servers running. It's also funded by the Free Software Foundation Europe (FSFE). By relying on users' spare processing power, they can save the vast sums of money that Google must pay to run their own. YaCy doesn't need to advertise, which means that no search returns are ever done within the constraints of behavioural advertising, where users are grouped according to interests. Much more of the search content is "...determined by the users, not by commercial aspects of the Web portal operator" (Yacy, 2012c). It also means that "search requests are never stored, monitored or evaluated for commercial purposes" (YaCy, 2012c).

5. Potential breaches of privacy by the storage of vast user logs in central servers

Because of the aforementioned absence of user log storage and centralization of servers/tracking in YaCy, there is no data that can be gathered by third party sources to monitor either an individual user or even the "freeworld." No source can pinpoint the node from which a query originated, nor can they even see what such queries were, as "YaCy does not store words in clear text but only as word-hashes" (Christen, 2010). Additionally, every search query is encrypted on its way out from a node and on its way back to a node (YaCy, 2012c).

3.2.3 YaCy's limitations

YaCy can solve the various censorship and filtering issues that occur with Google search, but it also has several crucial drawbacks. First, just as in any P2P network, the power of YaCy is directly correlated to how many people use it. The amount of websites searchable in the free-world relies on this. Similarly, users' activity rate greatly affects the speed at which search results are returned. It is not a search engine that can be both passively used and successful. It requires that its users be active by contributing to the freeworld index. "[T]he quantity and quality of the results will depend on the number of peers connected at the time" (YaCy, 2012a).

As shown in section 2.1.1 Google searchers have a habit of "satisficing." If users were to do this in YaCy, it could never reach the level of power and variety it would need to make it a viable alternative to Google. Not only do users have to actively contribute, but there is also a learning curve involved in taking part. It has been pointed out that YaCy is only for expert users and

early adopters (Kanjilal, 2011). Based upon the “satisficing” of search, where users are happy with a decent search return in the shortest amount of time, the fact that YaCy requires a user to download it, install it, and then learn their way around a complex graphical user interface (shown in Fig 1.0) to use it, average searchers will likely be unwilling to put in the required effort to switch to YaCy.

Fig 1.0

Perhaps YaCy’s strongest feature is also one of its biggest drawbacks. Because it doesn’t store any of its information in central locations it can’t assess data to improve itself. Google is well known for its recursive learning algorithms that consistently up-date the quality of search (Levy, 2011, pp. 46). YaCy is not capable of being a recursive learning machine. By using YaCy, users sacrifice constant innovation for the benefits of guaranteed privacy and increased autonomy.

Concluding Remarks

To conclude, we can ascertain that YaCy provides users more autonomy with the information they find, how they find it, how they organize it, and how they relate it to themselves through the sharing of queries and logs. They can break free from the constraints of search engines like Google which impede such autonomy through filtering/censorship practices, and a strong slant towards treating users as nodes of attention to be bought and sold. But first, users have to overcome the three crucial faults of YaCy: it relies on an active user base, it has a steep learning curve compared to traditional search, and it can’t innovate like Google.

Brin and Page started Google with the noble intent of making information more freely accessible to all (Levy, 2011), but they chose a monetary route that relied on vast server farms and an indexing method that structured search results in such a way as to actually suppress much of the information online. This is not to say that the PageRank algorithm isn’t extremely powerful at finding the most popular types of websites, but if the intent of making a search engine is so that information is more freely accessible, choosing a route that doesn’t elicit issues of user autonomy and monetization may have been as simple as changing the internal structure of the search engine from a central-server approach to a P2P approach. Judging from Page’s well documented obsession with the unfortunately poor life of inventor Nicola Tesla (Levy, 2011, pp. 107), it’s no stretch to deduce that Page had always intended to make money from Google. Indeed,

as Erich Schmidt (Google's CEO) has stated, Google is first and foremost an advertising system (Vogelstein, 2007).

Unfortunately, for the time being, YaCy is in a fledgling stage of development; it is only a potential alternative to the big search engines. There's no indication that - with its rather lacking user base, steep learning curve, and slow innovation - it could ever challenge Google's scale and accuracy. But it's important that these new forms of search continue to strive towards solving some of the core issues surrounding Google. The potential democratic power that can be allocated to the average user in their endeavor for information is cause enough to support these search engines, even if used alongside Google for the time being. And in the decades to come, we can hope that they will act as a strong foundation upon which to build a robust digital commons, free from excessive monitoring and censorship, and closer in structure to the original packet switching networks that made up the early internet. With their reliance on users for power, the only solution to building our ideal of a robust digital commons is participation, the bedrock of democracy itself.

It is my hope that this paper has expanded constructively upon the discussion surrounding new models of search, in particular P2P's contribution. In fact, I see it acting as a foundation on which future studies of P2P search can be conducted. To further analyze the efficacy of such search models, it is important that future studies delve deeper into the mechanics of ranking algorithms and privacy enhancement. In regards to ranking mechanics, we must conduct user studies to better understand the types of information users of P2P search are looking for and then compare the findings against search results. We also need to compare these to the search results of Google. P2P search can only be successful in so far as it is actually useful to the average citizen. And in regards to privacy enhancement, further research needs to be undertaken to understand the intricacies of encryption in P2P search technologies. How strong is it? Can it be exploited? Such studies, and others that will surely be developed, are integral to an understanding that will go towards challenging the monetary influenced search engines that have run away with the internet's wealth.

Notes

[1] This monetization has no doubt worked. Google's 2011 revenue from advertising was \$36 Billion USD (Google 2012a). However, there may have been alternative routes Google could have taken, ones that were nonproprietary.

[2] These logs are anonymized - with only the IP address of the user viewable - and then deleted after 9 months. However, if a user is signed in with their Google account, and if their

name is provided somewhere in that account, then the logs can be matched to that user's real name.

[3] In public forums, Google have actively tried to protect user data. Their bottom line rests upon maintaining user trust. The Electronic Frontier Foundation gave Google a gold star (gold star being the best score possible) in its analysis of how internet companies treat their users' privacy (EFF, 2006).

[4] The Operation Aurora attack wasn't isolated to Google - at-least 34 other technology companies fell victim (Cha and Nakashima, 2010).

[5] By "central" I simply imply that the logs are located together rather than truly distributed like in a P2P environment, even though they may be dispersed across the world for need of redundancy.

[6] Yippy even censors plenty of its own content for political and religious purposes (Judic, 2010),

[7] I would like to note that this may very well change, as the search environment evolves at an ever-quickening pace. Therefore, section 3.1 would be best understood within the timeframe it was written (July 2012).

[8] A possible issue is that the default setting in Yacy uses a PageRank type algorithm combined with the fact that most people use the default settings in their software. But most of Yacy's users can be considered "early adopters" as Yacy takes several steps to install and an elementary knowledge of routing Internet Protocol addresses (Ips). The result of this is a user base who would be more likely to adjust default settings.

[9] An interesting future study would analyze the top few pages of Yacy results in relation to those of Google, using a wide variety of, and a substantial amount of, queries to understand the type of content crawled. I hypothesize that Google's results would be much more organizational and populist in nature, where Yacy results would have more fringe information from the "outlying" nodes of the WWW.

Works Cited

Abadi, J., 2009. *Data Management in the Cloud: Limitations and Opportunities*. Ph. D. Yale University.

Alexa, 2012. Top 500 Sites on the Web. [online] Available at <<http://www.alex.com/topsites>> [Accessed June 15 2012]

Articles 19, 21, 23. Regulations on the Administration of Business Sites Providing Internet Services (Hulianwang shangwang fuwu guanye changsuo guanli tiaolie). Issued by the State Council on September 29, 2002, effective November 15, 2002.

Ball, P., 2006. *Critical Mass: How One Thing Leads to Another*. New York: Farrar Straus and Giroux.

Bamman, D., O'Connor, B., Smith., N., 2012. Censorship and deletion practices in Chinese social media. *First Monday* [online]. 17(3). Available at: <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3943/3169>>

BBC News, 2012. China Arrests Thousands in Latest Internet Crime Crackdown. *BBC News Technology* [online] (Last updated 11:59 AM on 26 July 2012) Available at: <<http://www.bbc.com/news/technology-18996811>>

Benkler, Y., 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press, pp. 1-35.

Brin, S., Page, L., 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In: *Proceedings of the 7th International World Wide Web Conference*. Brisbane, Australia 14-18 April 1998. pp. 107-117.

China Internet Network Information Center. 2009. Statistical Survey Report on the Internet Development in China [online]. Available at: <<http://www.cnnic.net.cn/uploadfiles/pdf/2009/3/23/131303.pdf>> [Accessed 28 June 2012]

Christen, M., 2010. Web Search By Te People, For The People. In: SFCS (Society for Free Culture and Software) Conference 2010. Göteborg, Sweden 5-7 Nov 2010.

Darnton, R., 2009. The Library in the Information Age: 6000 Year of Script. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 32-45.

DuckDuckGo, 2012. About [online]. Available at: <<http://duckduckgo.com/about.html>> [Accessed 4 Aug 2012]

Duguid, P., 2009. Search before grep: A Progress from Close to Open?. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 13-32.

EFF (Electronic Frontier Foundation), 2006. EFF Applauds Google Resistance to Government Subpoena, But Broader Privacy Concerns Remain. [press release], 19 Jan 2006, Available at: <<https://www.eff.org/press/archives/2006/01/19>> [Accessed 20 June 2012]

FAROO, 2008. FAROO presents Peer-to-peer We Search at the TechCrunch50 Conference. Press release, 7 Sept 2008.

FTC (Federal Trade Commission), 2007. Online behavioral advertising moving the discussion forward to possible self-regulatory principles. [online] Available at: <<http://www.ftc.gov/opa/2007/12/principles.shtm>> [Accessed 28 May 2012]

Feuz, M., Fuller., M., Stalder., F., 2011. Personal Web searching in the age of semantic capitalism: Diagnosing the mechanisms of personalisation. *First Monday* [online], 16 (2). Available at: <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3344/2766>> [Accessed 28 June 2012]

FNF (Free Network Foundation), 2012. About. [online] Available at: <http://freenetworkfoundation.org/?page_id=738> [Accessed June 5 2012]

Fraim, J., 2002. Internet Words and Culture. *First Monday* [online]. 7(6) Available at: <http://firstmonday.org/issues/issue7_6/fraim/index.html> [Accessed 28 June 2012]

FSFE (Free Software Foundation Europe), 2012. About. [online] Available at:
<<http://fsfe.org/about/about.en.html>>

Guan Z and Cutrell E, 2007. An eye tracking study of the effect of target rank on web search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, California,: ACM Press, pp. 417-20

Gerhart, S., 2004. Do Web search engines suppress controversy?. *First Monday* [online], 9(1). Available at:
<<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1111/1031>> [Accessed 20 May 2012]

Google, 2008. Corporate Information. [online] Available at:
<<http://www.google.com/corporate/tech.html>> [Accessed 2 April 2012].

Google, 2012a. AdWords. [online] Available at: <<http://www.adwords.google.co.uk>> [Accessed 2 May 2012]

Google, 2012b. An Update to Our Search Algorithms. *Google Inside Search* [blog], 10 Aug. Available at: <<http://insidesearch.blogspot.co.uk/2012/08/an-update-to-our-search-algorithms.html>> [Accessed 15 August 2012]

Google, 2012c. Transparency Report. [online] Available at: <<http://www.google.com/transparencyreport/>> [Accessed 10 June 2012]

Hafner, K., Richtel, M., 2006. Google resists U.S. subpoena of search data. *N.Y. Times*, 20 Jan.

Haggerty, D., Ericson, R., 2000. The surveillant assemblage. *British Journal of Sociology*. 51(4) pp.605-622.

Halavais, A., 2008. Search Engine Society (DMS - Digital Media and Society). Cambridge: Polity.

Head, A., Eisenberg, M., 2011. How college students use the Web to conduct everyday life research. *First Monday* [Online], 16(4). Available at: <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3484/2857>> [Accessed 25 May 2012]

Hiltzik, M., 2012. Google may not be evil, but it's also not trustworthy. *Los Angeles Times Business* [online]. 13 May. Available at:
<<http://articles.latimes.com/2012/may/13/business/la-fi-0513-hiltzik-20120513>>

Hindman, M., Tsioutsoulis, K., Johnson, A., 2003. Googlearchy: How a Few Heavily Linked Sites Dominate Politics on the Web. In: Annual Meeting of the Midwest Political Science Association. Chicago, USA 2003.

Hoofnagle, C.J., 2009. How policy makers, journalists and consumers should talk differently about Google and privacy. *First Monday* [online], 14 (4). Available at: <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2326/2156>>

Human Rights Watch, 2006. Race to the bottom: Corporate complicity in Chinese Internet censorship. *Human Rights Watch* (technical report), 10 August. Available at: <<http://www.unhcr.org/ref-world/docid/45cb138f2.html>> [Accessed 1 June 2012]

Internet World Stats, 2012. Top 20 Internet Countries with Highest Number of Users. Internet World Stats Usage and Population Statistics [online]. Available at: <<http://www.internetworld-stats.com/top20.htm>> [Accessed 5 July 2012]

Jakobsson, P., Stiernstedt, F., 2010. Pirates of Silicon Valley: State of exception and dispossession in Web 2.0. *First Monday* [online], 15 (7). Available at: <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2799/2577>> [Accessed 8 May 2012]

Jeanneney, J.N., 2006. Google and the Myth of Universal Knowledge: A View from Europe. Chicago: University of Chicago Press.

Judic, L., 2010. Search Engine Update: Who's Out There Shaking Things Up?. Search Engine Watch [online], 30 July. Available at: <<http://searchenginewatch.com/article/2050683/Search-Engine-Update-Whos-Out-There-Shaking-Things-Up>> [Accessed 5 Aug 2012]

Kanjilal, C., 2011. Ambitious Decentralized Projects that Aim to Create a Better Internet. *techie buzz* [online] 30 Nov. Available at: <http://techie-buzz.com/foss/decentralized-project-bitcoin-darknet-yacy-internet.html?tb_sp=1> [Accessed 12 Aug 2012]

Koomey, J., 2011. My new study of data center electricity use in 2010. *Koomey.com* [blog], 31 July. Available at: <<http://www.koomey.com/post/8323374335>> [Accessed May 18 2012]

Kostakis, V., 2010. Identifying and understanding the problems of *Wikipedia*'s peer governance: The case of inclusionists versus deletionists. *First Monday* [online], 15(3). Available at: <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2613/2479>> [Accessed 25 June 2012]

Lessig, L., 2006. CODE: Version 2.0. New York: Basic Books

Levy, S., 2011. In the Plex: How Google Thinks, Works, and Shapes Our Lives. New York: Simon & Schuster.

Loban, B., 2004. Between rhizomes and trees: P2P information systems. *First Monday* [online], 9(10). Available at: <http://firstmonday.org/issues/issue9_10/loban/index.html>

Lobet-Maris, C., 2009. From Trust to Tracks: A Technology Assessment Perspective Revisited. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 73-85.

MacKinnon, R., 2012. Consent of the Networked: The Worldwide Struggle for Internet Freedom. New York: Basic Books.

MacKinnon, R., 2009. Original government document ordering "Green Dam" software installation. *rconversation* [blog] 8 June. Available at: <<http://rconversation.blogs.com/rconversation/2009/06/original-government...>> [Accessed 21 March 2012]

Manovich, L., 2009. How to Follow Global Digital Cultures: Cultural Analytics for Beginners. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 198-213.

Mayer, K., 2009. On the Sociometry of Search Engines: A Historical Review of Methods. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 54-73.

McCracken, H., 2011. Duck Duck Go, the In-N-Out Burger of Search Engines. TIME Techland [online]. (Last updated 16 June 2011) Available at: <<http://techland.time.com/2011/06/16/duck-duck-go-the-in-n-out-burger-of-search-engines/>>

Metahaven, 2009. Peripheral Forces: On the Relevance of Marginality in Networks. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 185-198.

Michaels, D., 2006. Rescuing Science from Politics. in: W. Wanger, R. Seinzor ed. 2006. Rescuing Science from Politics. Cambridge: Cambridge University Press, pp. 120-126.

Mumford, L., 1964. Authoritarian and Democratic Technics. *Journal of Technology and Culture*, 5(1).

Newport, S (editor), 2005. Largest Computation. In: Guinness World Records. Guinness.

Noveck, BS., 2005. A democracy of groups. *First Monday* [online], 10(11). Available at: <http://firstmonday.org/issues/issue10_11/noveck/index.html> [Accessed May 18 2012]

OpenNet Initiative, 2009a. China's Green Dam: The Implications of Government Control Encroaching on the Home PC. Toronto: OpenNet Initiative. Available at: <<http://opennet.net/china-green-dam-the-implications-government-control-encroaching-home-pc>> [Accessed 5 July 2012]

OpenNet Initiative, 2009b. China. Toronto: OpenNet Initiative. Available at: <<http://opennet.net/research/profiles/china-including-hong-kong>> [Accessed 6 July 2012]

Pasquinelli, M., 2009. Google's PageRank: Diagram of the Cognitive Capitalism and Rentier of the Common Intellect. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 152-163.

Ratzan, L., 2006. Mining the Deep Web: Search strategies that work: How to become an enlightened searcher. Computerworld [blog], 11 Dec. Available at: <http://www.computerworld.com/s/article/9005757/Mining_the_Deep_Web_Search_strategies_that_work> [Accessed 24 May 2012]

Redner, S., 1998. How Popular is Your Paper? An Emperical Study of the Citation Distribution. *The European Physical Journal B*, 4(2) pp. 131-134.

Rieder, R., 2009. Democratizing Search? From Critique to Society-oriented Design. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 133-152.

Rigi, J., 2012. Peer to peer production as the alternative to capitalism: A new communist horizon. *Journal of Peer Production* [online], 1 Productive Negation. Available at: <<http://peerproduction.net/issues/issue-1/invited-comments/a-new-communist-horizon/>>

Rogers, R., 2009. The Googlization Question: Towards the Inculpable Engine? In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 173-185.

Röhle, T., 2007. Desperately seeking the consumer: Personalized search engines and the commercial exploitation of user data. *First Monday*, 12(9). Available at: < http://firstmonday.org/issues/issue12_9/rohle/index.html> [Accessed 16 May 2012]

Röhle, T., 2009. Dissecting the Gatekeepers: Relational Perspectives on the Power of Search Engines. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 117-133.

Rudomilov, I., Jelenik, I., 2011. Semantic P2P Search engine. In: Proceedings of the Federated Conference on Computer Science and Information, Szczecin, Poland 18-21 Sept 2011. pp. 991-995.

Rushe, D., 2012. Google reports 'alarming' rise in censorship by governments. *The Guardian Technology* [online] 18 June. Available at: <<http://www.guardian.co.uk/technology/2012/jun/18/google-reports-alarming-rise-censorship>>

Ryan, M., 2011. Peer-to-Peer Search Engine Yacy Takes on Google. *LockerGnome* [blog], 30 Nov. Available at: <<http://www.lockergnome.com/news/2011/11/30/peer-to-peer-search-engine-yacy-takes-on-google/>> [Accessed 25 June 2012]

Seeks, 2011. Service. [online]. Available at: <<http://www.seeks-project.info/site/>> [Accessed 26 July 2012]

SETI@Home, 2012. About SETI@Home [online]. Available at: <http://setiathome.berkeley.edu/sah_about.php> [Accessed 28 July 2012]

Shaker, L., 2006. In Google we trust: Information integrity in the digital age. *First Monday* [online], 11(4). Available at: <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1320/1240>> [Accessed 25 June 2012]

SourceForge, 2010. Seeks delivers new search engine paradigm. SourceForge [blog], 24 June. Available at: <<http://sourceforge.net/blog/seeks-delivers-new-search-engine-paradigm/>> [Accessed July 26 2012]

Stalder, F., Mayer, C., 2009. The Second Index: Search Engines, Personalization and Surveillance. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 98-117.

Stark D., 2009. The Sense of Dissonance: Accounts of Worth in Economic Life. New Jersey: Princeton University Press.

Svensson, J., Bannister, F., 2004. Pirates, sharks and moral crusaders: Social control in peer-to-peer networks. *First Monday* [online], 9(6). Available at:

<<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1154/1074>> [Accessed 25 May 2012]

Van Hoboken, J., 2009. Search Engine Law and Freedom of Expression: A European Perspective. In: K. Becker, F. Stalder, ed. 2009. Deep Search: The Politics of Search beyond Google. Innsbruck: Studienverlag, pp. 85-98.

Varian. H., 2008. Why Data Matters. Google Official Blog [blog], 4 March. Available at: <<http://googleblog.blogspot.co.uk/2008/03/why-data-matters.html>> [Accessed 24 June 2012]

Vogelstein, 2007. As Google Challenges Viacom and Microsoft, its CEO Feels Lucky. *Wired* [online]. Available at: <http://www.wired.com/techbiz/people/news/2007/04/mag_schmidt_qa?currentPage=all> [Accessed 8 April 2012]

Vascellaro, J., 2008. YouTube suit sets protocol to shield IDs. *Wall Street Journal*, 21 July.

Watts, D., 2003. 6 Degrees of Separation: The Science of a Connected Age. London: W.W. Norton & Company.

Weinberger, B., 2012. What is Google?. TechCrunch [blog], 15 July. Available at: <<http://techcrunch.com/2012/07/15/what-google-is/>> [Accessed August 3 2012]

WSIS (World Summit on the Information Society), 2003. Charter of Civil Rights for a Sustainable Knowledge Society, Version 3.0. Geneva, Switzerland 10-12 Dec 2003.

Yacy, 2012a. Technology. [online]. Available at: <<http://yacy.net/en/Technology.html>> [Accessed July 25 2012]

Yacy, 2012b. About. [online]. Available at: <<http://yacy.net/en/index.html>> [Accessed July 25 2012]

Yacy, 2012c. Philosophy. [online]. Available at: <<http://yacy.net/en/Philosophy.html>> [Accessed July 25 2012]

Yahoo!, 2012. The History of Yahoo! - How it all started. Yahoo! media relations [online]. Available at: <<http://docs.yahoo.com/info/misc/history.html>> [Accessed June 18 2012]

Zuckerman, H., Merton, R., 1971. Patterns of Evaluation in Science: Institutionalization, Structure and Functions of the Reference System. *Minerva* 9, no 1, pp.68-69

THIS IS THE ORIGINAL DRAFT SUBMISSION MADE TO THE JOURNAL. IT IS NOT THE DEFINITIVE VERSION OF THE ARTICLE AND IS BEING RELEASED AS BACKGROUND ONLY.