ETHICAL AND SOCIAL ASPECTS OF INTERNET SEARCH IN HISTORICAL PERSPECTIVE

Miroslav Vacura

Department of Philosophy / Department of Information and Knowledge Engineering
Faculty of Economics / Faculty of Informatics and Statistics
University of Economics, Prague
vacuram@vse.cz

Keywords

Information ethics, internet, web search

Abstract

At the dawn of the 21st century everyday life of the common citizen of industrialized society has become more and more intertwined with Internet technologies and services. One of the prominent Internet services is the web search and the first part of the paper describes its history, current state and emergence of dominant search websites. The second part focuses on ethical and social implications of the described process. The paper discusses social bias of web search, problem of visibility on the Internet and relevance to concept of political voice. Several proposed solutions are analyzed in the final sections of paper.

1. Introduction

Since the beginning of the 21st century everyday life of the common citizen of industrialized society has become more and more interconnected with Internet technologies and services. The Internet is slowly making its way into people's private lives. Email was followed by instant messaging applications like ICQ or MSN, and most recently by Internet telephony either SIP based or Skype. The most frequently used and probably the most important are still full text web search services. Along with the new technologies also new companies have risen from the ashes of the dot com bubble at the beginning of the century.

The process of commercialization of the Internet has led to retreat from its original ideas. The original vision of the Internet imagined it as a dense distributed network with no dominant websites; Internet standards and protocols were designed with this vision in mind (Abbate, 1998; Berners-Lee, Fischetti, 1999). Many people believed that new sources of online information would inform citizens more about politics and would help to involve previously inactive citizens into political participation. Early visionaries believed that the Internet would become a robust forum for political debates and that the openness of the Internet would allow ordinary citizens to publish their opinions along with professional journalists (Hindman, 2008:1).

The rest of the paper is structured as follows: Section 2 presents short historical overview of development of web search services, Section 3 analyzes relevant social and ethical issues, and Section 4 finishes with conclusions and acknowledgements.
2. Advent of search engines

In the beginning of the Internet era, a typical user spent most of his time “surfing” Internet websites which meant going from one web to another, and then to another based on the links present on previous webpages or in some cases manually maintained directories of websites usually focused on a specific theme. With the increasing amount of web pages more ambitious projects emerged – web global catalogues trying to include every important internet web site for any topic. These catalogues were created and maintained semi-automatically and were organized by a topic based hierarchy. Most of these global web directories have not survived until today, but one of them became very successful – Yahoo.

As the amount of Internet websites grew, manually and semi-automatically maintained catalogues were not flexible enough to provide complete and reliable reference of the web content. The number of fully automatic search engines was developed with a similar central idea – crawler software browsing through web space and collecting information about webpages into a central database paired with user interface and search software performing retrieval functions on the database content. Again, there were a handful of such web services, apart from experimental and academic software, one of the first was AliWeb in 1993, then WebCrawler, Infoseek, Lycos in 1994, Magellan, Excite, AltaVista in 1995, Inktomi, Northern Light, SavvySearch, Infind, and many others later. In 1998 the Google search web site was launched. In following years it has slowly become a dominant Internet search engine. Yahoo introduced Inktomi powered full text search in 2001, later migrated to Google technology and then implemented its own search engine in 2003-4. Since 2009 Yahoo uses Microsoft Bing technology.

Along with Yahoo and Google, the third very popular web site that includes web search service is Microsoft's MSN/Windows Live in 2009 renamed to Bing. Several MSN/Windows Live specific tools were developed and distributed as part of the Windows operating system and provided to users as an automatic update of the operating system or published as a free download add-on.

<table>
<thead>
<tr>
<th>Search engine</th>
<th>Share 2007</th>
<th>Share 2010</th>
<th>Share 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>53.6%</td>
<td>66.6%</td>
<td>66.7%</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>19.9%</td>
<td>16.0%</td>
<td>12.2%</td>
</tr>
<tr>
<td>MSN/Live/Bing</td>
<td>12.9%</td>
<td>12.0%</td>
<td>16.3%</td>
</tr>
</tbody>
</table>

Table 1

During the last few years the following three internet websites emerged: Google, Yahoo, Bing – as centers of the Internet traffic and dominant web services providers on the Internet. In the Table 1 there are their websites' shares of the full text Internet search service in 2007, 2010 and 2012 (comScore, 2008a, 2011, 2013). Google is leading by a significant margin since 2007 – the Internet search is Google's primary area of expertise (Nielsen, 2007).

Each of these websites had a different background and a different primary area of expertise. In Yahoo's case it was a large semi-automatically maintained hierarchical catalogue of websites, in case of Google it was an efficient and fast full text Internet search web service, and in case of Microsoft's MSN/Windows Live/Bing it was its strong position in desktop software and specifically its advantage as a developer of the most popular operating system. Over time these websites developed into a comprehensive suite of various web services. Such additional services typically include email, news, weather forecast, TV and cultural events program, photography sharing, discussion forums and many others. There are of course many specifics. Microsoft at first
developed the MSN/Windows Live in its core as a social networking website (based on the concept of social networking websites like Myspace or Facebook), enabling users to easily share information and communicate with their friends. Google tried to attract users by ability to customize their search homepage, providing online office suite applications and many other small but useful gadgets like its geographical map web application, online library or even house interior design application. This ability to develop and integrate this wide range of Internet applications leads some commentators to claim Google as a standard-bearer of Web 2.0 (O'Reilly, 2005). All these companies – Microsoft (as part of operating system), Google and Yahoo – provided instant messenger applications: Live Messenger, Google Talk and Yahoo Messenger. Since 2006 Live Messenger and Yahoo Messenger are compatible and users of these two instant messaging networks can communicate with each other. In 2013 Microsoft Live Messenger was discontinued and replaced by Skype internet phone and messaging service, acquired by Microsoft in 2011 (See Microsoft, 2011).

Such an approach aims at developing a complete framework of Internet applications that suits all needs of an average Internet user. The purpose of such web application suites is to keep users inside or to *internalize the Internet traffic*. A very small number of external links is offered and the user is encouraged to stay within the limits of the internal web and use only services that are provided on this website.

The advent of Web 2.0 is characterized mainly by social networking services and web applications. As we have already described Microsoft made social networking applications the core of its MSN/Windows Live website. It may seem that social networking sites are competitors for sites like Google and dominant Internet websites are probably well aware of the importance of social networking in the future of the Internet. In 2006 Google acquired YouTube (social networking website for video streaming) and recently Google developed OpenSocial interface aimed to be common ground for developing applications for social networking applications. Although only a couple of not so very well-known social networking sites joined this initiative at launch – Orkut, Salesforce, LinkedIn, Ning, Hi5, Plaxo, Friendster, Viadeo, later also MySpace, Bebo and SixApart announced their participation. It may be also noted that MySpace was acquired in 2005 by Fox Interactive Media. The Facebook social networking site founded in 2004 remains the only large social networking website that is still independent. In 2011 Google launched its own social networking service Google+.

In the light of these events another interesting questions may be considered: What websites do Internet users really regularly visit? Already in November 2008 Internet statistics captured totally 190 m. unique users in the USA. Google website was visited by 146 m. unique users, Yahoo website was visited by 143 m. users and websites operated by Microsoft were visited by 123 m. unique users (comScore, 2008b). We may conclude that the majority of users visit three dominant sites regularly or more typically – uses them as their primary point of departure even when visiting other websites. E.g. Google and two other dominant websites are the most important sources of visiting users for music related websites in the UK (Hopkins, 2006). Many users go to these specialized websites after searching for a keyword for example in Google and choosing from the presented list of results. In the light of presented statistics it may be also interesting to mention that there exist ongoing efforts of Microsoft to acquire Yahoo (Isidore, Lev-Ram, 2008).

### 3. Social and ethical relevance of Internet search

The reason why we describe this development is to illustrate the continuous process of centralization of the Internet services. Such process implies questions relating to many different
scientific areas – it may be interesting, for example from the point of view of economics, to ask whether the essential characteristics of the Internet itself necessarily result in forming of some kind of natural monopoly. Becoming a central site of Internet traffic is enormously expensive. Hindman (2008, p.84) points out that Google pays out billions of dollars annually to have other websites which send visitors to his web services. Similarly, costs of computer equipment are extremely high – already during the years 2003–2005 Google spent $1.33 billion on property and computer equipment. The total number of servers operated by Google has been in year 2009 estimated between 450 thousands and 1 million. Google does not make this information public, however Miller (2009) revealed, that its single container data center often holds more than 45 thousands servers and according to Shankland (2008) already in year 2008 Google had more than 35 such large data centers across the globe. Due to inability of standard database software to deal with amounts of data Google has to handle, a new database system called BigTable was internally developed and is used to manage as much as 6 petabytes of data across thousands of servers (Lai, 2009). Such circumstances make it extremely difficult and expensive to seriously compete with Google.

If we focus on ethical and social implications of the described process, the first problem we encounter concerns the description of the situation itself. It has been already pointed out by Moore (1985), one of founders of computer ethics, that there are often conceptual muddles that need to be sorted out. Also Johnson (2004, p.68) asks “How are we to conceptualize a search engine?” Technical development we have described in previous paragraphs results in technology and information artefacts that have many unique properties unlike anything else in human history. Johnson therefore believes that when we are dealing with issues like these, it is not the case of “simple” applied ethics, because it involves a complex conceptual analysis and interpretation of completely new phenomena not just applying existing ethical theory to a new situation. World Wide Web inventor Barners-Lee suggests that the complexity of the web has grown to the level of complexity of the human brain – there are $10^{11}$ webpages and there is a similar number of neurons in the brain. He says that now we do not fully understand the nature of the emergent systems that have cropped up on it (Marks, 2009).

The fact that search engines raise not merely technical issues but also political ones was recognized already by Introna and Nussbaum (2000, p.17). They focus on the ranking of websites in search results and explain the nature of the problem in what we can call sociologically and technologically based bias. The technological, software design of web search engines implicates preference of specific websites, “popular, large sites, whose designers have enough technical savvy to succeed in the ranking game”. There is also a socially or economically based preference of sites “whose proprietors are able to pay for various means of improving their sites’ position”.

The social bias can be also connected even with the ethnic or racial background and current demographic patterns of the Internet access and usage (Hoffman et al., 1997, Hoffman, Novak, 1998). The important outcome is that information relevant to some ethnic group, which is not numerous or for whatever reason does not use the web as intensively as others may be ranked lower in search results sets than information on more popular websites. This constitutes what we may call visibility on the Internet. If dominant web search pages are major sources of incoming users for many webpages then the visibility of a webpage is determined by the position which it has in the result sets returned for some typical queried key words. This search engines visibility is closely related to what is in political science called political voice and is one of its central concepts. It has been pointed out that clear, loud, and equal voice of citizens in politics is a requirement for meaningful democratic participation (Verba et al., 1995, p. 509; Hindman, 2008, p. 6).
Let us imagine a situation preceding presidential elections (let us say in a country like the USA). Most responsible voters try to find out relevant information about their candidates. There is a lot of relevant information on television and in newspapers but some people prefer to find such information on the Internet, and the importance of online information will probably even rise in future. Now what happens if a dominant search engine deliberately presents at the top of its result set webpages idealizing one of the candidates and pages containing mostly criticism and denouncement of other candidates? Such manipulation can be done in a way that is not easily recognizable. Does it have measurable effects on the results of elections? While there is a considerable amount of work trying to analyze the impact of new media on politics and democracy (see e.g. Abramson et al., 1990) the specific role of internet search engines has not yet been sufficiently analyzed. We already know that the link structure of the Internet, the element that is the most important for most of search engines is not itself politically neutral and Roger (2004, p. vii) has shown that it can be analyzed in terms of what he calls “politics of association”.

The exact working of a search engine and its algorithm are considered an industrial secret. Engineers and owners of the search providing company are free to modify it in any way they want. A question then may be: What is the legal status of a search service? What is the relation between a user and a service provider? Something like that is usually stated in a “terms of service” document however for example in the case of Google there is not anything mentioned regarding the characteristics of the search results. Is there any obligation (legal or moral) of a search service provider related to the set of results he presents to a user? Google describes its determination to provide correct results in one of its basic documents: “[our search results] …are unbiased and objective, and we do not accept payment for them or for inclusion or more frequent updating” (Page, Brin, 2004). However exact legal (and moral) status of such a statement is unclear.

One way of overcoming such possibilities of abuse is to use the open source approach to software development or at least a partial open source approach – like in case of Microsoft which made the source code of its Windows OS available to selected public institutions and government authorities (Microsoft, 2010). But such an approach seems to have a number of drawbacks in the case of search websites. Introna and Nussbaum (2000, p. 16) note that web search companies are loath to give out details of their webpage ranking algorithms for fear that abusers and spammers will use this knowledge to trick them. There are ongoing efforts of many individuals and companies to guess details of ranking algorithms, some even with scientific backing (Pringle et al., 1998).

Other authors suggest that some kind of regulation should take place on the Internet. However, such suggestions are made only in very general terms, without any specific regard to web search. Livingstone and Lunt (2007) say: “Access to, and the content of, the press, television, Internet, and so on should be evaluated, therefore, not in terms of what contents or services they provide but in terms of the possibilities they afford or impede.”

Anderson (1993, p. 141) similarly claims that there is a category of goods that should not be left entirely (if at all) to the marketplace because there are inherent ethical limitations of the market norms (See also Introna, Nussbaum, 2000, p. 23). There are goods for which this claim is uncontroversial, such as: person, body, friendship, political rights like the right to vote, but she controversially believes that the same applies to a much wider range of goods such as public spaces artistic endeavour, addictive drugs and reproductive capacities (See also Fabre, 2006). Introna and Nussbaum (2000) believe that also Internet search services belong to this specific category of goods and say that while for goods like cars or bottled salad dressing etc. the marketplace is a perfectly adequate distribution mechanism, for other goods this distribution fails to properly express values of the liberal democratic society committed to freedom, autonomy and welfare. Introna and Nussbaum therefore agree with Anderson (1993) in her substantial claim that goods belonging to
the category of political goods have to be distributed in accordance with public principles and not just by the market mechanism. The reason for such a conclusion is the belief that while retaining a full range of options in bottled salad dressings or cars has no impact on the political sphere, retaining visibility of a full range of political options expressed on the Web has key importance in maintaining the pluralistic democratic society (Anderson, 1993, p. 159; Introna, Nussbaum, 2000, p. 23). The argument may be reduced to this: while we may live in a perfectly democratic society with only one variety of salad dressing available, the democratic character of society would be endangered if there would be only one kind of a political opinion offered by search results of internet search services.

As a supportive argument Introna and Nussbaum (2000, p. 25) claim that the special character of search services is derived from special character of Web itself. The Web is a public good and it earns this character in many of the same ways as other public goods. The meaning of the term “public” itself signifies something that is not privately owned and the Web seems to be public at least in this sense. While its constituent parts – hardware and software, could be privately owned the Web as a whole is not privately owned by any particular entity. Similarly, it does not come under jurisdiction of any single sovereign state; therefore its character invokes a number of difficult legal and legislative dilemmas (See also Johnson, Post, 1996).

Gaus (2009, p. 5) says that many characteristics of the Web are similar to what is usually called “common pool resources” like fresh air or water – resources that are characterized by relatively open (public) access and private consumption. We can encounter similar classes of problems – just like pollution is a prime example of a common pool problem related to fresh air (over-use of the air’s ability to dissipate waste gasses leads to the depletion of that ability), the email spam is an example of a common pool problem related to the Internet (abuse and over-use of the email ability to efficiently and cheaply deliver messages leads to the depletion of that ability).

Another important point is contribution of availability of information to market effectiveness. To function properly and to maximize efficiency the free market presupposed that parties involved in market exchange have information about what they are exchanging. Economic theories of free market generally assume that both parties to an exchange are equally informed. Recent research focused on how asymmetric information can affect market transactions – if one party does not have access to full information regarding the subject of transaction, we can no longer suppose that market exchanges are truly mutually beneficial and maximizing efficiency (Gaus, 2009, p. 11; Sandler, 2001). The Web then may be seen as part of “market infrastructure” that ensures that everyone does have equal access to information and therefore ensures free market efficiency. The search service obviously plays an extremely important role with regard to this function of the Web; therefore as the necessary condition of the efficient function of the market it may not be seen just as one of many marketplace subjects.

While an asymmetric information problem is relatively uncontroversial the similar problem of asymmetric bargaining power lies at root of many current economic and political controversies. Many government regulations like labour laws regulating hours and factory conditions are justified by the claim that employers and workers have asymmetric bargaining power. While some inequality in bargaining power does not harm effectiveness of the market and the mutual benefit from the exchange there are others that seem to have such an effect. Gaus (2009, p. 12) describing such an economic situation cites Nozick (1974, p. 180), who argues: “a person may not appropriate the only water hole in a desert and charge what he will. Nor may he charge what he will if he possesses one, and unfortunately it happens that all the water holes in the desert dry up but his.” If now the hypothetical appropriator of a single source of water makes an offer of a glass of water for all your property this would be what is in economic theory called “coercive offer” – an offer that
exploits one’s bargaining power and cannot be refused. Such a situation on market results in a sort of exploitation of those in need and not in mutual advantage.

If we now extend this line of argumentation to finalize this article with question – what if there would be only a single comprehensive search service on the Web, would not we be in a situation similar to the one described above with a single source of water in the desert?

4. Conclusion

This paper provided short historical overview of development of web search services and described process of their centralization and of internalization of Internet traffic. On this foundation then the most relevant social and ethical issues implied by this development were analyzed: the conceptualization of the search engine, the bias of search related to Internet visibility and problem of political voice. Also some proposed solutions were shortly investigated like open source approach and regulation.

Acknowledgements

This work was supported by the European Commission under Contract FP7-287911 LinkedTV.

References

Miroslav Vacura


