

International Telecommunications Society  
ITS 19th European Regional Conference  
2-5 September 2007, Istanbul, Turkey

## ***Search Engines, the new bottleneck for content access***

*Prof. Dr. Nico van Eijk<sup>1</sup>*  
*Institute for Information Law (IViR, University of Amsterdam)*

*Version 2.0*

*Contact:*  
*e-mail: [vaneijk@ivir.nl](mailto:vaneijk@ivir.nl)*  
*tel.: +31 20 525 3931*

## **Abstract**

*The core function of a search engine is to make content and sources of information easily accessible (although the search results themselves may actually include parts of the underlying information). In an environment with unlimited amounts of information available on open platforms such as the internet, the availability or accessibility of content is no longer a major issue. The real question is how to find the information. Search engines are becoming the most important gateway used to find content: research shows that the average user considers them to be the most important intermediary in their search for content. They also believe that search engines are reliable. The high social impact of search engines is now evident.*

*The paper discusses the functionality of search engines and their underlying business model – which is changing to include the aggregation of content as well as access to it, hence making search engines a new player on the content market. The biased structure of and manipulation by search engines is also explored. The regulatory environment is assessed – at present, search engines largely fall outside the scope of (tele)communications regulation – and possible remedies are proposed.*



# Index

Abstract	2
Index	4
1. Search engines: we cannot do without them	6
2. How a search engine works	6
3. The search-engine market	8
4. Where does the money come from?	9
5. Manipulation of search results	11
5.1 Search engines	11
5.2 Information providers	12
6. Data retention and content aggregation	13
6.1 Data retention	14
6.2 Content aggregation	15
6.3 Other search engine involvement	15
7. Regulatory aspects	15
7.1 Freedom of expression	16
7.2 Competition law	16
7.3 Sector-specific regulation	18
7.4 Privacy	19
8. Analysis	20
Literature	22
Notes	24



## 1. Search engines: we cannot do without them

Search engines have become an essential part of the way in which access to digital information is made easier. They are used by virtually all internet users (in February 2007, US internet users conducted 6.9 billion searches), who moreover believe that searching through search engines are reliable and the best way of finding websites.<sup>2</sup> “Googling” has become an autonomous concept and an independent form of leisure activity, similar to zapping through television channels. Anybody who cannot be found via a search engine does not exist: “To exist is to be indexed by a search engine.”<sup>3</sup> Because of its prominent position, Google is often used as an example in the following paragraphs.

Number of searches in the United States		
Search Engine	01/2006	02/2007
Google	2.3 billion	3.3 billion
Yahoo	1.6 billion	2 billion
MSN	752.5 million	730 million
Others	827.5 million	870 million
Total	5.48 billion	6.9 billion

*Source: comScore Networks*

## 2. How a search engine works

The main function of a search engine is that of enabling access; it is a gateway to possibly relevant information on the internet. However, it is a two-directional gateway: from the information provider to the user and from the user to the information provider. A search engine determines which information provided by an information provider can be found by the end-user as well as what information the end-user will ultimately find. The search facility provided and the underlying search algorithm thus control supply and demand. Or to put it more simply: it is a bottle-neck with two bottles attached to it. How does a search engine work? Most search engines use more or less the same method to achieve search results.<sup>4</sup>

The process starts with searching the internet for information. This automated process uses intelligent “sleuths” called spiders, bots or crawlers. These sleuths surf the internet using criteria set previously by the search-engine provider.

The information found is thus made uniform and structured, laying the basis for its traceability. Then the information is indexed. This indexing determines the criteria for what are considered relevant words or combinations of words. Irrelevant information, such as fillers and punctuation marks, is deleted. At this stage the information is also streamlined in such a way that, for example, differences between singular and plural forms of words or variations due to declensions produce identical search results. Certain recognisable words, such as people’s names and basic concepts, are possibly identified. The rest of the information is then “weighted”, based on the frequency of words in a text and the contextual relevance or significance (or otherwise). This enriched information forms the ultimate basic material for the search engine.

When a search engine is consulted, a process is used that is largely the opposite of the indexing process. The end-user formulates a search question that is broken down and analysed by the search engine. In this process, non-relevant elements (such as fillers) are deleted, the relationships between the search terms are looked at (this can be indicated in the search query i.e. by using Boolean operators, e.g. AND, OR, NOT), and the relative importance of the search terms entered is charted. This leads to several search results, which are displayed on the end user’s screen.

The search engine process						
Searching the internet	Structuring collected data	Indexing data	Search query	Analysis query	Linking with index	Search result

It is by no means true that all information that is present on the internet is found and indexed by search engines. In literature, there are claims that individual search engines index only 16% of all the information present on the internet, and all the search engines together cover no more than 42% of all available information.<sup>5</sup> Other estimations contradict these low numbers, but this does not detract from the observation that only a limited amount of the information present is, or can be, indexed.

There are various reasons for this. Some of the information is hidden in files that cannot be indexed, such as text in graphics files. However, search engines are becoming increasingly intelligent and are increasingly capable of analysing more and more formats of information (e.g. Word, PDF and JPG-files). There is also information that the providers do not want to have included in search engines. News information that is rapidly refreshed, for example, is not suitable for inclusion in search engines, as the information quickly becomes obsolete (sometimes months pass before a spider attempts to re-index the site). There is also information that is accessible via the internet but that is not itself present on the internet, such as information stored in external databases. Moreover, the internet is still constantly growing and changing.

The model of collecting and ordering information and making information available is only one reflection of reality. What actually happens before a search result is made available is very complex and is characterised in an important way by the many subjective elements woven into the process (also see paragraph 5).

### **3. The search-engine market**

Not so long ago, at the beginning of the century, a lot of search engines were active, and it was the general assumption that competition between search engines would discipline the market. Both information providers and users would be able to benefit from this. Although the number of search engines is still significant, this cannot be said about their market shares.

Recent statistics on the US market show that Google, Yahoo, MSN/Livesearch and ASK together have a market share of 92%. All the other search engines account for the remaining 8% of the market. Google is clearly the market leader.



<b>Percentage of US searches among leading search engine providers</b>			
Domain	Mar 2007	Feb 2007	Mar 2006
www.google.com	64.13%	63.90%	58.33%
search.yahoo.com	21.26%	21.47%	22.30%
search.msn.com	9.15%*	9.30%*	13.09%
www.ask.com	3.48%	3.52%	3.99%

\*includes executed searches on Live.com and MSN Search

Source: Hitwise

There is an interesting difference between the US and Europe. Although an American company, Google is even more dominant in Europe. Recent figures about the Dutch market speak for themselves. Google has reached a 96% market share, whereas the second player, Ilse (Dutch), has a share of only 2%. The Dutch figures are extraordinary, but Google dominates in many European countries with a market share above 80%.

<b>Market share of search engines in the Netherlands</b>											
Search engine	02/02	05/02	01/03	08/03	02/04	10/04	01/05	04/05	01/06	10/06	02/07
Google	32	40	52	65	68	74	84	85	91	90	94
Ilse	19	16	14	17	19	14	9	8	5	4	2
Livesearch	4	3	5	6	4	3	2	3	2	1	1
Yahoo	3	2	1	1	1	4	1	1	0	0	0
Lycos	2	2	2	1	1	0	0	0	0	0	0

#### **4. Where does the money come from?**

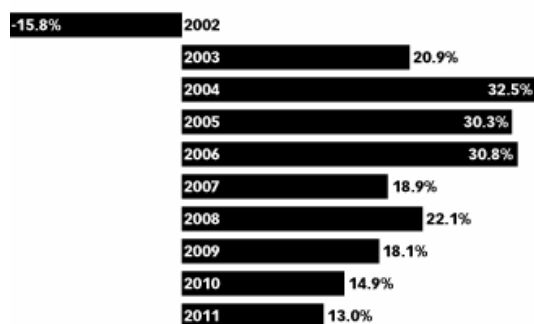
Search engines generate income mainly from one source: advertising. Again, we take Google as an example. Google generates almost all of its income from advertising. This income is generated mainly by “Google AdWords”. AdWords enables advertisers to create their own advertisements and state how much money they are willing to spend. They are then charged on the basis of the number of times that the advertisement is clicked on. The advertisements appear on the Google web site next to the results of a search request. Google decides which advertisement appears when and does this mainly in relation to the search request.

The second source of income consists of placing the advertisements on third parties’ websites. This is done via the AdSense program, which has two variations: “AdSense

for search” and “AdSense for content”. With “AdSense for search”, advertisements are placed in relation to search requests on third parties’ web sites. With “AdSense for content”, advertisements are linked to the content of websites. For AdSense, Google has a revenue-sharing model, with some of the advertising income generated going to the information providers. These providers are thus in a position to take this into account when putting together the content of their website and to “optimise” the content.

Just to illustrate the financial impact: according to industry data for 2005, the four largest search engines/portals had captured more than half that year’s US internet ad spending of \$12.5 billion. In 2007, projections suggest that two-thirds of the \$19.5 billion spent online will go to Google, Yahoo, AOL and MSN. Google alone reported a total advertising income for 2006 of almost \$10.5 billion.

**US Online Advertising Spending Growth, 2002-2011 (% increase/decrease vs. prior year)**



Note: eMarketer benchmarks its US online advertising spending projections against the Interactive Advertising Bureau (IAB)/PricewaterhouseCoopers (PwC) data, for which the last full year measured was 2005; online ad data includes categories as defined by the IAB/PwC benchmark—display ads (such as banners), paid search ads (including contextual text links), rich media (including video), classified ads (including sponsorships, referrals/lead generation, e-mail (embedded ads only) and slotting fees  
Source: eMarketer, February 2007

080791

www.eMarketer.com

**US Online Advertising Revenues\* at Top Four Portals As a Percent of Total Online Advertising Spending, 2004-2007**

	2004	2005	2006	2007
Google	13.1%	19.2%	25.0%	32.1%
Yahoo!	18.4%	19.4%	18.3%	18.7%
AOL	6.8%	7.2%	7.5%	9.1%
MSN	9.4%	7.8%	6.7%	6.8%
<b>Total for top four portals</b>	<b>47.8%</b>	<b>53.7%</b>	<b>57.4%</b>	<b>66.6%</b>

Note: Numbers may not add up to total due to rounding; \*net of traffic acquisition costs (TAC)  
Source: company reports, 2004-2007; eMarketer calculations, February 2007

081313

www.eMarketer.com

Syndicating ad space (related to search results and other available data) is now being extended to become a more general mechanism to allocate advertising slots in other media like radio, TV and print. Again, Google is an active market player in this respect. It offers the possibility to use the Adwords mechanism to sell airtime to radio advertisers (“AdioAds”). Already 1600 radio stations – including the 675 Clear Channel stations – use the service. More recently, Google announced the acquisition of DoubleClick, one of the leading companies in digital marketing. The announcement caused quite some reactions about the possible negative effects on the market and with regards to privacy. It is a sign on the wall that companies like Microsoft and AT&T

where amongst those who expressed their concerns. This horizontal extension of its market should generate further advertising-related income and contribute to the diversification of revenue resources. The transaction is still under review by the (US and EU) competition authorities.

Certain search engines (ie. Yahoo) offer the possibility to influence search results and/or ranking positions. This is not a dominant activity, but remains often unclear for the user.<sup>6</sup>

## **5. Manipulation of search results**

The manipulation of search results takes at least two forms: manipulation by the search engine and manipulation by information providers by boosting their ranking in the search results.

### **5.1 Search engines**

The first form of manipulation is carried out by search-engine providers. They draw up the criteria on the basis of which the information present on the internet is collected, ordered and made available. Information that is not searched for is not found. If a spider is instructed to ignore certain information, this information will never appear as the result of a search action. The analysis of a search query and the answer to be given are determined by the algorithm that the search engine uses. This algorithm is the true secret to the way the process works, and it is the ultimate manipulation tool. It resembles to some extent the secret recipe for Coca-Cola.

Here are a few examples from practice to illustrate the manipulation by search engines: Some search engines offer the opportunity of “buying” a high position on the list of search results. There are different variations of this. The simplest method involves literally selling the position. Other search engines priority-index the pages of paying parties, so that they rank higher in the list of search results.

For commercial or policy reasons, some search engines - using filters - deliberately do not reproduce any certain results. For example, it is claimed that Google does not make certain search results available in the case of search queries from specific countries.<sup>7</sup> Furthermore, search engines can be under legal obligations not to provide certain search results. Criteria for exclusion can originate from legislation or based on jurisprudence. For example, in Germany and France restrictions exist on the portrayal/promotion of Nazi-related material (the famous Yahoo case). Courts regularly interfere based on trademark, copyright or unfair business practices regulation. Research shows that the results of search requests differ, not only depending on the search engine used, but also depending on whether Google.com, Google.de or Google.fr is used.<sup>8</sup>

There are search engines that, in addition to automated systems, also use a human factor: search results are manually adjusted by their own employees on the basis of more detailed criteria that have been formulated, both subjectively and otherwise.

Finally, the relationship between search and advertising income has already been mentioned in paragraph four. The need to optimize revenues causes search engines to take this relationship into account.

## **5.2 Information providers**

The second form of manipulation is manipulation by information providers. They can do this by paying for a higher ranking in some cases or by exercising direct influence on the search-engine provider, but more often it is a matter of cleverly designing the information provider's own web information to create a profile in such a way that the information is placed high up on the list of search results by the search engines. In doing this, they attempt to anticipate the search engine's algorithm (to the extent that this is actually known). A classic example is the manipulation of one's own metatags by adding attractive search words that have nothing to do with one's own service provision (such as football, pornography or the brand names of competitors).

However, search engines are becoming increasingly clever and are often capable of "neutralising" the effects of manipulated metatagging. More advanced methods are

therefore currently used to attract greater attention. Fake sites are being set up, for example, that contain a lot of references to one's own site in order to influence page-ranking systems. Popular sites are being copied and included invisibly in one's own site so that unsuspecting users end up at other sites than those they intended to access. These and other forms of manipulation or deception are known as spamdexing, cloaking, link farming, webring, redirects, doorway pages, page-jacking, etc. All such methods aim to improve the ranking in the search results.

These manipulation techniques are combated by the search engines but not always successfully. At Google, the ultimate sanction is the exclusion of the offender, whose pages are then no longer indexed. The party concerned can then no longer be found via the search engine. The offenders are not just shady characters: they include governments and reputable companies, which use agencies to optimise the search results. An entire industry has emerged around this optimisation of search results. Under the name "search engine marketing" companies offer services aimed at improving rankings. They are also called SEOs, "search engine optimisers", a nice euphemism. Search engines in general do have policies on optimisation and "allow" certain types of manipulation by information providers.

## **6. Data retention and content aggregation**

The functionality of search engines is to a large extent determined by the nature and extent of the underlying data. The systems not only gather information about the data available on the internet, they also link that to what they know about the people submitting search queries. This means that the query itself plays an additional but crucial role.

As well as that processing and analysis itself, this paragraph also looks at the fact that, in certain cases, search engines are developing a vertical relationship in respect of the content they are processing and analysing.

## **6.1 Data retention**

In the first instance, a search engine is dependent upon data generated by third parties. That is the information available on the internet, in the form of websites and the associated data like metatags. The engines interpret that information, which results in them recording a large amount of selected data. That is then saved so that, amongst other things, a more accurate interpretation can be provided and hence a better search result generated. This process is described in section 2 above.

As well as gathering information from the internet, user data is also generated. This consists of data made available by users themselves. That may come from submitted forms specifying personal preferences, but can also be derived from user-authorized analysis of personal documents such as e-mails (as is the case with Gmail, Google's e-mail service) or the use of online or offline applications like Google Desktop, Picasa and Google Docs & Spreadsheets.<sup>9</sup>

Thirdly, there is the data generated by the search queries themselves. In principle, these provide information about both the user – such as personal preferences, possibly combined with personal data – and what they are looking for.

If all the data mentioned are recorded, it creates a vast database. The size of that is determined by such factors as:

- a) when data recording began;
- b) what data is selected;
- c) how long the data is retained;
- d) how and when data is re-evaluated; and
- e) when aggregated data is deleted.

Although the phenomenon as such is not unfamiliar – data warehousing and data mining are well-known terms, after all – relatively little is known about the data recorded by search engines. They are very coy about this aspect of their activities. We shall return to the sensitivities associated with data retention when discussing the regulatory aspects of the issue.

## **6.2 Content aggregation**

Several search engines are seeking vertical integration. This trend is reflected in their efforts to own, acquire or otherwise control content or its associated exploitation rights. In this respect, Google is a striking example. It is building a database of world literature, Google Books, by digitising the contents of libraries. Out-of-copyright works are being made available online in their entirety; in the case of books still subject to copyright protection, only an excerpt known as a “snippet” can be viewed. Another case in point is the company’s acquisition of YouTube, the website on which companies and individuals can post videos for viewing by other internet users. And a third example is Google’s activities in the field of mapping and geographical information.

As well as acquiring content directly in this way, search engines are also entering into special or preferential relationships with information providers. These can be based either upon the “manipulation” model described earlier – privileging certain providers in return for payment – or upon some form of revenue sharing (see section 5).

## **6.3 Other search engine involvement**

Without further discussing it, search engines have activities in many other areas inside and outside the vertical value chain. For example, search engines actively participate in the debate about network neutrality. They clearly seek control over the underlying (tele)communications infrastructure as was recently illustrated again by the interest of Google for acquiring frequencies.

# **7. Regulatory aspects**

With the growing role of search engines, the question increasingly arises as to where to position them in law.<sup>10</sup> The myth of the self-regulating internet, the idea that it is “different”, seems to have been exploded. The next-generation internet, the much-hyped “Web 2.0” will definitely bridge the gap between the “old” and the “new” worlds as far as its regulatory aspects are concerned. It might be somewhat controversial to put it this way, but the internet is becoming embedded into the day-to-day business of regulation.

This is a sign of the internet's maturity and of its growing social and economic importance.<sup>11</sup>

Nevertheless, search engines are still largely “lost in law”. The applicability of existing legal concepts needs further testing, while sector-specific rules such as European media regulation or the European regulatory framework for the communications sector have not been written with the phenomenon of the search engine in mind.

A myriad of topics could be discussed under the heading “regulatory aspects”. Within the framework of this paper, however, only a limited number of aspects will be looked into – with an emphasis on the European regulatory perspective.<sup>12</sup> First of all, the question can be raised as to whether or not generic regulation might be or become relevant. We will look briefly at two aspects of this: freedom of expression and competition. Secondly, does sector-specific regulation come into play? And more particularly, do existing regulatory frameworks such as the European directives on audiovisual media services, the communications sector or privacy apply to search engines.

### **7.1 Freedom of expression**

Given their role in the information society, it goes without saying that freedom of expression is at the heart of the regulatory context pertaining to search engines. But, in particular as laid down in Article 10 of the European Convention on Human Rights (and Article 11 of the EU Charter of Fundamental Rights), that freedom does not directly cover the core activity of search engines. This has to do with the fact that Article 10 deals with the freedom to hold opinions and to receive and impart information. Search engines are primarily making information which is already available accessible. None the less, in my view this *making information accessible* is so closely linked with the basic aspects of freedom of expression that it should be treated similarly.<sup>13</sup>

### **7.2 Competition law**

It goes without saying that the generic national and European rules on competition apply to search engines. Abuse of a dominant position is prohibited and the European



Commission has specific powers to control mergers. But it is also quite clear that, under the present market conditions as described above, the position of one search engine in particular has begun to draw attention in that respect: Google.

It is difficult to say whether Google is abusing its market power at the present time. Before that can be done, we first need to establish what market search engines are actually operating in. More research is then going to be needed to reveal whether there is any abuse of power. Nevertheless, we can confidently identify some market areas in which there is a potential for abuse.

- a) *Inclusion in search results.* Information providers could object to the fact that they are being excluded from or incorrectly included in the results generated by searches. Thus far, no European case law exists to establish whether or not there is any entitlement to such inclusion. Under US law, search engines have successfully claimed that obligations to include specific search results infringe their freedom of expression (i.e. the famous Kinderstart case).
  
- b) *Preferential treatment for in-house information services.* Quite apart from the issue of whether other providers of information services are disadvantaged, it may be that the search engine's own services are given preferential treatment. Such a situation seems more likely the greater a search engine's interest in specific content becomes. One specific existing example is Google searches for video files, where results on Google Video and YouTube are - allegedly - given a preferred position.<sup>14</sup>
  
- c) *Access to the advertising market.* The business model adopted by search engines is driven predominantly by advertising. Large shares of the search market imply a concentration of so-called "eyeballs" – a phenomenon already familiar in the broadcasting market. This entails the risk that prices will be driven up, bias will occur and intransparancies become part of the advertising model.

Viewed from a merger's point of view, these three examples give rise to a number of pertinent questions. Competition in the marketplace could be affected adversely if, for example, (a) other search machines were taken over, (b) there were a takeover within

the vertical business column (content) or (c) there were a horizontal takeover in the advertising brokerage market.

Within competition law, there is also the issue of whether search engines qualify as an “essential facility” (the term “natural monopoly” has even been used!). Essential facilities are primarily a feature of network-related sectors; whether a service counts as one depends in part upon whether substitution is possible. And one important factor in determining that is how high the barriers to entry are. In the case of search engines, it can be stated that in principle those barriers are very low indeed and that setting up a new service is by no means a complicated procedure. This is a point of view I have adopted in the past, but it has to be said now that there is good reason to review that opinion. In particular, Google’s dominant position raises the question of whether relevant substitution really is possible. Let me give just one example. If the database built up by Google is indeed significant in its own right, then we have to ask whether other market players are still in any position to put together comparable databases of their own.

### **7.3 Sector-specific regulation**

What about the applicability of sector-specific regulation? The present European involvement with both the media and the telecommunications sector does not really take search engines into account.

Both the Television without Frontiers Directive and its successor, the Audiovisual Media Services (AVMS) Directive regulate primarily traditional television broadcasting and explicitly exclude services like search engines.<sup>15</sup> The framework for the communications sector has a similar handicap. Under the definitions in its core “Framework”-directive<sup>16</sup>, only electronic communications services are covered. That means services which consist “wholly or mainly in the conveyance of signals on electronic communications networks”. Service providing or exercising editorial control over content are excluded.

In my view, search engines have characteristics of both information and communications services. They are a good example of convergence in the information

society. But the information service aspects dominate: it is an understatement to see search engines as a mere directory service.

#### **7.4 Privacy**

The same applies to privacy as to freedom of expression. It is a right which enjoys constitutional protection under Article 8 of the European Convention on Human Rights and Articles 7 and 8 of the EU Charter. European law on this matter is further defined in a general privacy directive and a special directive applicable to the telecommunications sector.<sup>17</sup>

In general terms, the European privacy rules are easy to describe. They are based upon the principle that a minimum of personal data should be stored and processed, and that there must exist a direct relationship between what is done with data and the reason why it has been collected. Moreover, permission is required to gather data and the person involved must be able to verify and correct the information held. In all cases, proportionality is required. And compliance is regulated. The national regulators in Europe are member of an official working party,<sup>18</sup> which has recently launched an investigation into Google's observance of the European privacy regulations. This has prompted a correspondence<sup>19</sup> with the company, including a reference by the working party to the Resolution on Privacy Protection and Search Engines adopted at the 28th International Data Protection and Privacy Commissioners'.<sup>20</sup> This resolution more or less transposes the mentioned general characteristics into conditions relevant to the situation of search engines. The agenda has thus been set, with the working party indicating that it has now begun a general investigation of search engines. "Taking into account the current situation initiated by the 'Google case'," it says, "the Working Party will deal with search engines in general, because this issue affects an ever growing number of users."<sup>21</sup>

The privacy directive for the communications sector contains more detailed rules, specifically covering that industry's services. As well as upholding the confidentiality of communications, it regulates such matters as the use of traffic and location data. As mentioned earlier, these rules are not specifically tailored to the search-engine industry either and it is quite uncertain whether the directive applies to them at all. As far as is

known, no regulator has yet issued an opinion on that applicability. What is certain is that some other services frequently provided by search engine operators – such as e-mail – are governed by the directive. However, in this respect search engine operators do not substantially differ from traditional internet service providers.

## **8. Analysis**

As stated at the beginning of this paper, search engines represent an essential part of the way in which digital information is made easily accessible. But they have also become a bottleneck in access to information, with both its providers and users dependent upon the engine's intermediary function. At the same time, the way in which search engines work poses quite a few challenges. Nevertheless, they are able to generate serious revenues, primarily through advertising. But new elements are now being added, covering both vertical and horizontal issues – control over content, expansion into other advertising markets and marketing areas, and so on. Meanwhile, Google's dominant position in the market cannot be ignored. Policy makers and regulators are becoming increasingly aware of the role played by search engines in society, and the possible effects of reduced competition in the sector.<sup>22</sup>

The interests at stake are huge, certainly in a situation where market dominance is a factor. It is possible that there may eventually be some role for competition law here, but more pressing and increasingly relevant is the question of whether sector-specific regulation is needed for search engines.

From a European perspective, that could take its lead from the industry-specific frameworks applied to the telecommunications sector. However, the rules as they currently stand simply do not take into account a phenomenon like the search engine. Despite that, it is quite possible to investigate whether existing legal concepts like “significant market power” should be applied in this domain. Search engines with significant market power could be required to comply with obligations in respect of such matters as access, non-discrimination, transparency and accountability. Even where processes of a commercially confidential nature are at issue, that should not stand

in the way of independent audits. They could, for example, establish whether search results are indeed generated in an objective way. They could also investigate whether recorded data is being stored and processed correctly. (The existing privacy regulations might in fact sufficient for this to be done already, but so far they have never been invoked to justify checks or audits of search engines).

At the same time, the universal service/public good aspects of search engines need to be borne in mind.<sup>23</sup> Their users are entitled to minimum guarantees in respect of the way their operators work: they need to be properly informed, and misleading them has to be prevented.

## Literature

### **Elkin-Koren 2001**

N. Elkin-Koren, "Let the Crawlers Crawl: on virtual gatekeepers and the right to exclude indexing", *University of Dayton Law Review*, 26 U Dayton L Rev. 179, Winter 2001.

### **Fallows 2005**

D. Fallows/Pew Internet & American Life Project, *Search Engine Users*, Washington 2005.

### **Gasser 2006**

Urs Gasser, "Regulating Search Engines: Taking Stock and Looking Ahead", in: 8 *Yale J. L. & Tech. 201*, 2006, pp. 124-157.

### **Grimmelmann 2007**

J. Grimmelmann, "The Structure of Search Engine Law" in: 93 *Iowa L Rev.* (2007, forthcoming). [http://works.bepress.com/cgi/viewcontent.cgi?article=1012&context=james\\_grimmelmann](http://works.bepress.com/cgi/viewcontent.cgi?article=1012&context=james_grimmelmann).

### **Introna & Nissenbaum 2000**

L. Introna & H. Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters", in: *The Information Society*, 1 July 2000, vol. 16, no. 3, pp. 169-185, p. 171.

### **Lawrence & Giles 1999**

S. Lawrence & C. L. Giles, "Accessibility of information on the web", in: *Nature*, no. 400, 1999, pp. 107-109.

### **Louderback 2007**

J. Louderback, "Google's Gambit", in: *PC Magazine*, 17/7/2007, p. 7.

**Nicholson 2005**

S. Nicholson, "How Much of It is Real? Analysis of Paid Placement in Web Search Engine Results", in: *Journal of the American Society for Information Science and Technology*, 2005.

**Rainie & Shermak 2005**

L. Rainie & J. Shermak, "Search Engine use November 2005", memo Pew Internet & American Life Project/Comscore Communications, 2005,  
[http://www.pewinternet.org/pdfs/PIP\\_SearchData\\_1105.pdf](http://www.pewinternet.org/pdfs/PIP_SearchData_1105.pdf).

**Schulz, Held & Laudien 2005**

W. Schulz, T. Held & A. Laudien, "Search Engines as Gatekeepers of Public Communication: Analysis of the German framework applicable to internet search engines including media law and anti trust law", in: *German Law Journal*, vol. 06, no. 10, 2005, pp. 1419-1433.

**Van Eijk 2004**

N. A. N. M. van Eijk, "Regulating Old Values in the Digital Age", in: C. Möller & A. Amouroux (eds.), *The Media Freedom Internet Cookbook*, Vienna: OSCE 2004, pp. 31-38.

**Van Eijk 2006**

N. A. N. M. van Eijk, "Search engines: Seek and ye shall find? The position of search engines in law", *IRIS plus* (supplement to *IRIS - Legal observations of the European Audiovisual Observatory*),  
[http://www.obs.coe.int/oea\\_publ/iris/iris\\_plus/iplus2\\_2006.pdf.en](http://www.obs.coe.int/oea_publ/iris/iris_plus/iplus2_2006.pdf.en)) 2006-2, pp. 2-8.

**Zittrain & Edelman 2003**

J. Zittrain & B. Edelman, "Documentation of Internet Filtering Worldwide", in: C. Hardy & C. Möller (eds.), *Spreading the Word on the Internet*, OSCE: Vienna 2003, pp. 137-148.

## Notes

<sup>1</sup> Nico van Eijk is professor by special appointment of Media and Telecommunications Law. This paper contains updated parts of his inaugural address, of which an edited version was published in English (“[Search engines: Seek and ye shall find? The position of search engines in law](#)”, *IRIS plus* (supplement to *IRIS - Legal observations of the European Audiovisual Observatory*), 2006-2)).

<sup>2</sup> See, inter alia: Rainie & Shermak 2005.

<sup>3</sup> Introna & H. Nissenbaum 2000, p. 171.

<sup>4</sup> Liddy 2002, pp. 197-208.

<sup>5</sup> Lawrence & Giles 1999, pp. 107-109.

<sup>6</sup> See: Nicholson 2005. Also: <http://blogoscoped.com/archive/2007-07-16-n41.html> en <http://www.accc.gov.au/content/index.phtml/itemId/792088/fromItemId/142>.

<sup>7</sup> Zittrain & Edelman 2003.

<sup>8</sup> See, inter alia: Zittrain & Edelman 2003, p. 137-148.

<sup>9</sup> From the privacy notice of Google Docs & Spreadsheets: “*Account activity*. You need a Google Account to use Google Docs & Spreadsheets. Google asks for some personal information when you create a Google Account, including your email address and a password, which is used to protect your account from unauthorized access. Google's servers automatically record certain information about your use of Google Docs & Spreadsheets. Similar to other web services, Google records information such as account activity (e.g., storage usage, number of log-ins, actions taken), data displayed or clicked on (e.g., UI elements, links), and other log information (e.g., browser type, IP address, date and time of access, cookie ID, referrer URL).; *Content*. Google Docs & Spreadsheets stores, processes and maintains your documents and previous versions of those documents in order to provide the service to you... We use this information internally to deliver the best possible service to you, such as improving the Google Docs & Spreadsheets user interface and maintaining a consistent and reliable user experience. Files you create with Google Docs & Spreadsheets may, if you choose, be read, copied, used and redistributed by people you know or, again if you choose, by people you do not know. Information you disclose using the chat function of Google Docs & Spreadsheets may be read, copied, used and redistributed by people participating in the chat. Use care when including sensitive personal information in documents you share or



in chat sessions, such as social security numbers, financial account information, home addresses or phone numbers. You may terminate your use of Google Docs & Spreadsheets at any time. You may permanently delete any files you create in Google Docs & Spreadsheets. Because of the way we maintain this service, residual copies of your files and other information associated with your account may remain on our servers for three weeks.”

<sup>10</sup> On the legal aspects of search engines, see, inter alia: Elkin-Koren 2001; Schulz, Held & Laudien 2005; Grimmelmann 2007.

<sup>11</sup> See: Van Eijk 2004.

<sup>12</sup> To mention some of the legal issues which fall outside the scope of this paper: general liability issues, copyright, trademark, unfair business practices, criminal law aspects (including data retention) and e-commerce. We also overlook the issue of jurisdiction and assume that search engines – although mostly of US origin – have to comply with European regulation.

<sup>13</sup> Van Eijk 2006, p. 5.

<sup>14</sup> See: Louderback 2007. “Although there are thousands of useful video sources on the Net, Google delivers results only from its own YouTube and Google Video – along with third party MetaCafe. That’s just wrong, and...”

<sup>15</sup> EC Council Directive 89/552/EEC on the co-ordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the pursuit of television broadcasting activities, adopted on 3 October 1989, OJ L 298, 17 October 1989, p.23, as amended by Directive 97/36/EC of the European Parliament and of the Council of 30 June 1997 amending Council Directive 89/552/EEC on the co-ordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the pursuit of television broadcasting activities, adopted on 30 June 1997, OJ L 202, 30 July 1997, p. 60. The latest draft of the AVMS Directive:

[http://ec.europa.eu/avpolicy/reg/tvwf/modernisation/proposal\\_2005/index\\_en.htm](http://ec.europa.eu/avpolicy/reg/tvwf/modernisation/proposal_2005/index_en.htm).

<sup>16</sup> Directive 2002/21/EC of the European Parliament and of the Council of 7 March 2002 on a common regulatory framework for electronic communications networks and services (Framework Directive) OJ L 108/33 (24.04.2002).

<sup>17</sup> Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and

on the free movement of such data, OJ L 281, 23/11/1995 pp. 0031-0050; Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) OJ L 201/37 (31.07.2002).

<sup>18</sup> [http://ec.europa.eu/justice\\_home/fsj/privacy/workinggroup/index\\_en.htm](http://ec.europa.eu/justice_home/fsj/privacy/workinggroup/index_en.htm).

<sup>19</sup> See: [http://ec.europa.eu/justice\\_home/fsj/privacy/news/docs/pr\\_google\\_16\\_05\\_07\\_en](http://ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_google_16_05_07_en).

<sup>20</sup> d.d. 2/3 November 2006. Text of the resolution:

[http://ec.europa.eu/justice\\_home/fsj/privacy/news/docs/pr\\_google\\_annex\\_16\\_05\\_07\\_en.pdf](http://ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_google_annex_16_05_07_en.pdf).

<sup>21</sup> Article 29 Data Protection Working Party, press release, Brussels, 21 June 2007.

<sup>22</sup> Which has led to new support for creating European alternatives (The German Theseus and French Quaero-initiatives).

<sup>23</sup> Introna & Nissenbaum 2000.