



## UvA-DARE (Digital Academic Repository)

### Search engine freedom: on the implications of the right to freedom of expression for the legal governance of Web search engines

van Hoboken, J.V.J.

**Publication date**  
2012

[Link to publication](#)

#### **Citation for published version (APA):**

van Hoboken, J. V. J. (2012). *Search engine freedom: on the implications of the right to freedom of expression for the legal governance of Web search engines*. [Thesis, fully internal, Universiteit van Amsterdam].

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## **Chapter 3: Web search engines in the public networked information environment**

### **3.1 Web search engines: Basics**

#### **3.1.1 Definition**

As a legal category in information law the search engine is not well-defined. There are a number of legal categories that include search engines, for example 'information location tool',<sup>72</sup> but these definitions only serve specific legal contexts. In fact, it is unclear, and an interesting general question to which this research contributes, whether search engines should be a separate legal category altogether.

Search engine can be more easily defined from a functional perspective, in which case the following definition could be given for the search engine that is the subject of this study:

*an information retrieval system for the public networked information environment.*

Throughout this study the term public networked information environment is used to denote the collection of information which is publicly available on the Internet, and on the World Wide Web in particular.<sup>73</sup> Broadly speaking, search engines help end-users to find and effectively retrieve this information. In the following chapter the questions of what search engines are, what they do and how they do it, will be answered in more detail in order to have a better understanding of search engines for the remainder of this study.

The goal of this chapter is to properly conceptualize search engines from a functional perspective. The next section will present (1) the basic information flows between search engines, their users and information providers, (2) the typical architecture of a commercial search engine and (3) its user interface. In the section after that, the broader context in which search engines exist and operate will be discussed with reference to the layered model for networked communications and the value chains in which they play an important role. In the final section the two different functional roles which search engines perform in these value chains will be discussed in more depth, namely their role with regard to the end-users of search engines on the one hand and with regard to information providers and advertisers on the other hand.

#### **3.1.2 Basic information flows**

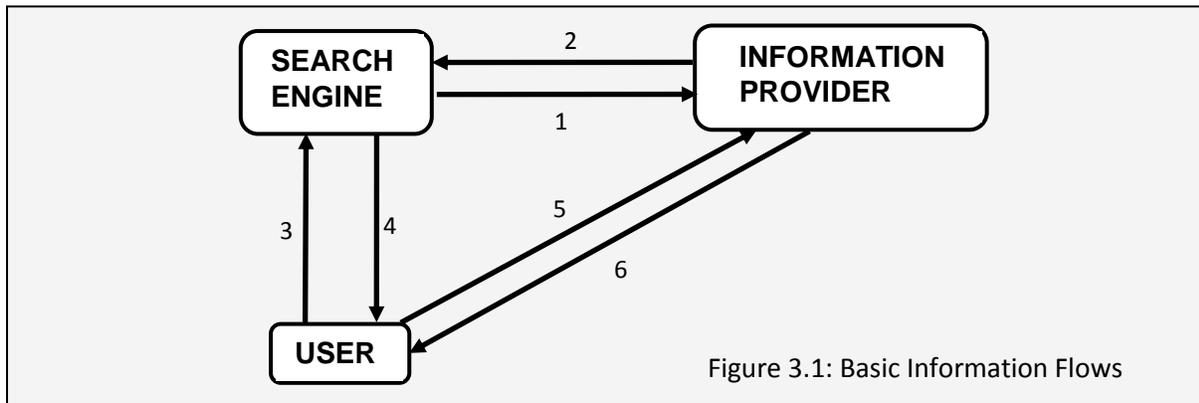
Figure 3.1 below shows the typical and most basic information flows induced by the operation of a search engine on the Web. The search engine is positioned in its intermediary position between the online information provider on the one hand and the Internet user on the other hand. It aggregates information and stores it in its index (2), typically with the use of sophisticated crawling software that makes automatic requests for the available online material (1). On the basis of that index and its subsequent analysis, it provides a service to end-users, whose input, in the form of search queries and other data (3) results in an output by the search engine, in the form of a ranked selection of references to and descriptions of certain information providers (4). From this selection the user can follow specific

---

<sup>72</sup> See Section 9.3.2 for the legal context and discussion of the 'information location tool' category.

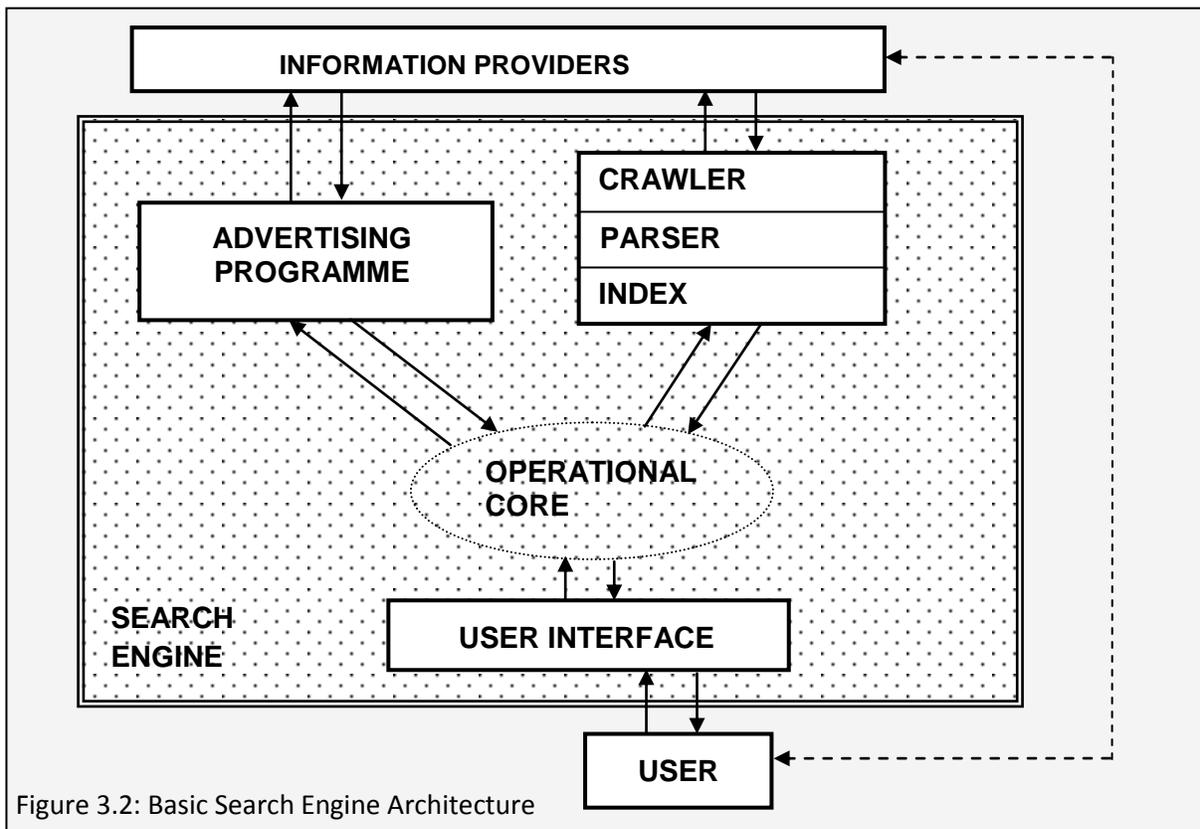
<sup>73</sup> The term 'networked information environment' is commonly used in the legal and scientific literature to denote the Internet as an information environment more generally. See Benkler 2006. See also Lynch 2001, pp. 12–17.

references and make an information request to a specific information provider (5) to receive the information provider's full information offering (6).



### 3.2.3. Basic Architecture

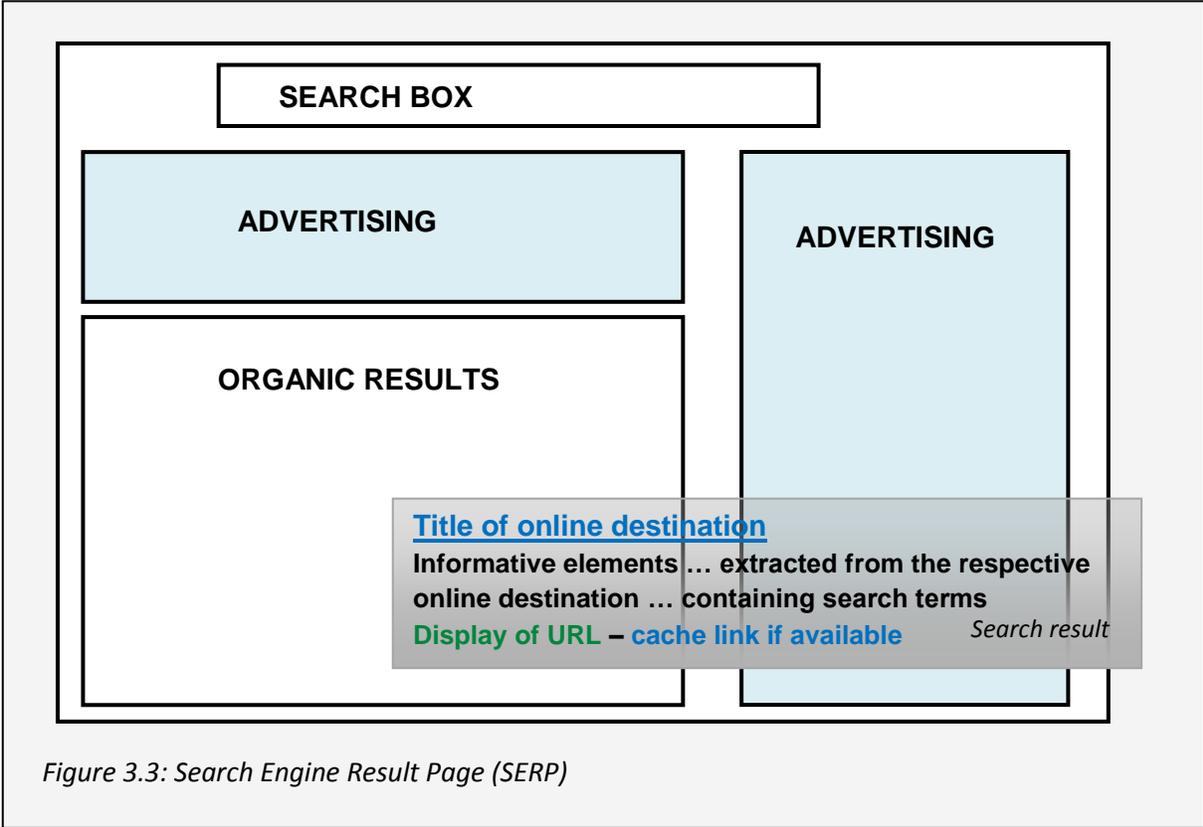
Schematically, the typical basic functional architecture of a search engine consists of a crawler, a parser, an index and a user interface (see figure 3.2).



The crawler, parser and index make up the technical back office of the search engine. The crawler interacts with the information providers. It is a complex computer program that looks for information on

the Internet, according to a set of criteria which tell it where to go and when. Over the years crawlers have become more sophisticated. They are now able to access more and more of the material that is available online, such as dynamically stored Web content, which used to be part of the so-called hidden Web.<sup>74</sup> The parser is the processing tool between the crawler and the index. It systematically stores the retrieved information in the index. The pieces of content the crawler finds are not the same in size, sort, language, code, and other characteristics, so the parser need to normalize them for the index. It also extracts a number of related data and meta-data that can be useful for the search engine’s technology.

The index is an ordered list of references to pieces of content on the net, but it is also more than that. The index is a large and complex database of which the references and keywords are basic elements. The index also contains information that is needed to apply the ranking algorithms. These lie at the heart of the search engine’s technology and determine which references are presented to users and in what order.



The user interface is the (layout of the) website offered to the user. The user interface design shapes the actual user experience.<sup>75</sup> The basic elements of the search engine interface are the search box and the search engine result page (SERP). For commercial search engines, the search engine result page (SERP)

<sup>74</sup> See generally, Olston & Najork, 2010. pp. 175–246.

<sup>75</sup> For a comprehensive and readable account of the various aspects of search engine user design, see Hearst 2009. Hearst does not discuss the optimization of the user interface from the perspective of advertising revenue optimization. See also Manning et al 2009; Höchstötter & Lewandowski 2009. For a discussion of transparency and advertising, see Section 10.3.1.

typically consists of a list of organic results, also called algorithmic or natural results, on the left hand side of the page and a list of sponsored results on the right, and in some instances at the top of the page. In addition, all major search engines incorporate a growing variety of relatively similar additional features such as a link to advanced features, image or video search, or a different language version. Some of these additional features can have some regulatory relevance, since they can be used to inform users about the ways they can control their search process. Examples of such features are Google's SafeSearch feature, which filters for adult content, or the link to its privacy policy, which Google somewhat recently added to its homepage.<sup>76</sup> Figure 3.3 below shows the typical layout of the result page as well as the typical layout of a search result, also called a search hit.<sup>77</sup> The layout of search advertisement has proliferated widely on the Web outside of the context of search engines due to contextual advertisement programs like AdSense, which use partner websites' content to place relevant advertisements on those websites.

### **3.2 Search engines in their context**

#### **3.2.1. Search engines and the layered model for the networked communications environment**

As discussed in Chapter 2, search engines exist in the context of the World Wide Web, the open hypertext structured information environment made possible by the Internet. Search engines have become central to the functioning of the public networked information environment, but from a technical perspective they are neither essential for this environment to exist nor for online material to be accessible to users. Conversely, it is true that the information flows induced by the use of search engines are made possible by the various protocols of the Internet's application layer, such as the HyperText Transfer Protocol (HTTP) which structures the service as it is made available to users. However, none of the layers of the layered model of networked communication (see Figure 3.4 below) contain elements specifically designed for search engines, even though these services play such a central role to the organization and navigation of content that is available on the network.

The layered model for the technical design of networked communications systems has inspired a functional layered model for such systems that contains three horizontal functional layers and which has gained some traction in regulatory debates about the networked communications environment.<sup>78</sup> These layers are a layer of physical infrastructure, a transport/logical/code layer and a content layer. Both the technical TCP/IP model and the functional model are shown in Figure 3.4 below. The horizontal character of this layered model can be contrasted with the traditional vertical regulatory models (silos) for various types of media or forms of communication, models which have been eroded by the phenomenon of convergence.

If one looks at the networked communications environment and the way in which the various communications network, technology and service providers map onto it, the functional layered model is sometimes used to conceptualize the role of specific entities in the communications environment on the

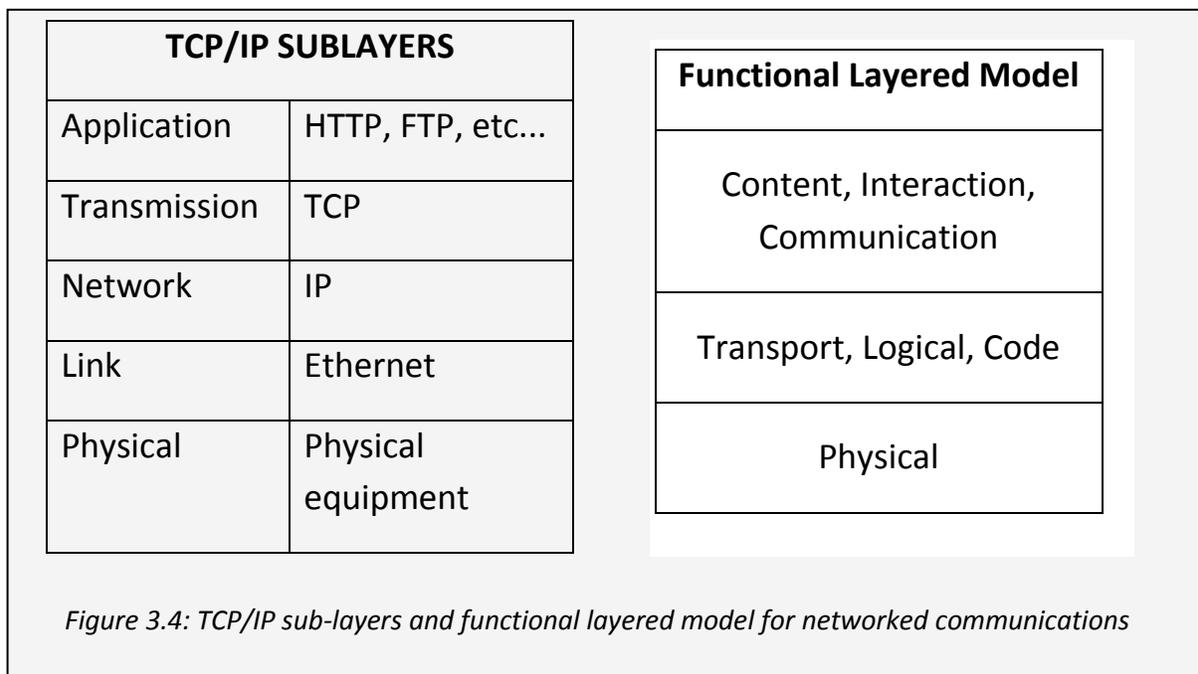
---

<sup>76</sup> See Anderson 2008.

<sup>77</sup> For a discussion of search user interface design, see Hearst 2009.

<sup>78</sup> For a discussion, see Solum and Chung 2003; See also Benkler 2000; Werbach 2002.

one hand. On the other hand it is used to frame the different regulatory questions that arise in the context of networked communications, more specifically in which layer specific legal issues can or should be addressed.



It should be noted, however, that the functional layers are an abstraction. Most importantly, the market structure and the practices of major players in the networked communications environment do not map nicely into the various layers of the model. The recent debate about net neutrality and the preservation of the end-to-end principle is a good example of this.<sup>79</sup> On the other hand, there are various forms of regulatory spillover, such as the enforcement of laws relating to the legality of content by targeting the Domain Name System (DNS) in the application protocol layer or even further down into the functionality of TCP/IP.<sup>80</sup>

If one uses the layered model for networked communications to look at search engines, they would seem to map principally onto the transport/logical/code layer. Web search engines are complex systems of software, typically server-based, made accessible for users of the network. But, search engines have a rather unique link with the content layer as well. First, Web search engines derive their functionality from the existence of publicly accessible content on the World Wide Web. Without the open and unstructured dynamics of content creation on the Web, search engines wouldn't have the pivotal role that they have today. And second, search engines can be argued to consume and produce 'content' on their own, namely information about information, or meta- information.

<sup>79</sup> For a discussion, see Van Schewick 2010. See also Wu 2003.

<sup>80</sup> For a discussion of Internet filtering by access providers, see Section 6.5.

Search engines present Internet users with human-readable representations of the networked information environment, while at the same time depending on various forms of machine-readable meta-information offered by other information providers or extracted from them or other places themselves. This relation to the content layer should definitely not be taken as a necessary or even sufficient argument for treating search engines, from a regulatory perspective, as traditional content providers (when they publish references for end-users) or as consumers of content (when they aggregate content for their index). What it does mean is that the conceptualization and regulatory treatment of this category of meta-information or 'meta-content' may be one of the key elements in properly solving the legal issues arising in the context of Web search engines.<sup>81</sup>

### **3.2.2. Search engines in the Internet communications 'chain'**

Another way to conceptualize search engines in the networked information environment is to position them in the chain or, more accurately, the network of communications on the Internet. From this perspective, a Web search engine, just like end-users and information providers, lives on the borders of the network, which itself is made up by basic communication services, such as hosting providers, access providers, and the Internet backbone at the highest level. Major search engines like Google have more than one location from which they provide their search service and typically connect to the Internet in multiple ways at a much higher level than a basic website. Still, this does not change the basics of this representation of the search engine in the network of Internet communications in Figure 3.5 below.

If anything, this representation clarifies the relation of the principal actors in the context of search engines to the Internet as a whole. But apart from that, it doesn't clarify the actual importance of search engines in terms of what is happening on the network. To really understand the role and importance of search engines in the public information environment it is more useful to modify this representation of search engines in the network of communication and focus on the flow of value instead of the flow of actual data.

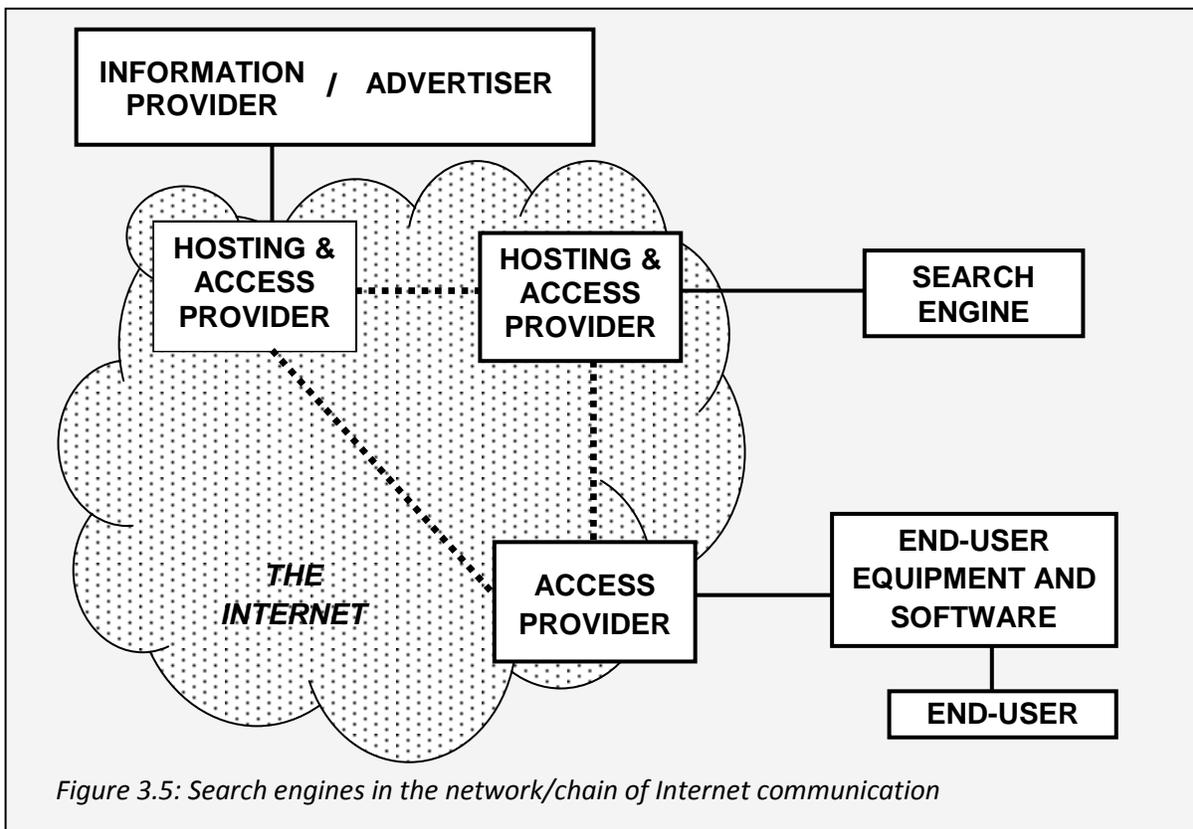
Over the last 15 years, Web search engines have become central brokers in many of the partly overlapping value chains in the networked online information environment. For instance, search engines connect end-users to informative online publications, political groups and various forms of e-commerce. These value chains in which the search engine operates can be broadly summarized from two different perspectives, namely a value chain flowing from content and service providers towards end-users and a partly corresponding but opposite value chain from end-users to information providers, both with search engines somewhere in the middle. In addition to clarifying the function of various entities in the information environment, the value chain perspective can be useful from a regulatory and policy perspective, because it helps to clarify the reasons for many of the conflicts between different entities in the value chain, conflicts that often arise because of the interest in control over such value.

The value chains in Figure 3.6 below are generalizations and simplifications. First, many entities that perform an important role in the current networked information environment are left out, such as hosting providers. In addition, a search engine provider may be accessible through the use of its website

---

<sup>81</sup> Compare Winkler 1997, pp. 35-36.

but it can also be built into the end-user's equipment's operating system, which is common in the case of mobile devices. Notably, the generalized flow of value does not necessarily represent the way in which money flows between the various entities in the chains. An in-depth analysis of the economic dynamics of these value chains is beyond the scope of this study.

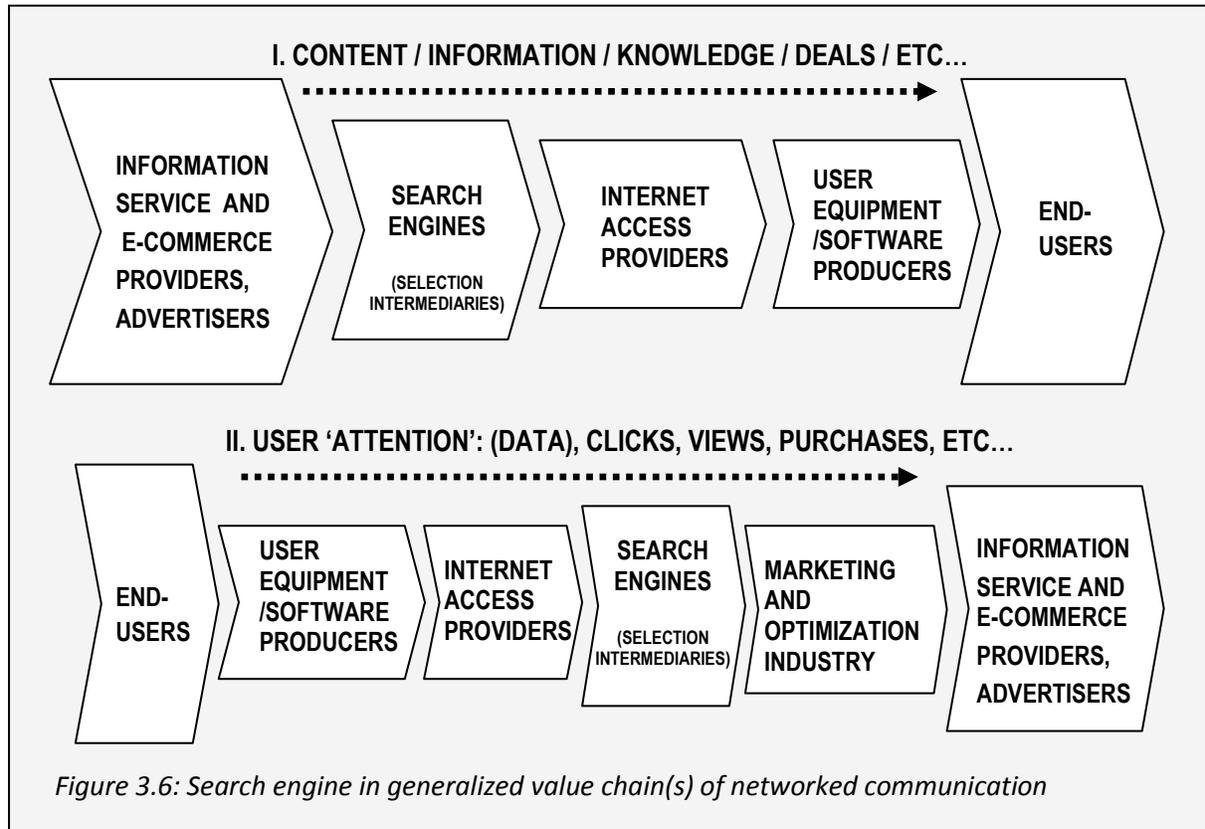


Finally, the search engine category in the value chain can be generalized into a broader category of 'selection intermediaries' which organize – or in the spirit of the Web's open information structure, let end-users organize – online materials and destinations by mapping, ordering, ranking, selecting, excluding, validating, and valuating them.<sup>82</sup> These selection intermediaries shape the relative accessibility of online material. As a category it includes not only search engines and directories, but other phenomena like portals, recommendation and bookmarking tools and services (delicious, StumbleUpon), social networking sites (Facebook), micro-blogging sites (Twitter), and news aggregators (Digg).

The first of the two value chains (I.) in Figure 3.6 is a generalization for the usual value chain for the flow of content, information and data to end-users. On the left, content is produced and published or made available online, services are offered and goods are sold on specific locations on the Web. All of these partly rely on the operations of search engines to find their way to the end-user. Between the end-user and the search engines, the access providers provide the essential service for end-users of enabling

<sup>82</sup> See Chandler 2008. See also Van Hoboken 2009.

them to go online in the first place. In addition, user equipment and software operating on this equipment can have an impact on the consumption of 'content' by end-users. Internet filters are a good example as well as browsers, toolbars and operating systems. Similarly, in the case of the mobile Internet the user equipment and mobile operating systems have been an important point of control in the value chain.



The second value chain (II.) is meant to illustrate the way in which the user represents the value that various information and service providers are competing for.<sup>83</sup> Search engines like Google and related advertisement networks such as AdSense are amongst the primary services to structure and sell the end-user's attention to those who are willing to pay for it. The portals we discussed in the last chapter play a similar role. Search engines have established a highly effective and lucrative monetization stream for end-user traffic, on which information, service, and e-commerce providers as well as advertisers have come to depend. Search engines auction the targeted user attention and information needs on their platforms, effectively selling the clicks of users in combination with other data about users and their activity.

As a result of the existence of this second opposite value chain, the search engine marketing (SEM) industry and the search engine optimization (SEO) industry play an important role in between the search engines and the information providers and advertisers. They help to optimize the traffic to their clients'

<sup>83</sup> See Van Couvering 2009. See also Bermejo 2007; Röhle 2007; Elmer 2004.

websites by structuring search advertising campaigns and optimizing their presence in organic results. User attention and user data in the form of click streams have become a basic value flow on the Web, and search engines are amongst the big players and amongst the big targets in the growing Internet marketing industry. Simply put, a website with sufficient traffic can make money on the Internet and search engines are important means to provide such traffic. Notably, this reality has had a major impact on the search engine industry as a whole, since search engines haven't always been successful in discerning between genuinely or just seemingly useful destinations for their users.

To summarize, the competition for search engine users induces information providers, including other search services, to pay for search engine advertising on the one hand and to invest in optimization of their ranking in natural search results on the other hand. However, some information providers will be, quite predictably, willing to optimize traffic from search engines to their sites for less benign reasons. This includes information providers that do not have genuine information offerings themselves and merely function as real estate for optimization instruments and advertisements. It includes information providers that offer information or services that would not have been selected by the search engine or ranked as prominently for particular queries if the search providers would have carefully evaluated the relevance manually. To give just one example, adult content providers can lure end-users to their website by wrongly suggesting they have adult content relating to often searched for celebrities.<sup>84</sup>

The existence of related and opposite value chains illustrate well the basic conflict of interest between information providers and end-users that a commercial search engine, as a matchmaker between different types of supply and demand in the public networked information environment, has to reconcile. On the one hand, user demand for information, knowledge and attractive e-commerce offers must be met by search engines in their competition to satisfy end-users. On the other hand, the profitability of the search platform is directly related to its ability to draw attention to those that are willing to invest in reaching an audience. Obviously, such willingness is not always a good predictor for the ultimate value of references for end-users.

To summarize shortly, in this section three perspectives on search engines in their context have been presented to come to a better understanding of the role of search engines in the public networked information environment. The next section will discuss the different functions the search engine performs with regard to end-users and information providers, the direct stakeholders in communications through search engines, in more detail.

### **3.3 Web search engines: a functional perspective**

#### **3.3.1 End-users: Information, Navigation and Transaction**

The information retrieval literature has developed several models to understand human interaction with search engines, generally focusing on people's reasons for using information retrieval systems and their strategies to arrive at satisfactory results when interacting with these systems.<sup>85</sup> From the perspective of

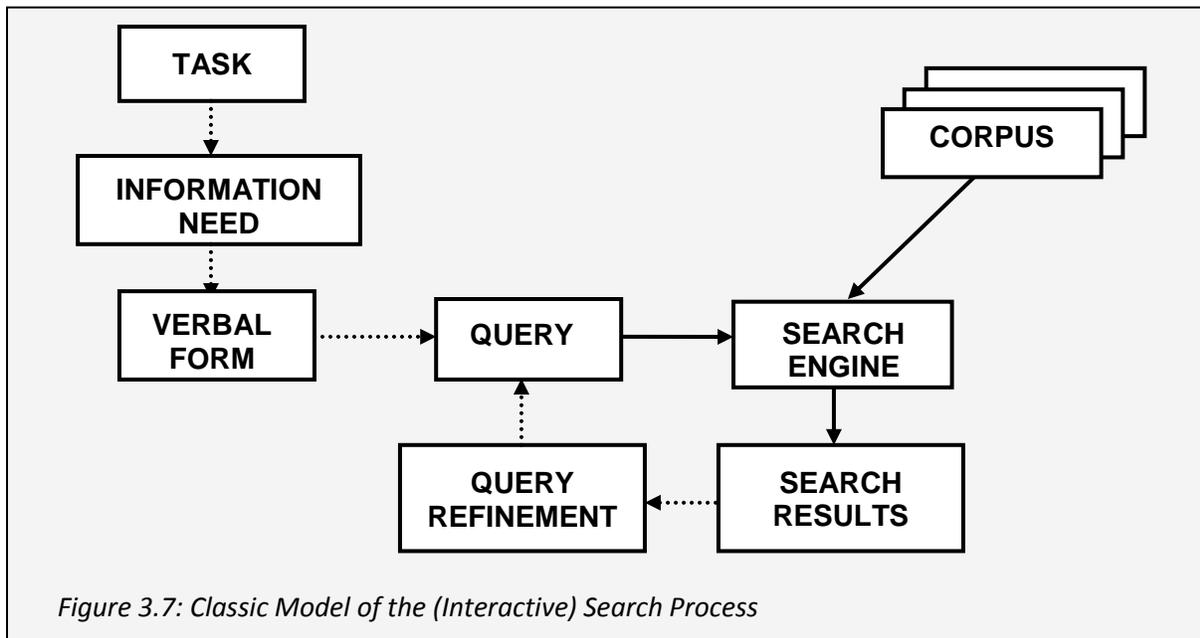
---

<sup>84</sup> See Chapter 8 and 10 for further discussion of the legal issues relating to so-called search engine spam and search engine manipulation from the perspective of search engine regulation and freedom of expression.

<sup>85</sup> See Hearst 2009.

search engine operations, these models play a crucial role in the interpretation of user queries, the design of search engines, and the way in which further interaction, such as query reformulation and evaluation, is facilitated.

An important general conclusion from this literature, that can help to understand the role of search engines in the networked information environment, is that there are many different situations in which Internet users choose to use a search engine in the first place.<sup>86</sup> Classic information retrieval systems, such as those offered in the academic realm or by libraries, had mostly focused on facilitating information needs of specialized users. The typical model for the interactive search process between users and the search engine in such information retrieval systems is shown in Figure 3.7 below. The information needs of the users in the classic information retrieval system could be broadly categorized as *informational*: a user wanted to find certain information that was presumed to be present in the corpus of information the information retrieval system was providing access to.



However, in the context of Web search engines it was observed that as a consequence of the open and more varied nature of the Web as an information environment, the information needs of end-users when engaging in a search process were more varied also. More specifically, Broder concluded that Web search engines were dealing with two additional types of frequent information needs, which he categorized as *navigational* and *transactional*.<sup>87</sup> He called a query *navigational* if a user wanted to find a specific web site which he knew or assumed to be present on the Web. He called a query *transactional* if the user aimed to reach a destination where further interaction would take place, such as making a purchase or obtaining a resource. Further research has offered additional insight into these different

<sup>86</sup> See Broder 2002. For a discussion, see Hearst 2009.

<sup>87</sup> See Broder 2002.

categories and the way in which search engines could respond more effectively to the different reasons end-users had for interacting with search engines online.<sup>88</sup>

The broad categorization of user needs in the information retrieval literature helps to clarify the rather broad societal role which search engines have come to fulfill.<sup>89</sup> First, in responding to the navigational needs of end-users search engines help users navigate to specific online locations. They respond to these navigational queries by directing users to the 'home page' of various organizations, institutions, companies or persons. Importantly, if the user's information need is known to be navigational, an example could be the query [University of Amsterdam], there is only one right answer that the search engine should give, namely the website of the University of Amsterdam. Evidently, search engines are the primary online destination for gaining access to this simple type of navigational information.

Second, search engines help end-users to find information about specific topics. They do so by returning search results directing to web sites that have information about these topics, in the form of general information, specific answers, advice, or lists of other relevant sources of information. In this category of informational queries, search engines help users to learn something about a topic, say [freedom of expression] or [Barbie]. They steer users to web sites that help them to find answers to specific questions, such as [EU Member States] and [gold price]), or obtain representative lists of certain entities [Amsterdam universities]. Finally, they help users to gain advice before making decisions or advice with regard to their problems, underlying queries such as [stop smoking], [vote Obama], [cheap tickets], or [cure headache].

Third, search engines direct users to purchasing opportunities, services and resources online, in the form of e-commerce sites [flowers delivery], [batelle the search], online services and entertainment [malcolm gladwell], [south park], [french anthem mp3], [el clasico streaming link], and other resources [weather report], [map Amsterdam], [apple pie recipe], [material girl lyrics]. This type of queries is directly related to the presence of directly available resources on the Web, which search engines help to place at their users' fingertips. The information need of the user is not related to reaching a specific defined destination (navigational) or generally obtaining further information on a topic (informational), but in reaching a destination that allows for the interaction he or she is interested in.

As is hopefully apparent from the examples given above, search engines will have to guess about the actual information needs of their users as expressed in different queries. The same query [Barbie] could be navigational: the user wants to access the Barbie home page of Mattel; informational: (the user wants to read about the history of Barbie or get information about product safety and quality for an article in a parent magazine; or transactional: the user wants to find a picture of a Barbie for a presentation about the history of toys or to simply buy a Barbie online. To a large extent, this type of intelligent guessing is precisely what offering a search engine is all about: to select and rank a list of online resources that has a good – or better, as high as possible – chance of satisfying the demand of the user as imperfectly expressed in a search query.

---

<sup>88</sup> See e.g. Rose & Levinson 2004. See also Jansen et al 2008.

<sup>89</sup> Examples are loosely based on Rose & Levinson 2004.

The basic model of information needs described above implies, for instance, that search engines will typically want to treat queries consisting of the name of an entity as navigational, returning the authoritative result in the first place.<sup>90</sup> For other types of queries, similar strategies can be observed and expected. The availability of the online collaborative encyclopedia Wikipedia has most likely been a particular blessing for satisfying informational queries. However, these are anecdotal characterizations of industry practices. The actual practices of general purpose search engines like Google are a lot more complicated and, for various reasons, remain largely undocumented. What has become clear is that over the last decade the analysis of historic user data has come to play a very important role for search engines in making informed guesses that satisfy their users and optimize themselves as marketing platforms. The legal issues that arise from the often opaque use of large amounts of user data from the perspective of freedom of expression will be discussed in detail in Chapter 10.

The value chains in which search engines operate may help to clarify some of the ways in which search engines have developed overarching strategies of discerning and responding to the various information needs of their users. For instance, the commercial nature of search engines has had an obvious impact on the way in which search engines would like to satisfy certain information needs of users. For instance, transactional information needs may be more likely to attract advertiser bidding. More specifically, commercial search engines as well as advertisers can be expected to be particularly interested in optimizing their strategies with regard to queries with high consumer intent.<sup>91</sup> At the same time, navigational queries are attractive because they still allow search engines to present end-users with alternatives, the reference to which search engines can easily justify with reference to query purpose ambiguity. Finally, the position of search engines in various value chains of online resources can also explain specific vertical integration strategies by search engines. Seeing that they direct users attention away to popular third party resource collections, search engines may come to the conclusion they would be better off to develop their own competing resources. If we take Google's search service, Google Maps (launched in 2004 to compete with AOL's MapQuest), Youtube, and the unsuccessful project called 'knol' (commonly interpreted as a means to compete with Wikipedia), may all serve as examples.

In summary, search engines go well beyond the function of what could be seen as a simple telephone directory for the Web. They help users with a large variety of quite different information needs by actively selecting and ranking lists of online destinations. These information needs of users range from political, medical and educational to commercial, domestic and recreational. Clearly, this shows not only the societal breadth of the function of search engines in our information environment but also hints at the variety of important public and private interests that are tied to their operation. The fundamental legal questions relating to the role of search engines in providing access will be discussed in detail in Chapter 8-10.

---

<sup>90</sup> See e.g. the leaked document relating to Google's human search quality evaluation program, Google 2007b. For a discussion see Sullivan 2010a.

<sup>91</sup> For a discussion of the way in which search engines can strategically respond to such queries, see Dai et al 2006.

On a more fundamental level, the way in which search engines end up selecting and ranking results for their users can be characterized as the expression of a range of underlying judgments about the relevance of various kinds of information and destinations in relation to the relative importance of the perceived needs of their users. Hence, on the one hand, the search medium is a facilitator to the communicative process between end-users and information providers. It does not itself, in its role as a search engine, publish information on the Web, but merely provides ranked collections of references to third party content to end-users. On the other hand, the search engine does have to make a variety of choices about the way in which 'relevant' Web destinations will be selected, ranked and presented to its users on the basis of its analysis of the material on the Web and its interpretation of the perceived information needs as expressed by the users' queries. In the chapters that follow the important question will be addressed of to what extent these type of choices, the existence of which give some of the search engine's operations an *editorial* character, must be qualified under the right to freedom of expression.

### **3.3.2 Search engines as forums for information providers and marketing platforms**

As previously stated, the search engine is one of the most important ways for information providers to reach an audience online. Information providers, online services, e-commerce providers, and advertisers have come to depend on general purpose search engines to establish an online presence and/or be successful as a business. Understandably, information providers and search engines each have their own preferred terms and conditions for carriage.

From the perspective of the information provider, one can discern three ways in which they tend to end up in the index of general purpose search engines. The first and standard way is that they simply allow their publicly available website to be indexed and ranked by the search engine in the organic or natural results. The information provider just puts his content online and waits for the search engine's crawlers to come by and include it. In this context it is important to mention that information providers can instruct search engines with the use of the *robots.txt* protocol, a de facto standard in the industry, whether or not they want to be indexed in the first place.

There are several services that can help to speed up the process of being included in search engines and major directories, and some search engines, including Google, offer a special set of tools for webmasters with which they can monitor the way they are included in the index. Markedly, whereas some information providers will simply sit back and let search engines include them in their service, others will use all the available means to optimize the way in which search engines help them to reach end-users through their organic listings. Notably, the most important aspect to optimize for information providers is not their plain inclusion in the index but their selection as relevant and receiving optimal ranking in response to relevant user queries.

The second way in which information providers end up in search engine listings is by way of participation in paid placement programs. As discussed in Chapter 2, most search engines are commercial and have developed a monetization strategy that is based on the auctioning of targeted advertising space tied to specific end-user queries. Search engines cater to the information needs of

users, who are not simply looking for the location of a website or information about a topic, but also as the starting point to make a purchase. In addition, the granularity of targeting which search engines can provide is highly attractive. Google's AdWords program has been the most successful paid listing program online and includes a growing range of additional choices to select and optimize paid placement of advertisements in response to particular queries by particular groups and types of users.

The third way in which search engines could end up referencing to certain information is by contracting with information providers to be able to include them in their index, or to include them more effectively. The deals between Google and Twitter about (more) effective retrieval of tweets by Google users can serve as an example.

Notably, search engines are themselves an example of an information provider that sells its information to be included in other search engines. Many search services and portals contract other search engines for search results and provide their own information service with them. Ask stopped producing its own search results in 2010 and is said to pay Google for its organic search results.<sup>92</sup> Yahoo has contracted with Microsoft for organic results. Generally speaking there are hundreds of search engines but only a small percentage creates its own index. Many vertical search engines operate like this. How often general purpose search engines enter into these kinds of arrangements is not well documented. Most commonly, search engines like Google would simply buy the owner of this exclusive set of information and related technology and integrate it into its service.<sup>93</sup>

### **3.4 Conclusion**

This chapter has first clarified the basic elements of the search engine architecture, most notably the crawler, the index and the user interface. Furthermore, the role of search engines in the networked communications environment has been discussed, with reference to the layered model for networked communications and in respect to two interdependent and partly opposite value chains. And finally, the different functions the search engine performs with regard to end-users and information providers, the direct stakeholders in the communications through search engines, have been discussed in more detail.

The analysis of the position of search media in the layered model of networked communications, clarifies that search engines map both to the top of the application and services-layer as well as to the content layer.<sup>94</sup> On the one hand, Web search engines are complex systems of software, typically server-based, made accessible for users of the network in their Web browsers. On the other hand, search engines have a rather unique link with the content layer. Search engines can be argued to consume and produce a specific 'content' of their own, namely information about information, or meta-information.

The representation of the search engine in view of the essential value chains in the public networked information environment offers more insight into the important position search media find themselves in. The first value chain in which the search engine plays an important role is the flow and control over

---

<sup>92</sup> Notably, the Ask website does not refer to the actual producer of the references it shows to its users and it refused to clarify the actual source of organic results when asked for it by specialized media. See Sullivan 2010b.

<sup>93</sup> See Chapter 2 for an overview of business developments in the search engine industry.

<sup>94</sup> See also Wu 2010, pp. 279-289.

knowledge, information, data, news, offers, etc., from all sorts of online information and service providers to end-users. The second value chain, which is of greater importance from a business perspective, represents the flow and control over user attention and activity, in the form of their page views, clicks, purchases and personal data. In both of these value chains, search media, and what can be denoted as selection intermediaries more generally, have established themselves as one of the central mediating institutions. More specifically, search media are uniquely situated to negotiate between these different kinds of value, a position from which they derive most of their power.<sup>95</sup> Search engine users retrieve value flowing through them including through search engines, in the first chain, in return for which they subject themselves to the extraction of value flowing away from them to information providers in the second value chain.

In comparison to traditional information retrieval systems, in which the information needs of users were typically restricted to the *informational*, Web search media tend to serve two additional types of user needs, namely *navigational* and *transactional*. Navigational queries are the type of queries with which users aim to reach a specific online destination which they know or simply assume to exist. By satisfying navigational queries, search engines help Internet users to reach the home page of various institutions, organizations, companies or persons. Informational queries represent the user needs that are directed at learning something about a certain topic. Transactional queries represent the type of user needs which are directed at reaching a destination where the user will be able to use or consume a resource.

The conclusion follows that the function of search media goes well beyond the function of what could be seen as a simple telephone directory for the Web. They help users with a large variety of quite different information needs, by selecting and ranking lists of online destinations. These information needs range from political, medical and educational to commercial, domestic and recreational. This shows not only the societal breadth of the function of search engines in our public networked information environment but also hints at the large variety of public and private interests that are tied to their operation.

Finally, search engines end up selecting and ranking results for their users. The choices of how to do that can be seen as the expression of a range of underlying judgments about the relevance of various kinds of information and destinations in relation to the relative importance of the perceived needs of their users. To perform its function as matchmaker between information providers and end-users, the search engine has to make a variety of choices about the way in which 'relevant' Web destinations will be selected, ranked and presented to its users on the basis of its analysis of the material on the Web and its interpretation of the perceived information needs as expressed by the users' queries. Notably, this type of choices may give some of the search engine's operations an *editorial* character. In the chapters that follow the question will be addressed, amongst others, of what that means from the perspective of the right to freedom of expression.

---

<sup>95</sup> See Röhle 2009. See also Wu 2010, pp 279-289.