The London School of Economics and Political Science

Search Engine Bias

The Structuration of Traffic on the World-Wide Web

Elizabeth Jane Van Couvering

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Abstract

Search engines are essential components of the World Wide Web; both commercially and in terms of everyday usage, their importance is hard to overstate. This thesis examines the question of why there is bias in search engine results – bias that invites users to click on links to large websites, commercial websites, websites based in certain countries, and websites written in certain languages.

In this thesis, the historical development of the search engine industry is traced. Search engines first emerged as prototypical technological startups emanating from Silicon Valley, followed by the acquisition of search engine companies by major US media corporations and their development into portals. The subsequent development of pay-per-click advertising is central to the current industry structure, an oligarchy of virtually integrated companies managing networks of syndicated advertising and traffic distribution. The study also shows a global landscape in which search production is concentrated in and caters for large global advertising markets, leaving the rest of the world with patchy and uneven search results coverage.

The analysis of interviews with senior search engine engineers indicates that issues of quality are addressed in terms of customer service and relevance in their discourse, while the analysis of documents, interviews with search marketers, and participant observation within a search engine marketing firm showed that producers and marketers had complex relationships that combine aspects of collaboration, competition, and indifference.

The results of the study offer a basis for the synthesis of insights of the political economy of media and communication and the social studies of technology tradition, emphasising the importance of culture in constructing and maintaining both local structures and wider systems. In the case of search engines, the evidence indicates that the culture of the technological entrepreneur is very effective in creating a new mega-business, but less successful in encouraging a debate on issues of the public good or public responsibility as they relate to the search engine industry.

Acknowledgements

Search Engine Bias:

The Structuration of Traffic on the World-Wide Web

Table of Contents

List of Figures

List of Tables

1

Bias in Internet Search Engines

1.1 Introduction

The new 'electronic monks' will not be guardians of the scarce information; but they will be guiding users to the relevant information. The old monks may have been defending the information storehouse; the new electronic monks will provide the compass to guide users through the ocean of information. But in each case, it is the monk who stands at the gateway to the information that will lead to knowledge and understanding. (Melody, 2003, p. 11)

The thesis begins from a general interest in the role of information and communication technologies (ICTs) in creating what some have called the Information Society, a new type of social and economic organization akin to the agricultural or industrial society, but based upon wide availability of information and a new dominance of the role of information in everyday life (Webster, 2002). The internet, it has been argued, has a key role to play in this transformation. According one prominent theorist writing about the Information Age, Manuel Castells:

> "The internet is the fabric of our lives. If information technology is the present-day equivalent of electricity in the industrial era, in our age the internet could be likened both to the electrical grid and the electric engine because of its ability to distribute the power of information throughout the entire realm of human activity" (Castells, 2001, p. 1)

Nevertheless, despite the increasing importance of information in understanding modern society, critical scholars Herman and McChesney make the claim that the public interest in the internet is being left behind, giving citizens "no rights in cyberspace beyond those it

[they] can exercise as capitalists or consumers" (Herman & McChesney, 1997, p. 133). It seems to me that the very centrality of the internet to the information society demands scholarly inquiry into how the network is being used outside its commercial context. Therefore, this research focuses on an area of the internet where the public interest may be being undermined by commercial activities; an area which has no direct analogue in the non-internet world and yet is recognisable as a media market: the internet search engine.

Search engines, also sometimes called portals, are particularly interesting because of their centrality as both starting points and guides in the emerging content structure of the World Wide Web[1]. As the quote from William Melody that introduces this section indicates, those who stand at the gateway to information exercise a particular power over that information even when their power derives from an ability to guide the search engine user.

This thesis is not a study of internet infrastructure in the traditional sense of fibre-optic cables or computer hardware. Rather, it is a study of some of the structuring principles of online content. To take a television analogy, it focuses neither on particular shows nor on audience reception, but rather on practices akin to scheduling. To take a newspaper analogy, it focuses not on the text of particular stories but rather on layout and distribution.

1.2 Researching search engines and identifying a gap

Existing research on search engines is not an integrated field and covers a wide range of disciplines.[1] In their systematization of search engine literature, Machill et. al. (2008) divide the field into five areas: search engine policy and regulation; search engine economics; search engine quality and technology; user behaviour and competence; and search engines and journalism. Halavais, in his important introductory text *Search Engine Society*, uses his chapters to implicitly create following divisions: engines (about the technology and history of search); attention (about search engine advertising); knowledge and democracy (about search engine skills, search engines in areas other than the US, and search engines in

---

[1] 'Search engine' here is used in a broad sense, and refers to that genre of Web site which prominently feature Web search facilities and in addition may include news, directory services, shopping, and other 'portal' services. Using this definition, for example, Yahoo!, Google, Excite, and MSN are all 'search engines.'

relation to existing knowledge institutions); censorship; privacy; sociable search (about search and social networks); and future finding (about the future technical development of search).

My own typology is as follows, where I divide the field of existing research into five perspectives, as Machill does, but with slightly different emphasis and including some issues raised by Halavais. First comes the *information retrieval* perspective, where the search engine is studied as a complex programming problem. From a disciplinary standpoint, this perspective has its roots in library and computer science. Second, *information literacy*, where the interactions between search engines and user skills are of paramount concern. This perspective draws both from sociology and literacy but also from disciplines like human-computer interaction for eye-tracking studies and from computer science and statistics in terms of log-file analysis. Third, *media economics and online marketing*, where the search engine's commercial structure and its use as a marketing tool are of central concern. This perspective has contributions from media studies, business studies, and occasionally economics. Fourth, there is emerging research on how various communities interact with search engines, including journalists and academics, as well as various publics, in an area that might broadly be referred to as *search engine effects*. This perspective draws so far from journalism studies, science studies and to some extent computer-mediated communication studies. Fifth, *search engines and society*, where public concerns relating to search engines are being debated, such as censorship, privacy, and bias. Key contributions here have been from law, political science, and media studies. Thus while there may be said to be broad agreement that something in search engines is worth studying, exactly what and exactly how depend upon the researcher.

Nevertheless, the area of search engine studies (studies, that is, which are not focused upon search algorithms or programming) is beginning to develop a series of key works that have been published for the most part in the past five years. In addition to Halavais's book, there are also two important edited collections upon the subject of search, one edited by Spink and Zimmer (2008) and one edited by Machill (2007) and partly in German, and two important journal issues, an issue of the Journal of Computer-Mediated Communication edited by Hargittai (2007) and an issue of the Yale Journal of Law and Technology in which both Goldman (2006) and Gasser (2006) published articles. Finally,

two useful popular books also cover the history of the whole search industry (Battelle, 2005) and Google in particular (Vise & Malseed, 2005). And, although technical, Brin and Page's description of Google's PageRank algorithm (1998) and Kleinberg's description of the HITS algorithm (1998), both central to modern search engines, must also be considered core works.

Taking these perspectives in turn, the following section summarises some of the core issues for search engine research in these areas.

1.2.1 INFORMATION RETRIEVAL

Within the information retrieval perspective, one main concern is to optimize the quality of search results. It is known that search engines give a particular slant to the Web. The results retrieved represent only a portion of the Web and over-represent popular sites, commercial sites and American sites[2] (Bergman, 2001; Kleinberg, 1998; Kleinberg & Lawrence, 2001; Lawrence & Giles, 1999; Lewandowski, 2008b; Vaughan, 2004; Vaughan & Thelwall, 2004). Search engines have ongoing problems keeping their indexes "fresh," e.g., populated with more or less current content (Lewandowski, 2008a). Höchstötter and Lewandowski (2009) reviewed the composition of results on the average search results page for five separate engines, and came to the conclusion that Google and Yahoo favoured their own subsidiaries, with Google, for example, consistently returning far more YouTube links than other searches. Many of those who have reported these findings have tended to see these issues as technical problems. For example, Vaughan and Thelwall maintain that the over-representation of American sites "…is due not to deliberate choices of the search engines but occurs as a natural result of cumulative advantage effects of US sites on the Web" (Vaughan & Thelwall, 2004, p. 693). Kleinberg (2001), a mathematician, sees the Web as a "self-organizing" system.

1.2.2 INFORMATION LITERACY

Information literacy studies, which concentrate on the varying abilities of the ordinary user to use the search engine, suggest that many users may be ill-equipped to counter the bias of search (Hargittai, 2002). Log file analysis reveals a stable finding that most people

---

[2] Interestingly enough, these studies did not indicate that the language of a web site was in and of itself a factor in whether it was more or less likely to be found by search engines.

use very short, imprecise search terms and rarely click beyond the first page of results (Jansen & Spink, 2005, 2006; H. C. Ozmutlu, Spink, & Ozmutlu, 2002; S. Ozmutlu, Spink, & Ozmutlu, 2004; Spink, 2002; Spink, Jansen, Wolfram, & Saracevic, 2002; Spink, Wolfram, Jansen, & Saracevic, 2001; Wolfram, Spink, Jansen, & Saracevic, 2001). They also tend to believe that what comes at the top of the search results must be the best result, even when researchers have deliberately scrambled the results, leading Keane (2008) to suggest that it is the users who are biased, that is to say, overly influenced by the "trustworthy" brand of the search engine. Appendix E presents the literature on internet navigation and search engine usage in some detail, particularly focusing on a very large longitudinal study by Beauvisage (2004) including both log analysis and interviews, and lends empirical support to a conclusion that bias in search particularly affects inexperienced users, users under time pressure, and those searching for small websites.

1.2.3 Media Economics and Online Marketing

Rogers (2000) raised the issue of pressure by advertisers to add paid-for links to search results ('preferred placements' in Rogers's terms), a new and highly controversial development at the time he was writing. It is now commonplace. Rogers remarks that preferred placements challenged the idea that search engines were the objective 'librarian' of the Web-as-library metaphor, or the 'telephone book' of the Web-as-communication-network metaphor, and instead highlighted the "significance of the Web as a realm for media strategists *operating* the network, as opposed to mere information disseminators and/or creative actors *using* the network" (Rogers, 2000, p. 13). Halavais (2009) raises the point that search is an "attention economy" and dependent upon advertising. Bermejo (2009) agrees, but he argues that there is nothing fundamentally different about the "audience commodity" constructed by search engines and that constructed by television. I discuss this question further in Chapter 4.

Machill et al. (2003) point to the effects of other significant actors seeking to use or influence the rankings of a particular search query, and include web site owners/operators, search engine optimization companies (SEO's), search engine operators, and advertisers (of various varieties e.g., banners and buttons as well as paid results listings). In addition to the less than transparent tactics that Machill et al. focus on, there are also many other strategies employed by these stakeholders – for example, web

site providers may change the copy or technical structure of their site to be more search-engine friendly on the advice of search-engine optimizing companies.

Research in the online marketing tradition suggests that marketers are very aware of and active in the manipulation of search results in order to benefit the organizations they work for. For example, one article aims to teach search-engine optimization (SEO) "style" to public relations (PR) agents to help ensure the articles that they place are found online (Fallhauber, 2009); another considers how best the tourism industry can represent destinations in search engine results (Xiang, Wober, & Fesenmaier, 2008). Popular books also offer advice to marketers and business owners seeking to improve their position in the search engine rankings (Fleischer, 2009; K. B. Jones, 2008). Much of the technical infrastructure of search is driven by the quest for profit, so some papers deal with the way in which search technology is transforming and being transformed by technology, for example, Röhle's (2007) article on personalized web search and marketing and Zimmer's (2008) article about Search 2.0.

1.2.4 SEARCH ENGINE EFFECTS

The study of search engine effects as I have termed it, has concentrated primarily on academics and journalists as key groups in society whose daily work is information rich (lawyers would also fit into this category although I have yet to see a comparable study). Machill and Beiler (2009) have pointed out the way in which search engines have altered journalists' routines and their sources, potentially for the worse. Carlson (2007) suggested that search engines are a positive influence, disrupting the ordered routines of journalists looking for information, empowering non-traditional journalists such as bloggers and further that news search engines decontextualise news items, making news more difficult for journalists to package and control.

A similar ambiguity is seen in the studies of the effects of search engines on academics, with some negative and some positive effects noted. Brown et al.'s (2008) study of a scientific literature review for a fossil amphibian noted that search engines, even those seemingly targeted to scientists such as scholar.google.com and specialist engines like JSTOR and GeoRef, are inadequate for high-quality scientific research in some fields, with the implication that over-reliance on them will produce shoddy work. Hellsten et al.

(2006) pointed out that the quest to constantly retrieve the freshest document has detrimental effects for scholars and others who seek to do research through search engines, as semantic networks are degraded through the loss of older documents, versions, and links. Halavais (2009) considers that the position of academics on the one hand has improved by the freer circulation of knowledge and the ability to find authors of particular papers; but also that gold-standard approaches such as peer-reviewed search journals are less accessible online and may therefore decline in use.

1.2.5 SEARCH ENGINES AND SOCIETY

Although information retrieval specialists had been aware of the biases in search, it was Introna & Nissenbaum (2000) whose work first drew scholarly attention to the idea of search engine bias. Their concern for the quality of search engine results from a public perspective, rather than from an information retrieval perspective, was echoed by others including Hargittai (2000), Moshowitz & Kawaguchi (2002; , 2005), Goldman (2006) and Diaz (2008).

Law and policy approaches to the analysis of search engine developments have tended to concentrate on privacy and intellectual property issues. However, many other issues are also involved in the regulation of search engines (Aljifri & Sanchez Navarro, 2004; Gasser, 2006), especially issues of free speech and what does or does not qualify as protected speech. According to several reviews of case law, there may well be problems of "unaccountable discretion" (Grimmelman, 2007) with search engine results, but there is currently little recourse in the courts for individuals or companies who feel that they have been wrongly presented in these results. The legal and policy debates are presented further in Chapter 5, section 5.7.3, but it is worth pointing out that regulation varies by country, from relatively un-regulated as in the United States, to semi-regulated, as in Germany, to heavily regulated by the state, as in China (see Machill, Beiler, & Zenker, 2008, pp. 592-595). Concerns about censorship, particularly in China, and search engine participation in censoring have run high. In February 2006, the Yahoo and Google were summoned to Washington to account for their actions in censoring search and reveal users details to the Chinese, and were heavily criticized by members of the House Committee on International Relations. Various technical tools have been made available

to help end users understand when their results have been censored (see, for example, Meiss & Menczer, 2008)

Taken all together, these scholars have raised the point that we have, in search engine results, an extremely valuable element of the structure of internet content which is prone to a range of biases, some systemic and some affecting particular search results (e.g., for economically valuable search terms). This, in turn, raises concerns about the ability of ordinary users to find the websites they are seeking and for specialists to discover information of necessary quality. The technical explanation – that search engines have results that are "not deliberate" owing to the "self-organizing" structure of the Web – is arguably insufficient to explain the search engine bias technology scholars have observed.

One area of study that is notably absent in the literature reviewed above is any study of the production processes of search engines, apart from technical documentation of various algorithms. There are studies of programming, but not of programmers. There are studies of optimization, but not of optimizers. The algorithms powering the search engines are, of course, the product of particular engineers working in a particular environment at a particular time making particular choices, not all of which will be strictly technical. Further actions are then taken by others in the system such as website owners and marketers. If the bias of search engines is to be mitigated, whether by technological, economic or social means, we must understand more about why that bias occurs, and it is this gap in the literature that the present thesis addresses.

1.2.6 IDENTIFYING THE RESEARCH QUESTION

Within the search engine companies themselves, various groups seek to develop, gain control of, change and modify the search engine results. We might identify the search engineering group, the user interface group, the product management group and the paid listings group among these (these, and a range of other administrative units, are all a part of Google's organisational structure). However, it is not necessarily that simple. For example, most search engines operate internationally and tensions between different units can run high: in 1998, when I was working in Excite's European office, the director of product management was told by the US search engineering team that a hoped-for change could not be made because "Your entire traffic is less than our error pages." Owing to a

lack of study on search engine production processes, however, it is unclear which groups are relevant to the study of bias and where boundaries are to be drawn.

In general, the process which produces search engine results should be amenable to sociological studies just as are other types of "factual" production, as Latour (1987) has shown in his study of the production of scientific knowledge, or Tuchman (1973) in her study of news production. But this type of study, to this writer's knowledge, has yet to be undertaken with regard to these key players in the internet. In the literature reviewed for this thesis, none of the sources retrieved concerned the production processes within the search engine organisations. Thus, the industrial and commercial processes that lead to the observed bias in search engine results are poorly understood. Hence the central research question of this thesis is: *"Why does bias arise in search engine results?"*

The question of bias has a further implication when it comes to considering media in the information society. Traditional media are considered by many to be essential to a well-functioning democracy, and the justification of media regulation has to do primarily with the need for citizens to be able to make informed choices about public issues (Feintuck, 1999). To do this, it is argued, citizens need the right to freedom of expression, so that they can find out the 'truth' about events, learn to articulate their viewpoints, and participate in the public sphere in which, through informed debate, important choices are made about matters of public concern[3]. An additional argument has been made that media form an essential part of our shared culture, which should be denied to no one. The idea of the public service media therefore essentially rests on three key principles: universal access, diversity of sources, and quality of the media product (Feintuck, 1999, p. 73). Feintuck argues further that these principles were developed in an historical context in which the general threat to freedom of expression arose from a repressive state, whereas in the modern context, a greater concern for regulators should be limiting the ability of commercial interests to use "freedom of expression" as a justification for purely commercial purposes when in fact their actions constrict the provision of information; in other words, acting against the principles upon which the freedom of expression is based (Feintuck, 1999, pp. 13-14).

---

[3] Obviously, this is an idealized portrait of the functioning of democracy, which nevertheless continues to be at the heart of media regulation.

As the patterns of usage of television are changing and the internet becomes a more predominant source of information and culture for more and more people, it is arguable that the public interest criteria for regulation of media may need to be transferred or translated into the online space, since the public interest may not be best served by the commercial search services which exist and, as in other media, market forces may not sustain online media of sufficient quality. This thesis does not recommend policy solutions, but a recurring theme is the suggestion of the need for a public debate with regard to search engine results and the public interest in the way the search engine market is developing.

1.3 Investigating bias in internet search engines

Internet search engines are amongst the most complex software constructions of the modern era. The original search engines were used primarily on library collections, which were among the first large collections of data to be transferred to digital media. Library collections had several features that influenced the development of the earliest search engines. First, they were structured, inasmuch as every book or journal in the collection had a series of identical properties, e.g., title, author, year of publication, place of publication, etc. Second, they were authoritative: each library book had first been through a publication process and, secondly, had been selected by a librarian. Third, they were relatively stable, with few volumes being added or retired per year. Fourth, the collection was small enough to be viewed as a totality. The search engines that developed from these collections had several essential features that related to the nature of these collections and against which their success came to be measured. The search engine should ideally retrieve *all* and *only* those records that are related to the term the user entered and should display the complete record. Web search engines superficially resemble these early engines in that they display results in response to a user query, but the collection they operate on is fundamentally different: web pages and other web content are unstructured, unselective, massively large, constantly changing, and in many different languages. Initially, search algorithms producing excellent results on library collections resulted in incredibly slow searches on the Web, returning millions of out-of-date documents of which many were only tangentially related to the user query. The problem of creating a search engine powerful enough to search the whole Web reliably is

one not to be understated. The very complexity of the algorithms leads to potentially unforeseen consequences.

The technical challenge outlined above is exacerbated by the monetary value of search engine results. The large search engines are amongst the Web's most valuable businesses primarily because of their ability to send online businesses potential customers. In addition to direct advertising, which comprises the largest source of income for these companies, there is also a large trade in search engine marketing in which companies seek to improve their ranking in response to a particular query or set of queries on the part of the search engine user. Thus, the collection being searched by the search engines is also being altered and refreshed by many other users of the system.

Bias therefore may be introduced by a number of mechanisms, as discussed below.

1.3.1 The mechanisms of search Engine bias

On one level, the research question – *why is there bias in search engine results?* can be answered relatively simply with reference to the technical structure of the search engine and the behaviours observed in search engine producers and search engine optimisers (SEOs). As a prelude to the analysis in this chapter, this section reviews what has been learned about the mechanisms that produce search engine bias (for another review of biasing mechanisms, see Diaz, 2008).

In order to discuss biasing mechanisms, this section first reviews the technical structure of a search engines. Simplifying what is a very complex technology, a search engine can be divided into three elements: the index, built up from the journeys of the crawler or spider; the algorithm, which matches the user's query to the content in the index; and the display mechanism, which shows the user the results. Alongside this engine is typically an analogous advertising engine, in which the index is built up of paid entries, and matching algorithm and display may be configured slightly differently. This configuration is represented in diagrammatic form, in Figure 1.

Figure 1: Elements of a search engine
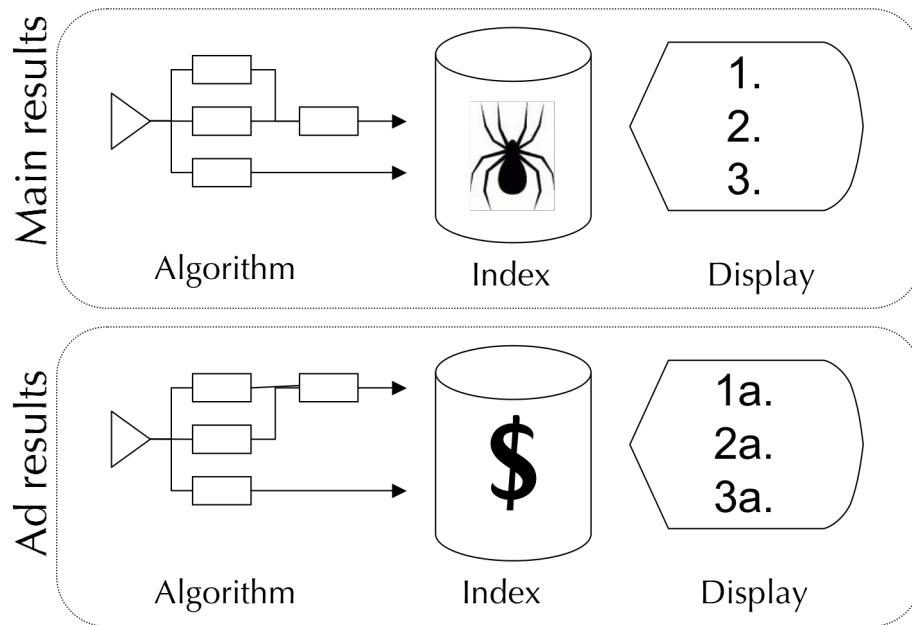
Source: Author.

Each of these elements of the search engine is subjected to pressures which cause it to unfairly represent certain classes of websites. The first form of bias, *indexical bias*, is linked to the index, which is the search engine's representation of the Web. In order to be found by a search engine, a Web page must be included in the index. Four main elements seem to affect inclusion or exclusion:

1. The seed list. When developing a new area of search – for example, for a new country – search engines use a "seed list" of Web sites from which the crawler begins its traversal of the Web. The contents of this seed list, which according to one informant consists typically of the conventional wisdom on the "important" sites in the area, have the potential to influence the results.

2. The crawling algorithm. Each search engine has a computer program which traverses the Web in search of new content. The structure of that crawl has implications for what is and is not included in the index (Cothey, 2004).

3. Index cleaning. Once the initial index is assembled it is processed at the search engine, where entries that do not meet various quality criteria are either masked or deleted. Chapter 7 explains some reasons why indices are cleaned.

4. Paid inclusion. Certain search engines, notably Yahoo, maintain a paid inclusion policy as described in Chapter 7, allowing advertisers to fast-track multiple URLs for inclusion in the index.

The second type of bias, *ranking* or *algorithmic bias*, refers to the search engine's propensity to rank certain sites higher than other sites, given that both sites are included in its index. Ranking algorithms are closely guarded trade secrets. However, modern ranking systems are based both on the Web page ranked (in other words its text, title, domain name and other elements) and also on its position within the Web in terms of inlinks and outlinks to other websites (see Langville & Meyer, 2006). Thus each site is assigned an "authority" value independent of the content of the Web page which then contributes to relevance and, therefore, ranking judgements. Interestingly, this is described in a Web video with two relevance engineers published by the Microsoft Development Network (Scoble, 2005). Chapter 6 deals with relevance decisions in some detail. Chapter 6 also provides anecdotal evidence of algorithm "tweaking" or "tuning" to match results more closely to those desired – for example, when current events cause a particular query to increase in popularity, the results for that query may be changed to represent sites that match that event. Chapter 7, on the other hand, deals with search engine marketers' (SEMs') efforts to match their websites to the ranking algorithm in order to ensure a high rank, through a variety of techniques including content changes, technical changes, domain name changes, co-operations with other websites for links, and even the development of whole networks of websites. These activities may be viewed as legitimate or illegitimate on the part of the search engine providers, but there is no doubt they affect the rank order of search results.

The third form of bias, *display bias*, refers to the visual preferencing of one site over another. Of course, ranking itself could be classified as display bias, but in general this type of bias would be shown through the use of larger fonts, colours, showing text for some results but not for others, etc. In practice, this type of bias seems restricted to the balance between 'natural' search results and advertising-generated results, or to the intermixing of advertising and non-advertising results in smaller search engines and meta-search services.

Some of these elements of bias in search engine results are under the control of search engines producers but others are outside their control. That bias exists is clear; that it has

a detrimental effect on users is arguably clear; but what is less explicable is the conviction within the search engine industry and elsewhere (see, for example, Goldman, 2006), that it simply *is not an issue*. In effect, this may be said to be the underlying contributor to search engine bias: the widespread agreement that there is no problem and that, therefore, no action need be taken.

1.3.2 A FRAMEWORK FOR INVESTIGATING BIAS

Designing a theoretical framework and a method for investigating bias in this complex environment – one in which bias may be an unintended consequence, a designed-in feature, or a result of commercial activity, or all three – was a central issue in this thesis. The conceptual framework, outlined in Chapter 2, draws from a range of different traditions in order to address this complexity. The creation of a technological artifact as a social *practice* is the first element. A technological artifact, from this point of view, is not static but rather a shifting cluster of meanings realized out of the interactions of individuals within groups and also by the interactions of groups with each other. Of course, one of the features of an artifact is that it has some physical existence, and may be passed from one group to another. The cluster of meanings which it enacts – the *technology-in-practice*, to use Orlikowski's (2000) term – may be very different from group to group and, in this way, the artifact also functions as what Star (1989) calls a *boundary object*. The practice perspective in the social study of technology therefore calls attention to two main observable activities: the activity of humans constructing meanings and defining group characteristics through language, and the activity of interactions with the artifact. Wenger and Lave (1998) call this the *duality of participation and reification.*

The strength of the practice perspective in the conceptual framework for this study lies in its grounded nature which deals with observable phenomena in small groups. It is, however, less useful in investigating the commercial aspects of potential search engine bias or potential power asymmetries between groups. The historical and commercial context in which the practices described above take place, form the second major element of the conceptual framework. Here, the perspective is one in which the search engine is first and foremost conceived as a media business and insights from the theoretical tradition of the *political economy of communications* are used to structure this investigation. The thesis focuses on two central processes identified within the political economy

tradition: commodification and spatialisation (Mosco, 1996). *Commodification* refers to the way in which cultural resources become part of the logic of capital – in other words, how previously free products (such as information or entertainment) are created as goods that can be bought and sold. Of particular interest is the dynamic of advertising, where cultural resources appear as free to the public. It is possible in this regard, for example, that advertisers might become de-facto censors of cultural content and thereby contribute to search engine bias. The second process, *spatialisation*, refers to the ways in which media both occupy space physically and construct it for viewers, listeners, or users. Geography may play a role in search engine bias (cf. Vaughan & Thelwall, 2004) and the location and structure of search engine companies in relation to physical space is examined to shed light on this possibility.

The final element of the in the conceptual framework which is presented in Chapter 2 is the way in which the theoretical perspectives drawn from the social construction of technology tradition and the political economy tradition can be related to inform the issues at the heart of this thesis. Structuration theory, as proposed by Anthony Giddens (1984), suggests that practices are related to larger structures through at least three modes of interaction: interpretative schemes, facilities, and norms. In a dynamic duality, the actions of agents are said to either reinforce or subvert existing structures which, in turn, cohere into larger systems. The agent is theoretically free to act outside existing structures while, at the same time, remaining to some extent constrained by them. Equally, structures are theoretically relatively stable while, at the same time, the possibility of their disappearance or alteration always remains. Thus, the central research question for the study is examined through the lens of a conceptual framework which integrates the practice perspective with processes of commercialization and spatialisation, using structuration theory as a bridge.

This thesis makes a number of contributions to the growing field of internet studies and to research in the field the social studies of technology. It breaks new ground by investigating the search engine company as a *media* business, examining the industry structures and business models, the way in which search engines have been concentrating both in terms of location and in terms of ownership, and the transformation of advertising and commodification of web content that search advertising is bringing about.

It also yields an insight into the interpretative schemes of search engine producers, that is to say, the way in which search engine quality is conceptualized and framed. The results are used to extend the conceptualization of technological structuration as developed by Orlikowski by extending the analysis of conflict and power which is latent in her work. This is achieved by drawing on methods from the tradition of political economy of media and communications to reveal insights into search engines as a "technology-in-practice."

The first component of the empirical work is the presentation of the history of the search engine as a business which is the focus of Chapter 4, dealing with issues of commodification. Based on an analysis of company documents, press releases, and contemporary press accounts, the history of Web search engines is divided into three periods. The first period, from 1994-1997, was one in which search engines were developed in the pre-existing Silicon Valley/venture capital model. This period saw several search engines launch onto the stock market. The second period, from 1997-2002, saw both the internet boom and the dot-com crash. During this period, search engines struggled to find a funding model, transforming into "portals" and in many cases being acquired by large traditional media or telecommunication corporations. Nearly all the acquired engines were closed during the dot-com bust of 2000-2001. The third period, from 2002 until the time of writing in 2009, is characterized by a continuing trend towards consolidation funded by a sophisticated syndicated advertising market. This chapter analyses the changes in the economic model of search engines that accompany these three periods and sheds light on the current industry structure.

The analysis of the Web search engine industry is given a geographic dimension in Chapter 5, forming the second component of the empirical research, where issues of spatialisation are at the fore. The chapter examines search provision in four different countries (Japan, China, Germany and South Africa). Japan, China, and Germany had the largest internet population after the United States (US) at the time of the empirical research, while South Africa is an example of a much smaller market. The analysis of company documentation, ratings data, and press reports suggests the large American search engines were active in all of the countries at the time of writing. In China, where the advertising market is much weaker and an alternative funding model was needed, a local company, Baidu, has been able to compete with Google and Yahoo. In Japan,

Yahoo was hugely dominant; in Germany, Google was used for nearly all search queries. In South Africa, a two-tier model seemed to have developed, with local search engines being used for specific local queries while US search engines are used to search the general web. In addition, this chapter reviews the geographic distribution of the search engine production and marketing offices of the large US engines, and considers the impact of local regulatory regimes.

The third component of the empirical investigation is presented in Chapter 6 which is based on a series of interviews with senior search engine personnel. The focus of the chapter is on discursive practice and specifically on the construction of the idea of quality within the discourse of search engine personnel. What the interviews appeared to reveal were a set of related discourses which I call *technological schemas*. These are interconnected ways of situating the technological artifact – in this case the search engine – in a pre-existing web of meanings. The most evident technological schema was the "market" schema, in which the search engine's identity was that of a product, and the way in which quality appeared to be judged was based on "customer satisfaction." Within this schema, the speaker seemed to self-identify as an employee with a bounded ability to effect change. The second major schema, highly intertwined with the first, was identified as the "science/technology" schema. Here, the search engine had an identity more akin to a feat of engineering. The major quality criterion, "relevance," was borrowed from the scientific discourse of the earliest information retrieval systems. Interviewees using this schema seemed to position themselves as experts, fully capable of modifying the search engine. One interesting aspect of these two schemas was the way in which interviewees appeared to move between them depending upon their rhetorical ends, and the way in which they seemed to be used to bolster each other, with relevance equating to customer satisfaction and vice versa. Finally, a minor schema was identified as the "war" schema, in which the search engine seemed to be a battleground and the identity of the speaker seemed to be that of competitor with honoured opponents (other search engine companies) or guardian with actual enemies (spammers).

Chapter 7 presents the fourth and final component of the empirical work. The practice perspective is employed again in this chapter to examine the interaction of search engine marketers (SEMs) with search engine companies and with the search engine technology.

Contrary to the rhetorical positioning revealed in Chapter 6, the relationship between SEMs and search engine companies is shown to be complex and often supportive rather than conflicting. Although SEM tactics can be labeled "black hat" and "white hat" by members of the community to distinguish between legitimate activities (i.e., search engine company approved) and illegitimate activities (i.e., prohibited by the search engine companies), the boundaries between these activities seem to be unclear in practice. Some search engine customers, who advertise on the search engines pages, are also the same marketers who work to ensure their websites come to the top of the results through optimization. Again, these are often, also, the very same people involved in what the search engines portray as "spam" or "fraudulent" practices. Thus, the search engine companies may be providing tools and assistance to the same people they are involved in lawsuits with. The search engine results, serving as a boundary object, form a way in which these communities tacitly negotiate their relationships in their daily practices. Most overt conflicts seemed to concern the allocation of funds between the two groups, while overall, the analysis of these relationships suggests that there is agreement on the legitimacy of the Web as a marketing medium among the different groups.

In the concluding chapter, I suggest augmenting the initial conceptual framework with several elements. The first supplementary concept and observation that is of use is that rules and resources are mutually constitutive. This derives from the work of the anthropologist Sewell (1992). Sewell divides Giddens's modes of interaction into "rules" (interpretative scheme and norms) and "resources" (facilities), and argues that rules function to create resources. The second additional concept is developed from the insight that within capitalist societies technologically-enabled growth spurts, or techno-economic paradigms, are recurrent and that these contain both new rules (a new "common sense" of how business should be done) and new resources. The chapter then links the "common sense" of Perez (Perez, 2002) with the "interpretative schemes" of Giddens, so that taking these two points together, it appears that the interpretative scheme associated with the existing techno-economic paradigm may be a particularly powerful one because of its ability to constitute many new and previously overlooked resources. I suggest that interpretative schemes and by inference technological schemas are not stand-alone constructs, but rather embedded into local cultural complexes. In the case of the search engine, the local culture may be identified with the Silicon Valley or "internet culture" that

Castells (2001) describes, in which technological development is strongly intertwined with money-making and personal virtue, and communal values are expressed through the creation of free networks of association which form about a variety of interest topics, rather than being associated with particular ideologies or pre-existing institutions.

The implication of these observations is that search engine bias (and the relative lack of concern by both industry and academic writers) can be attributed to the construction of search as a solely, or primarily, commercial service. Nor is search simply any commercial service, but it is one which embodies to some extent the idea that technological quality and public virtue are exemplified by, and only by, commercial success. The exigencies and opportunities of the market, in turn, lead to an industrial situation in which search engine companies are concentrated industrially and extensive geographically; in which search algorithms are opaque to protect against competition and against the search marketer; and in which search quality can be reduced to issues of "relevance" and "customer satisfaction." Against this backdrop the concept of bias and its explanation may seem both futile and naïve. Yet, by using "bias" as a lever to open up the discussion and examine the implied tautology of success=quality, this thesis offers insights that are likely to engage with the concerns of academics, engineers, and policymakers in ways that will take debates about the meaning and implication of search engine 'quality' further.

1.4 Conclusion

To summarise for the reader's convenience: Chapter 2 sets out the theoretical underpinnings and conceptual framework of the research based on concepts drawn from the social construction of technology and from the political economy of media and communications, which are then linked building insights from structuration theory. Chapter 3 describes how the conceptual framework was operationalised and the research design and methodology for the study developed. It also discusses how the interviews, participant observation, and documentary analysis that form the evidentiary basis for the thesis were conducted. Chapter 4 reviews the historical development of the search engine industry and draws attention to the importance of online audience traffic to the new navigational media. Chapter 5 examines the development of search engines outside the United States, assessing the "global" qualities of search engines and how they operate

within larger capitalist structures. Chapter 6 is based on interviews with search engine producers and focuses on how they account for, and justify, quality-related changes to the search engine algorithm based on the ideas of relevance and customer satisfaction. Chapter 7 analyses the relationship between the search engine optimizer and the search engine production companies and discusses how they co-operate as well as compete, using the search engine results themselves as a co-ordination device. Chapter 8, the concluding chapter, also considers these empirical investigations in the light of the conceptual framework, emphasizing the importance of cultural factors in shaping economic resources as well as vice versa. It also outlines the theoretical and methodological contributions and limitations and suggests new directions for research.

II

The Dynamics of Technological Structuration

A Conceptual Framework for Understanding Search Engine Bias

The view that technology just changes, either following science or of its own accord, promotes a passive attitude towards technological change. It focuses our minds on how to adapt to technological change, not on how to shape it. It removes a vital aspect of how we live from the sphere of public discussion, choice, and politics. (MacKenzie & Wajcman, 1999, p. 5)

2.1 Introduction

The research question of this thesis is "Why does bias arise in search engine results?" Chapter 1 explained the choice of the subject matter for this thesis, suggesting that search engines are uniquely important agents in the content structure of the internet. This chapter develops the conceptual framework which both guides the research design and

methodology for the thesis presented in the next chapter and structures the analysis of the empirical research presented in Chapters 4-7.

The concept of bias is at the core of the research question and it forms the central axis of this chapter. As discussed in Chapter 1, Section 1.2, search engine bias can be both systemic or particular to a few results, but in any case it affects both ordinary user's ability to find the site they are seeking as well as their ability to be "heard" online if they themselves publish content online. "Bias" generally means the slanting or even misrepresentation of fact, so it carries negative connotations of misleading or even reprehensible behaviour on the part of those responsible. No one, to my knowledge, has suggested that search engine companies, in producing the results they display in response to users' queries, knowingly behave either reprehensibly or in a misleading fashion. Nonetheless, several academic studies from a computer science perspective have referred to search engine results as "biased" (Lawrence & Giles, 1999; Mowshowitz & Kawaguchi, 2002; Vaughan & Thelwall, 2004). In the first part of the chapter, I examine what bias means in the context of search engines, and reframe the issue of bias as being one of conflict.

By seeing bias as the expression of a conflict, we are led to consider the interest groups and processes involved in that conflict. The theoretical perspective that is presented here is based on two key strands. The first is the **social construction of technology**, particularly the idea of *technology-in-practice*, which conceives of technology and social processes as intertwined at a community, everyday level. The second is derived from the **political economy of media and communications** tradition which stresses the location of media products and practices within larger historical, social and economic processes, particularly capitalism. The two levels of analysis are related by developing a relational or dialectical approach to **technological structuration**, drawing on Anthony Giddens's structuration theory, which theorises technology as key to both constraining and enabling action, and on Wanda Orlikowski's extension of this theory into the technological domain.

This chapter proposes that the bias in search engine results, rather than being understood as a mainly technical issue, is best understood interrogated as a dynamic process of

interaction between and within the search engine industry as it is developing under capitalism, the engineers who produce the results, and search engine users of various types, including those – like search engine optimisers – who may exploit the engines for their own ends. Based on the conceptual framework presented in Section 2.5, I suggest that it is not inevitable that current search engines produce the particular biases they do, but rather that the process is both historically influenced and contingent. Rather, I argue that the bias in search engines can be seen as a conflict about what *should* come at the top of the search results screen, rather than being necessarily an intentional distortion of reality.

The epistemological stance adopted in this thesis does not begin from the deterministic premise that technology has "effects" which are inbuilt, or that the capitalist system will always develop particular kinds of technologies. Nor does it propose an instrumentalist view of technology, in which technology is simply regarded as a tool in the hands of the elite. Instead, it offers a dialectic framework in which technology (shaped as it is by capitalist relations) can also be co-opted and used as a means of resistance.

The structure of the chapter is as follows: it begins in Section 2.2 by discussing bias and developing a reformulation of the concept as a *normative conflict,* stressing the importance of different concepts of 'relevance' in developing search engines. Next, in Section 2.3, it moves on to consider the insights from the sociology of technology, where the resolution of conflict and the stability of socio-technical systems are key issues, as expressed in the concepts of closure and stabilization. In Section 2.4, a framework for analyzing larger social processes is provided by reviewing selected work in the political economy tradition, specifically as it has been applied to communications and the internet. Section 2.5 presents the conceptual framework and discusses other potential avenues.

2.2 Bias as a normative conflict

What kinds of bias do search engines display and why does it matter? According to researchers in the field computer science, search engines over-represent in their results sites that are more popular (that is, have more inbound and outbound links), sites that are older, sites that are based in the United States, and sites that are commercial (Lawrence & Giles, 1999; Mowshowitz & Kawaguchi, 2002; Vaughan & Thelwall, 2004). Why are

these biases important? Essentially, the argument is about the diversity of content accessible *in practice* online versus in theory. In one large-scale study based on crawling[4] results from the top search engines on a range of topics finds that in every category, a few large sites dominate linkage patterns, creating a "Googlearchy" of large and prominent sites to which users will most likely be funnelled (Hindman, Tsioutsiouliklis, & Johnson, 2003). Concerns about the contents of the search engines' first few pages are exacerbated by literacy issues: most users use short, unsophisticated queries and do not search beyond the first page. Many others are unaware of the distinction between paid-for search results (advertisements) and unsponsored results which come from the search engine algorithm (Hargittai, 2002; Hsieh-Yee, 2001; Machill, Neuberger, Schweiger, & Wirth, 2004; S. Ozmutlu, Spink, & Ozmutlu, 2004; Spink, Jansen, Wolfram, & Saracevic, 2002; Spink, Wolfram, Jansen, & Saracevic, 2001). Thus, search engine results pages, and particularly the first page of results, are important gateways to the information on the Web. As a result the observed bias of their results gives some scholars cause for concern (Introna & Nissenbaum, 2000; Machill, Neuberger, & Schindler, 2003).

The computer science researchers who have identified bias in search engine results have treated the evidence of bias primarily as a technical error, which requires a technical solution. In this way, they consider it to be different from the "bias" that we might associate with other forms of media (for example, news content). As one study puts it:

> "Detecting bias in an information retrieval system is different from analysing the content of a message. A retrieval system contains a set of items (typically title, citations, or brief subject descriptions) that represent messages, rather than the messages themselves. Bias is exhibited in the selection of items, rather than in the content of any particular message. The former may be termed *indexical bias*; the latter, *content bias*." (Mowshowitz & Kawaguchi, 2002, p. 143)

While indexical bias may be distinct from content bias, we can say that the assessment of both types of bias implies the comparison of messages or index items against a norm of what a perfect or ideal system ought to have. The norm is implied in the term bias itself, as is shown in the dictionary definition of statistical bias: "A systematic distortion of an expected statistical result due to a factor not allowed for in its derivation; also, a tendency

---

[4] "Crawling" is a term which describes the operation of the algorithm that search engines use to discover and record content such as text and images that are hosted on the World-Wide Web.

to produce such distortion." (Oxford English Dictionary, 1989). Within media and communications studies and particularly within journalism studies, bias has been the focus of extensive research (for example, Breed, 1999; Gieber, 1999; Herman & Chomsky, 1994; White, 1999). The norm against which news is generally judged is *objectivity*, expressed both in factual accuracy and the reporting of conflicting sides of any story. However, the concept of objectivity has tended to fade as an analytical subject, as the possibility of objective media representation has been called increasingly into question by postmodern thinkers. Instead, it has been argued that the news is socially constructed (Schudson, 1989; Tuchman, 1978), that the ideal of objectivity is a false dogma (although see Lichtenberg, 2000), and that no story could ever really represent all the facts and all the points of view. Interpretation on the part of the journalist is said to be inevitable. Is search engine bias, by analogy, also inevitable?

Let us imagine some kind of perfectly "objective" search engine. We conceive first of a universe of Web sites that is 'out there', though perhaps empirically difficult to catalogue. Second, we picture a computer algorithm that would match all the sites with given search criteria and then return the site most relevant to that search in order. So far so good. An interview at the Search Engine Blog website with Cindy McCaffrey, VP of Marketing at Google, suggests that staff at Google would argue they do just that, to the best of their abilities:

> "The future direction is to continue enhancing our search technology, and to provide innovative features and services that efficiently connect people to the information they're seeking…. We believe this will result in a search experience that will continue to surprise and delight users around the world. ..." (Da Vanzo, n.d.)

Still, it is always the case that one site must come top of the list in a search results page, and that other sites must be lower down the list. This ranking must also be in some sense subjective. If the user of this hypothetical engine searches for "apple," are they looking for the fruit or the computer brand? And if the fruit, do they want recipes, an apple seller near them, or a botanical description? Which is the best result for our 'unbiased' search engine to return? Each site seems to have a claim. And indeed, the debate within the search engine companies themselves is not cast in terms of objectivity or bias, but in terms of *relevance*, or the appropriateness of the search engine results to the initial query,

relative to a given intended audience (which may change, for example, across geographic boundaries). Thus the quest to build a "relevant" search engine implies a dynamism and flexibility across both time and space. It is arguably a balancing of different communities' needs for relevant results and expectations of relevant content. Objectivity and bias, in this context, are very slippery concepts.

Nonetheless, despite its slipperiness, its critique within the academic literature and its lack of use within the search engine community, this thesis employs the term bias as a central organising construct. It is a useful concept because it turns our attention to the choices that are being made in the name of 'good' and 'relevant' content and to issues of representation and fairness. All kinds of classification schemes have certain principles to enable their information to be sorted and, in some cases, ranked. For example, Bowker and Star examined a range of classification schemes and concluded that they are "significant sites of political and ethical work" (Bowker & Leigh Star, 2002, p. 319), precisely because they often appear to be societal givens and questions of objectivity or fairness seem irrelevant.

It seems that the search engine companies themselves recognize that the technological decisions they are making have ethical and moral consequences. For example, it was revealed at a recruiting session for Google in 2004 that Google has an internal ethics committee – although the exact function is unknown, the source mentioned the committee in connection with Google's ranking algorithm, PageRank (Orlowski, 2004). Occasionally search engine companies find themselves at the centre of a controversy, or at fault in a legal challenge where the moral or ethical nature of their results become central issues. One example from Google was the persistent place at the top of the ranking for the query term "Jew" of a hate site. Google issued the following disclaimer at that time:

> "Our search results are generated completely objectively and are independent of the beliefs and preferences of those who work at Google. Some people concerned about this issue have created online petitions to encourage us to remove particular links or otherwise adjust search results. Because of our objective and automated ranking system, Google cannot be influenced by these petitions. The only sites we omit are those we are legally compelled to remove or those maliciously attempting to manipulate our results." (Google, 2004a)

Here we see objectivity being called on as a defence of search results that, clearly, are more relevant for some than for others. Other commentators noted that the hate site, intentionally or not, had managed some very effective "search engine optimization" (Finkelstein, 2004). In other words, they had, through the use of content and links, made it exceptionally easy for Google to find their site.

Often, search engine optimisation is done for a fee by professional marketing organisations. In the same interview cited earlier, Google VP Cindy McCaffrey goes on to acknowledge that commercial considerations may play into Google's answer about which sites come first:

> "As long as there are people who need help promoting their sites on the web there will be a legitimate role for search engine marketers. The bottom line is we're all part of the web and our intent is to work with webmasters, search engine marketers and the like to ensure that the quality of Google search results continue to be the best on the web."
> (Da Vanzo, n.d.)

Whether "we're all part of the web" or not, different groups have different ideas about what should come first, second, and third when it comes to search listings. The story of the anti-semitic hate site continues, for example, with both anti-defamation and neo-Nazi activists attempting to "Google bomb"[5] the search engine in order to force the hate site down or up in the rankings as they wished.

The concept of bias, therefore, leads us to wider consideration both of the 'objective' nature (or not) of the technology andof the power relations and conflicts between different groups that have the search engine results as their focus. Who succeeds in getting their ideas into practice and how that is done, are the empirical subjects of this research. Thus the present study is concerned with examining search engine technology as an ongoing locus of this kind of normative conflict in which many actors, both large and small, have a role to play. The following section draws on insights from studies of science and technology to locate this issue of conflict in the broader context of technology use and development.

---

[5] Because Google's ranking is, in part, determined by the number of links to a given page, as well as the text of those links, it is possible, with enough links, to cause that page to appear highly as the result of a search term which in fact has little to do with the page in question. Thus "miserable failure" resulted in George W Bush's autobiography, at least for a time, due to the efforts of some link-activists.

2.3 Technology as a social practice

Search engines are feats of software engineering, among other things, and as discussed most of those who have identified search engine results as biased have put the existence of bias down to technical issues. But of course search engines do not program themselves, so this section turns to the social study of technology for help in developing a theoretical framework. Bearing in mind that the last section discussed bias as a by-product of a normative conflict, this section focuses on how these theories treat conflict.

It is a premise of the social study of technology that the various technologies we see around us must be explained with reference to social processes and social actors. This method of investigating, focusing on the historical, the contingent, and the social, challenges an earlier paradigm that treated technology as a "black box," following a trajectory which was best analysed by engineering rather than social science methodologies.[6] This tradition focuses principally on how technology overcomes conflict to achieve closure, or stabilisation, around particular inventions or products.

Pinch and Bijker, in an influential essay, laid out the framework for the SCOT (social construction of technology) approach (Pinch & Bijker, 1989), one of several within this tradition. Their interest was in the innovation process: how does an invention come to be accepted? They used the bicycle as an example: from the huge range of bicycles available in the late 19th century, including the pennyfarthing bicycle with its huge front wheel and tiny back wheel, how did the "safety" bicycle design we now use become the norm? They argued that technology, when introduced, is in a "destabilised" phase. Different groups with an interest, called "relevant social groups" are at this stage in conflict about what is central to the design and how to solve technological issues. A key concept here is *interpretative flexibility,* which denotes that different groups give a different account of the meaning or causes of technological issues. In the bicycle example, Pinch and Bijker argued that speed racers perceived the key limitation of the bicycle as its speed, while women riders, particularly, were concerned about its safety. This led to a range of reactions to technical issues, as each group perceived them differently. This, in turn, led

---

[6] This paradigm was particularly evident in economic studies of technology, where technology and technical change, while recognised as vital to economic growth, [e.g., see Rosenberg, 1982, 1994; Dosi et al 1988] were treated as independent variables.

to different variations in the technology in response to the perceived issues. Some manufacturers redesigned bicycles for speed, some for safety. Finally, one of these variations was selected through a *closure* mechanism (discussed in more detail below). Thus, it was argued that technological progress operates through a mechanism of first variation and then selection.

Scholars adopting this approach generally reject a progressive, determinist view of technology. They recast the history of technology as "multidirectional" rather than linear: "[O]ur argument [is] that the the 'successful' stages in the development are not the only possible ones" (Pinch & Bijker, 1989, p. 28). Successful innovation and invention are analysed as cultural achievements, instead of being the result of some "superior" technology.

How is stability maintained and closure achieved in this environment – or in other words, how is conflict resolved or "success" selected? Pinch and Bijker suggest two strategies of closure: rhetorical closure and closure by redefinition of the problem. Because they define the controversy as one of different meanings, *rhetorical closure* or persuasion may be sufficient: "[O]ne need not *solve* the problems in the common sense of that word. The key point is whether the relevant social groups *see* the problem as being solved." (Pinch & Bijker, 1989, p. 44, italics original). As a consequence of this, they suggest that advertising, for example, may be a legitimate closure strategy. *Redefinition of the problem*, a second strategy, also seems to have a marketing twist to it – by defining a different need, to which the technology deemed to be the solution, a nascent controversy can be silenced.

Rhetorical closure and redefinition are not the only mechanisms for closure. In another work, Bijker mentions closure of a controversy over the development of plastics by legal means through patent adjudication (the patent awarded to the manufacturers of Celluloid) and also through the development of a community of "Celluloid chemists." Later controversies regarding the Bakelite type of plastic were resolved first in the patent courts and second by incorporating the losing patent litigants into the Bakelite corporation and thus into the community of Bakelite producers which was "in the beginning almost totally congruent with the employees of the Bakelite Corporation." (Bijker, 1989, p. 176). Finally, even greater stability was achieved through finding powerful customers in the

radio and automobile industries and in industrial design. This process of closure is known as *enrolment* (Callon, 1987; Law, 1992).

Helpfully, then, these scholars characterise technology as a sociotechnical system, in which social, natural, and technical forces interact within the contingent development of any particular innovation[7]. Nevertheless, two significant critiques may be made of this perspective that suggest the needs to its extension if it is to provide a framework for the current study. First, it examines technologies that are artefactual in nature – bicycles, Bakelite, and bulbs, to quote the title of one of Bijker's works (Bijker, 1995); these are quite different in some respects to the kinds of technology that search engines represent, as discussed in the next section. Second, the analysis of the larger forces of capitalism which help to shape technology – indicated in the examples above by the foundation of companies, patent suits, the enrolment of customers, and the use of advertising – is not undertaken, as some scholars have pointed out (Klein & Kleinman, 2002; Winner, 1993). The rest of the chapter develops these observations in more detail and suggests how this theoretical framework can be augmented: first, by drawing on the *technologies-in-practice* conception as developed by Orlikowski and supplemented by Lave and Wenger's communities of practice approach; and second, by incorporating insights from the political economy of media and communications field as complementary analytical constructs.

2.3.1 TECHNOLOGIES-IN-PRACTICE: A CRITIQUE OF CLOSURE AND STABILISATION

Unlike many studies of technology, the current study is not primarily concerned with the process of innovation, whereby an artefact goes through a range of early forms to yield an eventual commercial success (or, more rarely in the literature, a failure). Nor is it concerned with an adoption process in which users take a fixed technological product and use, discard, or adapt it to their ends. Instead, it is concerned with an information technology *service*, involving an ongoing innovation, adoption and use process in which the idea of "stabilization" or "closure" remains open. SCOT in particular has been criticized

---

[7] This is a short overview and the social construction of technology tradition has generated a large body of empirical and theoretical research that deserves fuller treatment. There are excellent collections available to the interested reader (see for example Bijker, Hughes, & Pinch, 1989; Law & Hassard, 1999; MacKenzie & Wajcman, 1999).

for its emphasis on consensus and resolution as the usual endpoint of technological controversy (Hård, 1993).

The challenge in this study is to provide a picture of technology as a continuing and developing service, rather than as a material physical artefact. Designers, users, and a range of actors interact continuously through this service. Processes revealed by studies of technology devoted to innovation or adoption may therefore be shown to be fully present, partially present, or not at all present. This study investigates specifically how conflict over technology is managed on an *ongoing* basis with a technologically-based information service. Therefore, while accepting that rhetorical strategies, redefinitions of the problem and enrollment may take place, I want to leave open the idea that they necessarily lead to closure. To help understand technology as an ongoing practice, I turn to a theory proposed by Wanda Orlikowski (Orlikowski, 1992, 2000; Orlikowski & Gash, 1994), who links a social constructivist framework to the theory of structuration, as developed by Giddens (Giddens, 1984).

Orlikowski set out a "model for analyzing the nature and role of technology in organizations" (Orlikowski, 1992, p. 398) which she called the "structurational model of technology". She later expanded the model (Orlikowski, 2000) to develop a more sensitive idea of user interactions which she called the 'enactment' model or the *technology-in-practice* model. She reframes the role of technology "in terms of a mutual interaction between human agents and technology, and hence as both structural and socially constructed" (p. 403) . This model of technological structuration, as elaborated below, offers the groundwork for understanding technology at a micro-level in small groups and emphasises the relationship between the broad contextual level and the community level of technological practice.

The model of technological structuration is based on Giddens's structuration theory. Structuration theory elides the dichotomy between structures and human agents and, instead, concentrates on the reciprocal nature of structure and agency, emphasising the knowledgeable and reflective nature of actors who draw on structural properties (structure) in their habitual and ongoing interactions (agency) and, in so doing, reinforce structure (Giddens, 1984). Three such reciprocal relationships are highlighted. First,

between human meanings and interpretative schemes and institutionalized structures of signification. Second, between human power over both authoritative and allocative resources and institutionalized structures of domination. Third, between human norms and institutionalised structures of legitimation, including legal systems but also rituals, socialization practices, and tradition. Giddens expresses these relations through the diagram given as Figure 2:

Figure 2: The three reciprocal relations of the structuration process
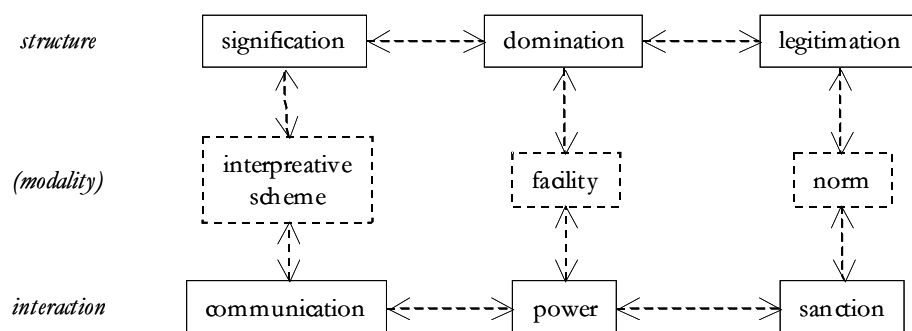
Source: Giddens (1984, p. 29).

How does technology come into these dynamics? Orlikowski bases her answer on two principles. The first principle is the *duality of technology*, which reflects on the recursive nature of human-technological interaction:

> "[T]echnology is physically constructed by actors working in a given social context, and technology is socially constructed by actors through the different meanings they attach to it and the various features they emphasize and use. However, it is also the case that once developed and deployed, technology tends to become reified and institutionalized, losing its connection with the human agents that constructed it or gave it meaning, and it appears to be part of the objective, structural properties of the organization." (Orlikowski, 1992, p. 406).

The second principle is that of *interpretative flexibility*, drawn from the work of Pinch and Bijker (1989), which Orlikowski defines as "the degree to which users of a technology are engaged in its constitution (physically and/or socially) during development or use." (1992, p. 409).

How do these principles relate to the three dynamics of structuration she draws from Giddens? First, Orlikowski stresses that technology is the product of human action.

Thus, in design, "human agents build into technology certain interpretative schemes…, certain facilities…, and certain norms…" (Orlikowski, 1992, p. 410); and in use, "human agents appropriate technology by assigning shared meanings to it, which influence their appropriation of the interpretative schemes, facilities, and norms designed into the technology…" (Orlikowski, 1992, p. 410). Therefore, in the first instance, humans are able to both design structures into and interpret structures out of technology. Secondly, Orlikowski says that technology is a medium of human action. By this she appears to mean that it both facilitates and constrains action – in other words, it has an 'impact' but that 'impact' is moderated by human agency, by the possibility of choosing to not use it or to use it otherwise. Thus, she acknowledges that technology makes certain kinds of social action easier and certain kinds more difficult.

Then, she describes two further 'influences' that she says are also particularly relevant: the institutional conditions and institutional consequences of interaction with technology. The first is explained as follows: "When acting on technology (whether designing, appropriating, modifying, or even resisting it), human agents are influenced by the institutional properties of their setting. They draw on existing stocks of knowledge, resources, and norms to perform their work. Often these influences are unarticulated, or reflected on only fleetingly…" (1992:411). The second concerns the way in which the use of technology itself acts upon an organisation's institutional properties. "Technology," says Orlikowski, citing Weick, "is an 'enacted environment' … whose construction and use is conditioned by an organization's structures of signification, domination, and legitimation. The appropriation and use of technology implies the change or reinforcement of these three institutional structures." (1992:411).

In other work, Orlikowski and Gash (1994) develop the related concept of the "technological frame of reference" which they define as "that subset of members' organisational frames that concern the assumptions, expectations, and knowledge they use to understand technology in organizations. This includes not only the nature and role of the technology, but the specific conditions, applications, and consequences of that technology in particular contexts" (p. 178). They link perspective this to Giddens's "interpretative scheme." A related concept developed by Bijker at a similar time, the "technological frame," is conceived more broadly as "all elements that influence the

interactions within relevant social groups and lead to the attribution of meanings to technical artefacts – and thus to constituting technology" (Bijker, 1995, p. 123). Other commentators offer helpful extensions of the concept. Klein and Kleinman emphasise the historical continuity of technological frames, arguing that they "are not likely to emerge de novo in the case of each new technology. Instead, they are likely to draw on cultural elements with historical resonances in the society at large or at least resonance among similarly socially located actors" (Klein & Kleinman, 2002, p. 40). Davidson and Pai suggest that these historically-situated technological frames can form part of the structures of domination, as "interpretative power is brought to bear when dominant frames form the basis from which others develop their understanding of technology" (Davidson & Pai, 2004, p. 482).

Orlikowski (2000) acknowledges the debt her work owes to theories of the social construction of technology. However, she suggests that this conceptual debt has led to some difficulties in applying structuration theory to the analysis of technology, in two major ways. These are "that technologies become 'stabilized' after development; and that they 'embody' structures which (re)present various social rules and political interests" (Orlikowski, 2000:405). These problems echo the difficulties highlighted earlier of applying an artefactual conception of technology to technologies with media-like properties. This conception leads to structure being reified as something more or less unchanging that is "embedded" or "inscribed" into technology by designers and thereafter "appropriated" by users.

Orlikowski proposes what she calls a "practice lens" for studying technology. In this, she emphasises the recursive and repetitive nature of technology use as the principle way in which structures become enacted, based on the properties of the technological artefact and the various capabilities, interpretative schemes, and norms of the users and of their situation. She modifies Giddens's diagram as shown in Figure 3 to explain this:

Figure 3: The enactment of technologies-in-practice
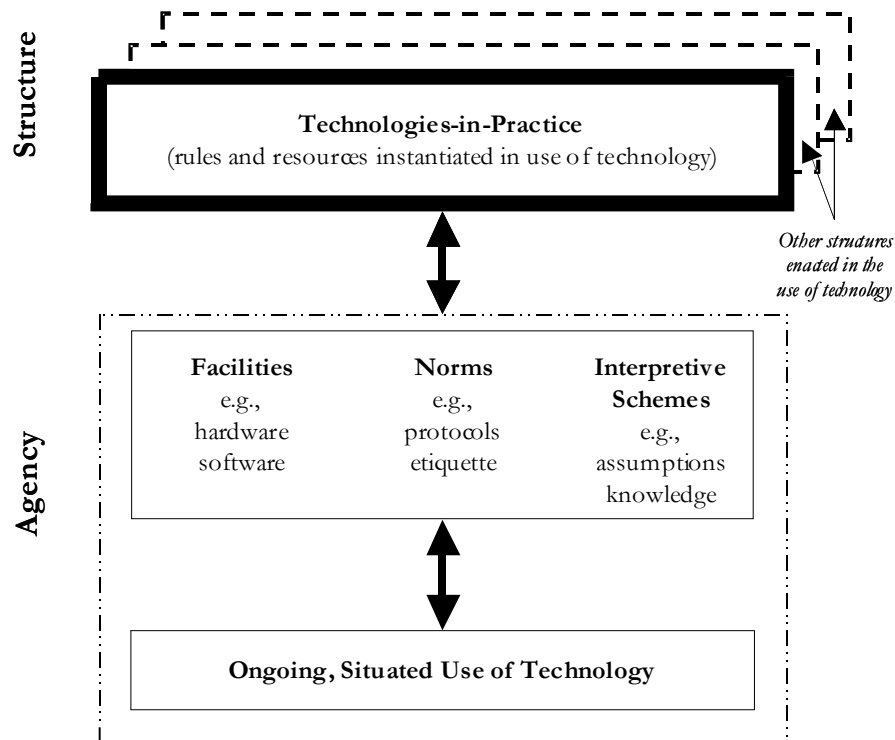
Source: Orlikowski, 2000: 410

This development of the technological structuration model is welcome because it facilitates a better engagement with technologies such as search engines which are difficult to consider as being stabilised and through which different kinds of structures are enacted by different user groups (e.g., designers, advertisers, search engine optimisers, single webmasters, 'ordinary' users, etc.).

An idea of group practice is of potential use in this context. The communities of practice model, developed by Lave and Wenger (Lave & Wenger, 1991; Wenger, 1998) can be used to complement Orlikowski's model, in which the group is conceived of as a somewhat unfocused entity, by developing a view of the appropriate group as one whose focus is a particular practice – for example, the production of search engine results. Conversely, it augments the view of practice, which in Giddens's original formulation as "practical knowledge" (Giddens, 1984:xxiii) is quite individualistic, by focusing on practices as group or communal activities.

Lave and Wengers' original focus was on learning, specifically on situated learning. They developed the communities of practice model to account for the way apprentices come to understand what it is to be a master. Wenger specifically provides a model which echoes the technology-in-practice model, when he discusses the duality of *participation* and *reification* in creating a community of practice. This conceptualization provides the major hook by which we can understand the relation of technology to practice.

According to Wenger, "*Practice is about meaning as an experience of everyday life*" (Wenger, 1998:52). What he seems to mean by this is that the practice of, for example, adjusting health insurance claims in a workplace setting involves *becoming* a 'claims adjuster,' a person with a both practical competence and a professional identity: someone who understands and negotiates meaning out of claims adjustment. It is worth quoting him at length on this:

> "Our engagement in practice may have patterns, but it is the production of such patterns anew that gives rise to an experience of meaning. When we sit down at lunch for the thousandth time with the same colleagues in the same cafeteria, we have seen it all before. We know all the steps. We may even know today's menu by heart; we may love it or we may dread it. And yet we eat again, we taste again. We may know our colleagues very well, and yet we repeatedly engage in conversations. All that we do or say may refer to what has been done and said in the past, and yet we produce again a new situation, an impression, an experience: we produce meanings that extend, redirect, dismiss, reinterpret, modify or confirm – in a word, negotiate anew – the histories of meanings of which they are part. In this sense, living is a constant process of *negotiation of meaning*." (Wenger, 1998:52-53, italics original)

This negotiation of meaning is essentially expressed in the two dimensions of practice that Wenger identifies: *participation* and *reification*. Participation refers to "both action and connection," that is to say, it is the active process of a member of a social community (and here Wenger specifically excludes non-humans, both technologies and animals)(Wenger, 1998:56). *Reification*, on the other hand, is defined as "the process of giving form to our experience by producing objects that congeal this experience into 'thingness'." (Wenger, 1998:58). Participation and reification form a duality in the sense of the duality of technology or of structure discussed above: a complementary and dynamic union. And in this complementarity, we can begin to understand the communicative content of artefacts

which "depends on how the work of negotiating meaning is distributed between reification and participation" (Wenger, 1998:64).

Let us consider now again at how the communities of practice approach can help us understand search engines. As we have seen, the social construction of technology (SCOT) framework emphasises the role of different groups in the historical processes of innovation and technology development. Specifically, certain "relevant social groups" are depicted most often as conflicting, with diverging values and norms that focus debate and development on particular areas (Bijker, 1995). Various processes of "closure," such as *rhetorical closure*, *redefinition of the problem* and *enrolment*, act to diminish conflict and achieve consensus-based artefacts. This depiction of the process of technology development requires and implies an interaction between the relevant social groups – and indeed, one definition of a relevant social group is that it is recognised by the other groups[8]. However, in the technology-in-practice model as developed by Orlikowski (2000) each relevant social group (for example, Lotus Notes developers and users in the sales group of Alpha or Zeta corporations) may give its own meaning to the artefact in question, incorporating it into a routine of practice, whether or not it recognises the other relevant social groups.

In this more pluralistic conception of the interaction between groups and technology, conflict between different value schemes is not necessarily resolved, but rather exists as a sort of tacit negotiation in which the designers, in effect, have the upper hand over the users. The technological frames of reference of each group are brought to bear upon the artefact in question; and routine, repeated interactions with the artefact lead it to embody certain structures – not universally, but in a particular given context. Thus, for example, the search engine results may be viewed as reified parts of the practice of different communities.

Further, these reified search engine results could be understood to mediate between different communities of practice as they are integrated into a range of routines. A useful way of conceptualising the mediating function of the search results is through the concept

[8] This has led to criticism about "invisible" groups, as this definition serves in some respects to reinforce a status quo in debate.

of the "boundary object." As set out by Star and Greisener (1989), boundary objects are material objects used in everyday community practice but importantly are shared with other communities. Star and Greisener give the following examples of boundary objects drawn from their study of the Berkeley Museum of Zoology: first, repository objects such as museums or libraries in which different groups may deposit information or objects; second, ideal type objects such as diagrams or atlases which can be created by different groups and shared with some characteristics being known to only one group or another and some to both; third, coincident boundary objects in which outlines are shared (such as on a map) but different properties are drawn within that outline; and fourth, standardized form objects, for example, a standardized method resulting in an index to which different communities can contribute and which all communities can use, albeit for different purposes. It is the function of all these boundary objects, according to Star and Greisener, to "translate" between different perspectives, and to serve as "a common coin which makes possible new kinds of joint endeavour" (1989, p. 414). In fact in some cases the boundary object makes possible the interaction between communities, since their creation and management is "a key process in developing and maintaining coherence across intersecting social worlds" (1989, p. 393)

Thus, one group may create the initial code, which functions as a 'resource' of stored information. But that resource, or boundary object, is instantiated into a structure with different rules (both interpretative and normative) in the different communities of practice. The meaning of the resource is collectively negotiated, and the local practices and local meanings may conflict with those of the original designers, which still having the function of co-ordinating action between the groups.

The boundary object concept is particularly helpful when considering the interaction between groups. The technology-in-practice framework suggests no particular mechanism by which different groups might inter-relate; rather, it implies that interaction between groups is outside technology and contained within the organization. The communities of practice perspective suggests that practices might become reified as part of the duality of reification and participation and that, in the context of search engines, the search engine results pages may be precisely such a reification of practice. However this situation is complicated by the fact that the search engine pages are the reification of the

practices of *multiple* communities – at a minimum, of the practices of the search engine producers and the search engine marketers.

Wenger, following on from his own work with health insurance claims processors, describes a situation reminiscent of that in which the search engine marketers or optimisers may find themselves:

> "Claims processors are not the designers of the rules and forms they use, yet they must absorb them into their practice. In an institutional environment such as a claims processing site, a very large portion of the reification involved in work practices comes from outside the communities of workers. Even so, however, reification must be re-appropriated into a local process in order to become meaningful." (Wenger, 1998:60)

Orlikowski provides a similar account of technology-in-practice as a socially situated phenomenon: "…people's interaction with technology will always enact other social structures along with the technology-in-practice, for example, a hierarchical authority structure within a large bureaucracy…" (Orlikowski, 2000:411).

Thus, we can suggest that in their reified form the search engine results provide a common thread which links together various communities of practice, both within and outside the organizations of the search engine producers. As Constant notes: "…[C]omplex, hierarchical systems imply multiple traditions of practice and multiple communities of practitioners." (Constant, 1989, p. 227). Each of these communities finds its own balance between participation and reification, and within each community, the search engine results page is differently appropriated into practice, which also represent different, potentially conflicting, negotiations of meaning.

While Orlikowski intends her theory to apply specifically to technology in organisations, much of what she says applies very well to media[9] which have a dual existence as text and as technology[10]. By developing a separate theory of 'technological' structuration, what she is implying is that Giddens sees the interactions that constitute the structuration process

---

[9] "Media" in this sense means television, radio, books, and Internet content and similar – communications delivered not face-to-face but via an intermediary technology, and often intended for multiple recipients; however the intermediary technology is often called the "medium," plural "media" and in practice the communicative content is also subsumed under "media" as well.

[10] Others have now also begun to try to redefine media as a complex of "practices" along these lines (Couldry, 2004).

as being unmediated. For, if he accounts for mediated interactions, then technology fits in well enough so that a specific theory of 'technological' structuration would not be needed. In the final chapter of *The Constitution of Society,* where Giddens is concerned with drawing his theory together and relating it to empirical research, he describes the nature of interaction as follows: "All social interaction is situated within time-space boundaries of co-presence (whether or not this be extended via media such as letters, telephone calls, etc.). Its situated character…is directly involved with the indexical nature of the 'bringing off' of mutually intelligible communication." (Giddens, 1984:332) This suggests that communication, whether mediated or unmediated, retains essentially the same characteristics and that no analytical distinction is needed. This seems particularly so as Giddens draws heavily from Goffman's (1969) work on interpersonal, face-to-face communication in his analyses of how the 'bringing-off' of communication is accomplished.

However elsewhere, Giddens classifies technology and media as *resources*, thus suggesting they are intimately linked with power and the structures of domination (Giddens, 1984:258). More than that, he gives media and information resources[11] a particular role in society as storage for resources: "The storage of authoritative and allocative resources may be understood as involving the retention and control of information or knowledge whereby social relations are perpetuated across time-space" (Giddens, 1984:261). In fact, Giddens continues:

> "It is the containers which store allocative and authoritative resources that generate the major types of structural principle in the constitution of societies…Information storage…is a fundamental phenomenon permitting time-space distanciation and a thread that ties together the various sorts of allocative and authoritative resources in reproduced structures of domination." (1984:262)

For Giddens these 'containers' are linked to the formation of complex societies: firstly in cities, then in nation states[12]. The mechanism by which this occurs is, presumably,

---

[11] Giddens seems to use the word 'technology' for the most part to refer to material artefacts, in that sense aligning it to the 'hardware' perspective Orlikowski discusses. Information and communication technologies (ICT's) seem to fall more generally in his conception of 'media and information resources.'

[12] In fact, Giddens uses the word 'containers' also to refer to cities and to nation-states themselves. However, previously he says "Storage presumes *media* of information representation, modes of information *retrieval* or recall and, as with all power resources, modes of its dissemination." (1984:261). Perhaps, therefore, a 'container' must contain all of these

through a structuration process in which storage media are involved in some fashion, but this is not developed in his text.

Orlikowski opens up Giddens's concept of interaction, moving from an essentially face-to-face perspective to one in which mediated interaction, specifically technologically-mediated interaction, is included. Thus, we can take her principle of *interpretative flexibility* to be a proxy for Giddens's "awareness and use of these phenomena [that is, facial expression, bodily gestures, linguistic and other media of communication] reflexively to influence or control the flow of the interaction" (Giddens 1984: 282). It is in the exercise of this reflexive influence or control or interpretative flexibility that the *duality of technology* is realized – in other words, the 'impact' model of technology is rejected in favour of an interactionist perspective[13]. But also, since Orlikowski rejects the purely social constructivist view of technology, she also implies that the technology in question is one of the actors in the interaction, albeit a non-reflexive one (and so of a different class than the human actors Giddens identifies)[14].

Taking Orlikowski's opening up of Giddens's structuration theory to technological interactions as a central building block of my conceptual framework, we can ask: what are search engine technologies *in practice* in organisations? What kinds of norms and interpretative schemes are brought to bear? What facilities are used? These are slightly different types of questions than are asked by many of those seeking to understand the innovation process in the context of software and service development. There remain, however, some important areas still to be adequately explained. These involve the larger question that Giddens alludes to about the role of information storage: what does it mean to have a privileged relationship to those containers that 'generate the major types of structural principles in the constitution of societies'? This leads back to the second critique of the SCOT literature: it does not easily deal with power or power inequalities, or

areas and might refer more generally to, for example, the 'network society' envisioned by Castells (1996). Still, ICTs would seem to be crucially implicated across both modes as a medium of information representation, and thus as an extension of resource storage.

[13] Perhaps this could be extended and referred to as *the duality of media* which would allow this elaboration to be applied to other types of mediated interaction such as television viewing, although that possibility remains outside the scope of this thesis.

[14] It is perhaps for this reason that certain studies, (including Hanseth & Monteiro, 1998; M. Jones, 1998; Walsham & Sahay, 1996), not including this current one, have used Giddens and Actor-Network Theory (ANT) together in the analysis of information systems, although their ontological foundations are quite different.

with ways in which specific forces of capitalism and technology are developmentally inter-related. The next section discusses this aspect, focusing on the political economy of media and communications to discern how we might understand search engine results in their capitalist context.

2.4 Conflict in context

Power dynamics are latent in each of the approaches I have discussed: SCOT, technological structuration and communities of practice. But since none deals explicitly with power, this concept is rarely carried forward in empirical research[15]. In this thesis, I develop a perspective which stresses the *conflict* latent in the discussion of search engine bias. This perspective contrasts with those where the emphasis is on consensus, closure, and stabilisation. Through conflict, some groups succeed, whereas others are less successful. Social theorists from Marx onwards have identified that certain social groups are more likely to succeed than others. From any political economy perspective, the position of groups with reference to each other is central. In this section I suggest that a strand of political economy of media and communication tradition, a macro-level theorical framework which is centrally concerned with power and the relative position of the media and technology helps to situate the overall analysis of search engine practices into the broader industrial context in which these engines are developed.

Beginning with SCOT, it seems intuitively clear that more powerful social groups would be more likely to achieve closure or stability on their own terms. History, geography, race, gender or age might privilege some groups. Winner, for example, criticizes the way in which relevant social groups are defined in these theories (that is to say, by mutual recognition of participants). What, he asks, about the social groups who are deemed "irrelevant" by the central actors (or actor-networks)? What about factors which lead to certain groups being effectively silenced? (Winner, 1993). Lave and Wenger also acknowledge that unequal power relations "must be included more systematically" in communities of practice work (1991:42).

---

[15] Mansell (2004) has identified this as a general weakness in studies of new media .

In Orlikowski's studies, although they include practices with different groups using the same technology, the dynamics of conflict and power tend to be underplayed. For example, she studies how Lotus *Notes* is adapted by different groups in different organisations, beginning with the designers of the product and extending into two different consulting organisations. She observes that many consultants in the firm 'Alpha' "remained sceptical and unmotivated" in their use of Notes (Orlikowski, 2000:416), using it at a "minimal, even perfunctory" level. Now, the use of Notes within Alpha had been mandated by their Chief Information Officer: "The CIO purchased thousands of copies of *Notes* for Alpha's consultants, and ordered his technology staff to install it…rapidly in all offices, so as to establish a critical mass of users as quickly as possible." It is possible to distinguish here the outlines of a conflict, however muted, between the perfunctory use of some consultants and the enthusiasm of the CIO. Orlikowski's approach, however, tends to emphasise the greater or lesser degree to which different communities of practice adopted the technology within their practices. The latent conflict is left unexamined. A potential conflict between the designer's wishes and the user's actual uses remains similarly uninvestigated, although a cultural explanation is employed to suggest that certain groups within Alpha (namely the IT department), and consultants in another firm, Zeta, shared a "collaborative" culture with the *Notes* designers, thus enabling them to use the technology more effectively (Orlikowski, 2000). Thus, the relative position of the CIO, the relation of Alpha and Zeta to other firms, and the positions of the various communities are not investigated within the framework of a micro-level analysis.

Potentially, structuration theory as described by Giddens provides a better framework within which to understand the initial position of the various actors with reference to their respective power relations. According to Giddens, social systems are composed of enacted structures and enacted structures draw their durability from social systems. He critiques the analytical separation of the macro- and micro-levels of analysis:

> "…[N]ot only do encounters 'slide away' in time but also once we start being concerned with how encounters are carried on by their participating actors, it becomes clear that no strip of interaction – even if it is plainly bracketed, temporally and spatially – can be understood on its own. Most aspects of interaction are sedimented in time, and sense can be made of them only by considering their routinized, repetitive character." (Giddens, 1984:142).

Giddens argues instead for a focus on "the modes of regionalization which channel, and are channeled by, the time-space paths that the members of a community or society follow in their day-to-day activities" (Giddens, 1984). Regionalization is a generic concept that refers to "the zoning of time-space in relation to routinized social practices" (Giddens, 1984:118). In other words, it can be seen as the spatial and temporal locus of a community of practice.

By using the model of technological structuration, we are able to suggest that technology-in-practice, as constituted by various communities of practice, contributes to the production and reproduction of social structure. In addition, these social structures contribute to the overall system. The practices themselves are initially constrained by the historically rooted rules and resources available to that community of practice. The link between the meso-level communities of practice and the macro-level system depends upon the individual structures enacted through practice. Understandings of that link can be changed if we incorporate a macro-level view of the system into the analysis.

As is the case in most studies of the sociology of technology, various strands of political economy of media and communications begin with a presumption that information and communications technologies (ICTs) are not the outcome of a process of "natural" development or evolution. Instead of stressing the dynamics of the local community, political economy theorists often suggest that the development of technology is intimately intertwined with the social, political, and economic context in which it arises. Indeed, some suggest that technological development is central to the capitalist system as a whole as the economic historian Robert Heilbroner writes:

> "Technology [is] a sociopolitical force within capitalism, not merely a lever of material change. The reason…is that technological change is the chief source of new areas of profitable accumulation…Capitalist economic history is thus written in bursts of accumulation largely brought on by technological change, followed by periods of slackening expansion as competition erodes profitability." (Heilbroner, 1997, p. 1324)

In this context, the quest for profit is understood to both direct technical developments in information and to be supported by them (H. Schiller, 1992; Webster, 2002).[16]

The political economy perspective encourages us to investigate biases in technical developments in information, such as search engines, within their capitalist context, so as to gain insight into major influences in their formation and dynamic development which potentially hold the key to their biases. Mosco, in his review of the field of the politial economy of media and communications, proposes three "entry points" into investigating communications activities in a capitalist context: *spatialization, commodification,* and *structuration* (Mosco, 1996). Structuration has already been discussed, but the concepts of spatialization and commodification can be used to guide the development of the conceptual framework for this study.

2.4.1 SPATIALIZATION

Spatialization, as Mosco uses the term, has two main elements. First, it refers to the material reality of communication processes and industries in space and time: to where they are located, to their history, and to the processes by which they govern the extent of their businesses. Secondly, it refers to the way in which communications processes aid the reshaping of space and time. The use of ICTs and processes in the spatial and temporal management of other firms, of the state, and of private lives, is also a concern which falls within this sphere of research.

Thus, the focus on spatialization involves an inquiry into the way ICT systems allow businesses to expand in space and time and to be more flexible. The expansion of the firm vertically, to cover supply and distribution, and horizontally to cover related products and services, helps to minimize risks from competition by creating a stable framework within which the firm can profitably do business. This has been operationalised traditionally through the study of industry concentration and ownership, by asking questions such as how large is a particular company? Who owns that company? How has it expanded or contracted across time? This tendency of firms to try to regulate their own environment instead of subjecting themselves to competition is, of course, also a concern

---

[16] Thus multinational corporations are the largest users of information technology (Mosco, 1996:182), and information technology companies are among the largest multinationals.

of neoclassical economists and regulators investigating market failure. Some political economists often investigate not simply ownership structures but also informal networks of alliances, partnership, and overlapping memberships of corporate boards of governance, as well as the relationship between national and international policy and corporate expansion (Melody, 2003; Murdock & Golding, 1999). In the media industries, particularly, a trend to converge into giant conglomerate companies has been documented (Herman & McChesney, 1997). Similar trends can be seen also in particular media industries. Film (Wasko, 1994) and advertising (Leiss, Kline, & Jhally, 1990; Mattelart, 1991) both show a similar pattern of expansion both horizontally and vertically, as do telecommunications companies in general and internet infrastructure companies such as Internet Service Providers (ISPs) (Mansell & Javary, 2004; Mansell & Steinmuller, 2000).

The focus on spatialization suggests research into centres of industrial power in the ICT industry – for example, into areas such as the Silicon Valley – where overlapping clusters of firms and support structures create spatially compact centres of the information and communication business. Within the USA, Mosco has pointed to "spatial agglomerations" like New York City (Mosco, 1999) where a complex of private funding is creating a new high tech centre at the expense, he argues, of public parks and public spaces. This expansion crosses, but does not ignore, national boundaries. Herbert Schiller raised the issue of the relation of the "home countries" of these media conglomerates, particularly the USA, to the rest of the world (H. Schiller, 1992). Thus, while some scholars have focused on the very localized qualities of some communications industries – for example Hollywood as the centre of film-making or Silicon valley as the heart of the internet industry (Wasko, 1994; Zook, 2005), many theorists have examined the transnational or global dynamics of the media and communications industries.

We can characterise three major areas in which the media system plays a part in global capitalism. These areas are analytically distinct but reinforce each other and, in practice, may be intertwined. First, the media can serve as an instrument of direct foreign policy. One example of this is the use of media for propaganda purposes during war or peacetime. Thussu, for example, highlights the use of the Voice of America and Radio Free Europe/Radio Liberty as elements of United States foreign policy during the Cold War (Thussu, 2000, pp. 31-35). A second example of foreign policy intervention is the

tying of World Bank, International Monetary Fund, or other development assistance to the privatisation or liberalisation of existing national media in developing countries (H. Schiller, 1998, p. 19). Mattelart (1994) extends his analysis of state action through media further to include overall security policy, for example, use of media technology in surveillance and its implications for privacy.

Second, the corporate media can serve to further the agenda of global capital. Private or elite financial media provide essential information to transnational corporations. But in addition to making international business more efficient, Herman and McChesney (1997) argue that advertising agencies (on behalf of large global advertisers and with home government support) act to diminish public broadcasting in peripheral nations in order to help open new markets. Indeed, so strong is this role that Herman & McChesney in the title of their book call the global media "the new missionaries of corporate capitalism" (Herman & McChesney, 1997). Matellart (1994), in a similar vein, sees the media as promoters of the ideology of progress which, among other things, ties the expansion of communication technology to development and educational goals.

The third area in which media have been identified as actors in global capitalism is in the establishment of what might be called a centre and a periphery of representation which serves to bolster both the interests of Western government (particularly the United States) and of transnational capitalism. Originally formulated in the 1970s, the thesis of media imperialism included the two dimensions previously discussed (direct foreign policy and corporate promotion), and also suggested that the media "promot[ed] an image of the USA and of the world that was favourable to American interests, and advertis[ed] American goods and services – directly through the provision of more channels for advertising, and indirectly through the display of consumer lifestyles." (Boyd-Barrett, 1998, p. 158) This strong view of media imperialism as a totalising system was also strongly criticised: in such a system, it was argued, the public are inevitably reduced to media dupes, unable to make their own choices, while evidence showed that interpretation and incorporation of media messages into daily life is a complex and often conscious project of the audience (see for example Ang, 1985, 1996; Livingstone, 1998). Matellart (1994) also highlights conflicting dynamics within the global media system. His account of the relation between culture and the media encompasses both the effort to

deterritorialize communication and to develop one consistent, hegemonic global culture as well as local efforts to reterritorialize and insist on difference.

These theorists, then, argue that both states and global capital use media communication as tools to further their geopolitical interests. These theories have been developed primarily in relation to traditional media: television, radio, the press and so forth. Communication infrastructure has also been studied with as an enabler of these processes. Schiller, for example, reviewed the development of satellite technology specifically as an instance of foreign policy intervention on behalf of American business, in order to break the European dominance on news which the British had, in part, established through their ownership of the transcontinental telegraph cables (H. Schiller, 1998, p. 20). Further, the development of communication infrastructure has enabled business offices to communicate more easily, making feasible the further spread of multinational corporations. Indeed, Mansell and Steinmuller have argued that the primary use of the internet was for business-to-business communication (Mansell & Steinmuller, 2000).

Despite this work, a treatment of new media forms such as search engines based on theories of spatialisation or global politics is not yet fully developed. This is a rich area and aspects of it are addressed specifically in relation to the bias in search engine results which arguably has a global dimension (see Chapter 5 for further discussion).

2.4.2 COMMODIFICATION

The previous section focused on the conceptualisation of time and space as one way to understand how some aspects of capitalism might affect the search engine industry and its results. This section focuses on commodification, or the ways in which goods and services which formerly were outside the profit system become integrated into it: in our case, this concerns the way search engines produce and fund their virtual services, and therefore this section focuses primarily on the dynamics of advertising.

Commodification generally refers to the process by which goods or services that previously were valued only for their usefulness or the pleasure they give – such as entertainment and recreation – are given a price and integrated into the profit economy or even "made available only on the condition that [they are] saleable" (Webster, 2002, p. 128). Commodification is an essential part of capitalist economic growth, which often

results from making saleable goods that were previously exchanged outside of market relations in the course of everyday life, rather than necessarily from an improvement or increase in production of existing goods (Heilbroner, 1985, p. 60). Many political economists investigate growth dynamics because their effect can be to remove information and other cultural goods from the grasp of the poorest by virtue of the price that is assigned to them. The analysis of growth dynamics, for many political economists, needs to be linked to values such as equity and freedom – for example the freedom to enjoy one's cultural heritage. Indeed, Garnham defines the role of political economy as "link[ing] the analysis of capitalism, both as an overall social formation and as a specific mode of production, to the normative definition and realisation of human liberation" (Garnham, 1990, p. 5). Communication and culture are specifically related to human freedom, first through a presumed right to enjoy one's heritage and, secondly, through democratic government, where equality of information is regarded an essential principle. Mosco (1996) therefore identifies the process of the commodification of culture and cultural life as a central analytical axis for the study of the political economy of media and communications.

Despite the centrality of information and culture to public life, some political economists point out that the information and entertainment industries prioritise business needs secondarily attending to the needs of the public or the "consumer". For example, the telephone network developed first in the United States within urban centers where business was concentrated, despite the urgent need for communication in rural areas (Fischer, 1992). After many struggles, regulation by the U.S. government using the principles of "universal access" now mandates that telephone companies provide basic services on an affordable basis to all members of the public – although businesses still receive more advanced services than the general public.

What services, then, do media and entertainment provide to business? From a business perspective a major function of the cultural industries is to deliver a sales message to the audience of the publication or, in other words, to provide advertising. Dallas Smythe, one of the first political economists in the Western world to focus on the media, referred to the sale of advertising as the sale of the "audience commodity," drawing attention to the way in which the individual's time and mental attention is part of the capitalist's profit

calculations (Smythe, 1977). Advertising was the crucial funding vehicle for most mass media in the United States and the wide availability of advertising vehicles has been linked historically to the spread of capitalism. Some institutional economists argue, indeed, that advertising is central to advanced capitalism, since without advertising demand for the vast productive capacities of modern industry is difficult to sustain. Galbraith, for example, argued that that job losses and economic collapse would ensue were individual consumers not persuaded that the continual purchase of new goods was necessary (Galbraith, 1985).

It has also been argued argued that advertising operates as a de-facto censor, or filter, of media programming (Herman & Chomsky, 1994). In cases where advertising is central to television production, for example, programs which have insufficient ratings, or an unwanted demographic, are often cut because the producers cannot make the program without the support of advertisers. For example, one study in the United States reports that advertisers pay a premium for young, affluent viewers versus older people on fixed incomes, at a ratio of 48:1, and persistently either discount or fail to fund shows viewed by blacks (Gandy, 2004).

The relation of advertising to journalism is no less fraught. In the United States, for example, while cases where a story is killed outright by a major advertiser may be rare, expert public relations can mean that well-funded institutions receive more than their share of news coverage (Herman & Chomsky, 1994). Public relations firms are integrated into large advertising conglomerates, and "integrated" or "through-the-line" strategies make plain that advertisers consider news to be part of an overall communication plan (Schudson, 1984, p. 100).

The dynamics of commodification require advertising to reach ever-larger markets in order to create demand for an ever-growing supply of products. Political economists argue that advertising is providing an essential service to business by providing the attention of potential consumers and that advertising is a generic influence on the kind of program produced (primarily those for affluent audiences). In certain instances, advertising and marketing may have an effect on specific programmes or items of news, for example by contributing press releases. The search engine industry (see Chapter 4), is

very strongly advertising based and similar dynamics (though not necessarily identical ones) may in evidence.

2.5 Conceptual framework

This chapter has outlined the theoretical perspective for this study of bias in search engines. It has suggested that bias should not be seen as deviation from an ideal but rather as an expression of a normative conflict about what *should* be on a search engine's results page. It draws insight from studies of the sociology of technology which suggest technological development is bound up in historical circumstances which can be revealed at a micro-level. However, rather than privilege stability and closure, notions which are the focus in many strands of work in this tradition, in this study technology development is seen as a locus of ongoing conflict, in which both individuals and wider social systems have a role to play. At a micro-level, search engines can be studied as a boundary object between different communities of practice, where the analysis of differing meanings of search results for the communities in question can be expected to throw light upon this conflict. The absence of a wider of power relations in most studies of the sociology of technology has been noted, particularly in terms of the relation of technology to capitalist dynamics. This observation suggests that bias needs to be studied from a perspective that enables a focus on the dynamics of spatialisation and commodification as developed within the study of the political economy of media and communication tradition. By drawing on structuration theory it is feasible to link the micro and macro levels of analysis to develop a conceptual framework for this study.

The conceptual framework is represented as a diagram in Figure 4. The outer circle represents the macro-level of analysis or the larger context of the material production of search engine resi;ts. This is the area where the issues of spatialisation and commodification come to the fore. Practice, the inner circle, operates at the level of the everyday. The processes of participation and reification help make up the *technologies-in-practice* within the communities of practice that are the object of inquiry. The insights from structuration are represented by the dotted boundary between the two circles: the issues of norms, interpretative schemes and facilities are important at both levels as is the dynamism of the system and its flexibility. Context does not determine practice, nor does

practice determine context, rather each is understood to have an effect on the other and to be open to change.

Figure 4: Conceptual framework: the linkage of practice and context through structuration
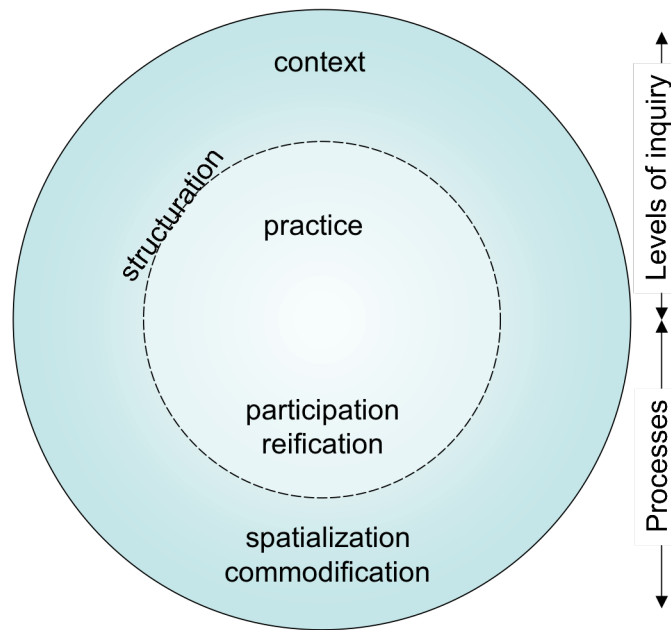
Source: author.

The processes of change within the framework are important, as research in the political economy tradition has often been criticized for an overly deterministic focus on economic factors (Schudson, 1989) and for its neglect of cultural form (Jhally, 1987). Within the field of media and communication studies, the key argument against the claims of political economy theorist has been that the consumer of media or information – the audience – is not a passive victim of ideology but actively constructs different interpretations of media texts (Ang, 1985, 1996; Hall, 1980). Most branches of SCOT also reject the idea of "effects" being built into technological forms, stressing the incorporation of technology into daily life (Bijker, 1995; Callon, 1986). Outright resistance, play, and transformation are all elements of the interaction of mass-produced corporate texts and technologies with people in everyday life (de Certeau, 1984; Feenberg, 1999, 2000). Nonetheless, the perspective of many writers in the tradition of the political economy of media and communications is that while the relationship between the strategies of the large corporations and governments and the tactics of the average citizen will be dynamic and

afford a range of possibilities, still it is *most likely* that the interests of business in a capitalist system will prevail and be visible in the emerging structures and operations of a given industry segment. This is an important perspective to consider alongside the social construction of technology arguments, where "society" is occasionally said not to exist or to be made up solely of individual interactions (Law, 1991). The pre-existing and persistent nature of certain dynamics within capitalism is not entirely due to local factors and cannot be fully explained based on a micro-level analysis.

The aim of this study is to construct a linkage between studies of communication and media, emphasising the political economy perspective, with studies of the construction of ICTs within the context of a social theory of the structuration process. Whereas Orlikowski applies her model to information technologies within organisations, conceptual framework for this study is applied to a specific media technology, i.e., search engines. The conceptual framework is applied in the analysis of the empirical research to yield provide a better understanding of the dynamics of power and conflict that can be shown to be at work in the production of search engine results.

2.5.1 POTENTIAL ALTERNATIVE FRAMEWORKS

I considered other perspectives for studying search engine bias. Very little work has been done on search engines from a sociological perspective. Rather, search engines are most commonly studied within two paradigms: information literacy or information retrieval. Studies using the information literacy perspective are focused primarily on the user and their interaction with search engines, either in terms of a quantitative analysis of search terms, or using experiments to assess user competence. While providing interesting evidence about the consequences of search engine bias, these studies rarely seek to evaluate search engines beyond an ease-of-use framework, nor do they typically question the rationale for the presentation of results. Information retrieval studies tend to concentrate on the performance of either existing or new search engines judged against a corpus of material either in a special collection or specially selected from the web. These studies provide important evidence for the existence of search bias but most often do not go beyond an analysis of the algorithm. Neither of these perspectives was deemed to be suitable to address the issue of why bias is embedded in search engine results.

Another approach considered was the use of the Actor-Network Theory (ANT) model, primarily developed by science and technology scholars Latour and Callon and sociologist Law (Callon, 1986, 1987; Latour, 1987, 1991, 1996; Law, 1987, 1991, 1992, 1997), which is linked to the social construction of technology perspective. ANT addresses the formation of humans, ideas, and machines into coherent but heterogeneous actor-networks. While different uses of ANT abound, the perspective is often applied to the analysis of a specific actor-network in a particular setting. Using the ANT methodology, machines are treated as equal to human actors, with both humans and non-humans being assumed to possess intentionality. In most applications of ANT, the task is to show how society is constructed through a series of actor-networks, and why certain actor-networks persist while others crumble. This perspective seems to suggest that there are no pre-existing power relationships but that each relationship is formed anew during the formation of an actor-network. The ANT framework makes it difficult to integrate large-scale power dynamics that I argue are important in a comprehensive analysis of search engine bias[17].

Another approach considered was the "grounded theory" approach (Strauss & Corbin, 1998), which would have been suitable for an in-depth ethnographic study of a single search engine. Grounded theory is a methodology for theory generation and is suitable when current theories seem no longer productive of new insights or when the research object is new (as is the case with search engines) and suitable theories have yet to be developed. Grounded theory was developed by Strauss as an analytical approach to a corpus of interviews and is primarily applicable to qualitative methods. The outputs of an application of grounded theory are a series of "dimensions" related to the research question. Each of the dimensions is suggested by the analysis of primary material and analysis is carried out immediately after each interview to develop and test the dimensions and their range. A series of interviews or interactions is carried out until no new dimensions are discovered in the analysis. The advantage of this method is that the theory generated is closely related to the experiences and interpretations of the research subjects. The disadvantages are that the method makes it difficult analyse and interpret non-qualitative data, such as financial or ratings data, or data which exist at a systemic level rather than at an experiential level. In addition this approach had some

---

[17] Its ontological equation of humans and non-humans, moreover, made it difficult to link to structuration theory, in which interpretative schemes are critical.

methodological disadvantages in that potentially a very large number of interviews or interactions is required. In this study, based on elite interviewees (see Chapter 3, section 3.4.1 for a discussion of interviewing elites), this method would have been extremely difficult to apply.

Each of the alternative theoretical frameworks and methodological strategies considered offered some advantages but none was as suitable for investigating the question of why bias becomes embedded in search engine results as those drawn upon and integrated in the conceptual framework outlined in the previous section.

2.6 Conclusion

This chapter has developed a conceptual framework to address the research question, *why does bias arise in search engine results?* The approach in this study is to examine both the context of search engines and their practice, drawing on theoretical perspectives in the social construction of technology and the political economy of media and communication traditions. In the light of the conceptual framework, the overall research question can be broken down into several sub-questions:

- What specific aspects and elements of the capitalist processes of spatialization and commodification are linked to search engine bias and how?

- How can we understand search engine results creation as a practice, in which members of different communities participate and from which the search engine results emerge as a reified object? Where does bias become embedded in that practice?

- How do the processes of the structuration of the search engine results influence the way bias is routinely embedded in everyday actions and in large-scale search engine industry processes?

The next chapter sets out the methodology for this study.

III

Follow the Results
A Methodology for Investigating Search Engines

3.1 Introduction

For this research, I comprehensively reviewed over a dozen annual financial reports, read ten years of press releases from eight companies (over 600 releases) and over 400 articles from both specialist and general press, interviewed eighteen experts in depth and conducted eight days of observation over six months within a small search engine optimization company. The process took approximately one and a half years. This chapter describes the way in which I conceptualised, organized, and conducted that research, the difficulties I faced in carrying it out, and the solutions at which I arrived. It begins by considering how the overall research question was developed, then reviews how the empirical research was designed, reflecting on key issues arising from both the conceptual framework and the empirical context. An empirical framework consisting of two levels of research – agents and actions, and structures – guided the overall methodology. The data gathering and analysis for both of these levels is described, including ethical issues that arose. Finally, I explain why several potential alternative research designs were not pursued.

Of necessity, this chapter presents the process of research as more tidy and linear than it was in practice. This is for the reader's convenience in making sense of what was a complex process. However, mindful of the ways in which the context of collection can affect the data, I indicate where the messiness of the research process seemed to become an issue in and of itself.

3.2 Development of the research question

I began my overall review of the topic of search engines with a simple question: what do we know about search engine production? This section reviews my approach to answering that question through a literature review and the development of the overall research question. This process contributed substantially to my effectiveness in the interviews I eventually conducted (described below). The method of the subject-specific literature review is not typically included as a part of the thesis, yet it is integral to the process. As the technological and informational resources available to researchers change, the literature review also changes; and a familiarity with technical resources is especially necessary for those working in emerging fields such as internet studies[18] (which encompasses work on search engines).

I drew my method of literature review partly from Hart's guide to the subject (Hart, 1998). Writing in the late 1990s, Hart addresses some of the electronic tools that were at that point beginning to become available, particularly stressing their ability to help the researcher complete a *systematic* and *comprehensive* review of a subject. Hart develops a several-stage model of interrogating the literature based on data sources: books, articles, and theses, in that order, supplemented by a recursive process developed through bibliographies and citation analysis, and aided by extensive notetaking (Hart, 1998, p. 35). In my own review, I modified Hart's approach to suit the tools available in the LSE's research environment, and to take advantage of my own learning style, which is suited to visual learning.

---

[18] "Internet studies" is loosely defined as the group of literature which takes specific elements of the internet as its object of study. These studies are frequently interdisciplinary and the objects of study (e.g., home pages, weblogs, social networking sites) change as the internet develops. At least one scholarly association, the Association of Internet Researchers (AOIR) exists to share and publish work on internet studies topics.

I began by seeking what might be termed the canonical sources on search engines: core books which set the scene for future research. I did not find any such books – indeed the only books available on search engines were guides for users[19]. I then turned to articles and theses. I used electronic databases to search for variations on the phrase "search engine" and "internet OR Web" in relevant syntax. The databases I used were the Web of Science, the International Bibliography of Social Sciences, ScienceDirect, ABI/Inform, and the Ingenta journals. These resources indexed journals from social science, computer science, and economic and business journals. I searched for dissertations on UMI Digital Dissertations (American/Canadian), Theses Online (British) and the Australian Digital Theses Program.

As I retrieved the relevant bibliographic records, I exported them into a bibliographic database held in the EndNote bibliographic manager, deleting duplicate references where they occurred. In the end, I had 1,070 articles and 40 doctoral theses. I skimmed the abstracts of these documents and developed an initial mind map which is included as Appendix B. My mind map was developed using principles from Buzan (1982). Essentially I started with the core concept, search engines, and as I reviewed each abstract I grouped it by theme. Alternatives to the mind map might include the concept trees, thematic maps, or subject relevance trees (Hart, 1998; Orna & Stevens, 1995). However, I wanted the material that I generated to be held on a computer, so that it would be possible to modify it easily and to link it to my other digital resources. Mind mapping programs are well developed in contrast to some of these alternative mapping tools. By using a mind mapping program – in this case, MindManager[20] - the map developed organically as I began to make sense of the mass of data. I have since modified this technique to include author names in the mind map, enabling me to easily reference the relevant publications in interesting areas (see Van Couvering, 2004) by searching for author names in my bibliographic database.

After reading and mapping the abstracts, I felt confident that little was known about search engine production. Of the articles I had discovered, fewer than ten referred

---

[19] For a popular account of the development of search engines and Google in particular, readers might now turn to Battelle (2005) and Vise and Malseed (2005). For more about the way search engine rankings are constructed, Langville and Meyer (2006) is a good but somewhat technical explanation.

[20] I have since used FreeMind, a similar application available freely in an open source version.

specifically to sociological or economic concerns. The large majority, perhaps three-quarters, were computer science or mathematical discussions focused on improving search engine algorithm programming or on developing specialist search engines. A second, smaller proportion referred to studies in which search engines had been used as a sampling tool, primarily in the medical field; for example, to survey the quality of information online for patients. Finally, about one tenth of the studies focused on user evaluations of search engines. A discussion of the most useful sources is found in Chapter 2, Section 2.2.

From this point on, my literature review methodology was more conventional: I reviewed the studies that addressed economic and sociological concerns primarily. Chief among their concerns was search engine bias, which helped to crystallize my own research question. I then used their bibliographies, and the bibliographies of the studies they referred to construct a reading list that provided the basis of my review of the search engine literature. The process of gathering reviews and mapping was comparatively efficient, taking approximately two weeks from start to finish. The more in-depth reading took much longer. However, the initial review and mapping project continued to be useful throughout as different aspects of search engine research have been helpful at different stages.

3.3 Research design

Once the research question was clear – why does bias arise search engine results? – it was time to design the empirical research. From the start of the research design process two overarching issues were apparent and these continued to influence the design up until the conclusion of research. The first issue was conceptual: with a conceptual framework that stresses both micro- and macro-factors as potential influencers in bias, how was I adequately to address these two very different levels of research? On what basis would I compare, contrast, or inter-relate them, or have assess the validity of my conclusions? The second issue was more practical: I knew there might be a problem of access to the elite teams of programmers and designers who make up the core of search engine production. My original research plan called for a period of observation within one of these teams as a key source of data. In the end, that proved not to be feasible and the issue of access continued to affect the amount and kinds of data I was able to collect.

Addressing the first issue, that of the inter-relationship of the different levels of research and how to relate and compare them, I followed Giddens who argues that empirical research into the process of structuration can be carried out at any one of the four levels implied by his work, shown in Figure 5:

Figure 5: Levels of research into the structuration process

Hermeneutic Elucidation of Frames of Meaning	(1)
Investigation of Context and Form of Practical Consciousness (The Unconscious)	(2)
Identification of Bounds of Knowledgeability	(3)
Specification of Institutional Orders	(4)
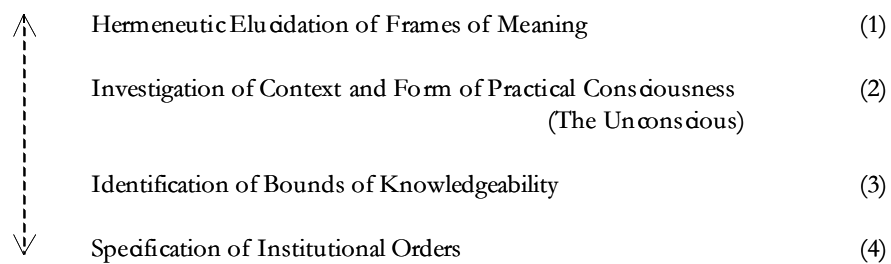
Source: Giddens, 1984: 327

Qualitative research is represented most often by levels (1) and (2) and quantitative research by levels (3) and (4). Giddens suggests that each needs the other in order to achieve a comprehensive view of the structuration process. However, he is somewhat less clear on concrete methodologies, although he stresses the need for contextual research which has within it an "ethnographic" aspect, or connection to the everyday, and is respectful of the skills of the people studied. This has led to criticisms that structuration theory is a theory incapable of generating either useful research questions or analytical frameworks. One critic says

> "…it would be unreasonable to expect structuration theory to generate either empirical research questions or appropriate categories for empirical analysis and… to transfer structurationist concepts directly into empirical analysis is misconceived [because ontological issues are at a premium]." (Gregson, 1989, p. 247)

Giddens replies to this criticism as follows: "Structuration theory is not intended as a method of research or even as a methodological approach," (Giddens, 1989, p. 296). While this may not seem very encouraging for an aspiring researcher, he does go on to say that "the framework of structuration theory both provides concepts relevant to empirical research and also warns against the pitfalls of some types of research procedure or interpretations of research results" (Giddens, 1989, p. 296). In other words, structuration

theory provides an "orientation," which focuses on how social practices are constituted and reconstituted. Structurationally oriented research can neither accept actors as wholly free agents nor structures as unchanging or rigid. Empirically this suggests a research design that can take account both of flexible structures and of agents' actions within those structures, and which encompasses both qualitative and quantitative methods. This general advice still leaves us some distance from a methodology for this project. I treat structuration theory, as Giddens does, as an orientational device, and will return to it in Chapter 8.

In Chapter 2, I discussed two research traditions that have guided the present research methodology. The first tradition is the social construction of technology, and the second is the political economy of media and communication. Methodologically, they are quite distinct. Nevertheless, the current research draws on both traditions to connect structures and agency in the investigation of search engine bias. Methods from the political economy tradition such as the analysis of financial data, company press releases, ratings data and expert interviews are employed to help shed light on the large-scale processes and historical power relations that shape the context of individual action. Discourse analysis of the accounts that producers give of the tradeoffs they encounter in the production process is used to yield insights into the implications of agents' actions on a micro-level. The link which brings the various actors and structures together is the thread of the search engine results pages: the research is primarily concerned with how these pages, which are what the everyday internet users sees and uses, produce biased results[21]. These results are the "thing" in which bias becomes visible, they are a nexus of economic transfer, since they are bought and sold, and they are a site of conflict as various parties try to influence them through non-commercial means.

Discussion of the application of structuration theory further suggest that the comparison of small-scale and large scale data will be meaningful as, at the analytical stage, we will be able to analyse agents' actions as being in some ways meaningfully constrained and, in other ways, meaningfully encouraged by the structures within which they are embedded.

---

[21] This is reminiscent of anthropologist George Marcus's exhortation to "follow the thing" in multi-sited ethnographic research: "Multi-sited research is designed around chains, paths, threads, conjunctions or juxtapositions of locations in which the ethnographer establishes some form of literal, physical presence, with an explicit, posited logic of association or connection among sites that in fact defines the argument of the ethnography" (Marcus, 1995, p. 105)

We will also be able to elicit a broader context for the interpretation of those small actions as they are writ large in the production and reproduction of structure.

A summary of the overall research design, with sub-questions and methods, is given in Table 1, for ease of reference.

Table 1: Research Design Levels and Methods

Level A – Agents – socially constructed meanings and everyday practices
Theoretical research sub-questions (as discussed in Chapter 2, section 2.6):
• How can we understand search engine results creation as a practice, in which members of different communities participate and from which the search engine results emerge as a reified object? Where does bias stand in relation to that practice?
• Following that, how is search engine bias routinely embedded into everyday actions, and how do those relate to large-scale processes? In other words, how can we understand the structuration of search engine bias?
Key sub-questions:
• How do search engine producers conceive of search engine results?
• How do they make decisions about where to allocate resources and how to make changes to the search engines?
• What notions of quality exist, and with what consequences for the search results?
• How do search engine optimisers interact with search engine results?
• How do search engine results act as a boundary object between optimisers and producers? How else do these communities interact?
Key methods:
• Discourse analysis based on interviews with search engine producers
• Observation of search engine optimisers
Level B – Structures – history, geography, and economics of search engines
Theoretical research sub-questions:
• What aspects and elements of the capitalist processes of spatialisation and commodification are linked to search engine bias and how?
• Following that, how is search engine bias routinely embedded into these processes, and how do those relate to everyday actions? In other words, how can we understand the structuration of search engine bias?
Key sub-questions:
• What is the ownership structure of the industry? Is the industry concentrated into certain companies and/or certain geographies? How has it developed over time?
• What is the role of advertising in the industry? How much does it contribute to revenue? How does it operate?
• How does the pure search product fit into the overall financial structure of the company? What contribution does it make to revenue?
• What is the relationship between global search provision and local search provision? Can we distinguish a centre or periphery in current search engine operations?
Key methods:
• Revenue analysis of financial reports
• Analysis of press releases and news coverage on search

- Review of relevant policy documents
- Analysis of ratings data
- Interviews with search engine producers, optimizers and distributors, as well as industry commentators

This section has outlined the overall research design for this project, which is conceived of in two layers: Layer A concerns agents and their actions and Layer B concerns structure. The research design is informed by structuration theory, which helps bring the findings from the two layers together during the analysis. The first layer, concerning agents and their actions, is primarily based on interviews with search engine producers whose discourse is analysed to investigate their interpretative schemes, as well as on observations and interviews with search engine optimisers. The structural layer is investigated using methods often employed in the political economy tradition: analysis of revenue streams and advertising, of ownership, and of relevant policy documents, supplemented by information from primary interviews with a variety of people (producers, optimizers, distributors, and commentators). The next section discusses how the data for the thesis was collected and analysed in detail, taking each level in turn.

3.4 Level A: Agents and their actions

The theories of the social construction of technology that are at the core of the conceptual framework for this thesis are based on the premise that the specificity of a particular technology is created within the social contexts in which technologies are developed and used. These social contexts – and, therefore, ultimately the meaning of the technology – can be investigated in different ways, but in general the focus is on small

groups of people and their interactions, practices, and routines. Often investigation into the social construction of technology is historical and the sources of information are primarily documentary (Bijker, 1995; Bijker, Hughes, & Pinch, 1989; Latour, 1987, 1996; Law, 1987; van den Belt & Rip, 1987). Others analysts concentrate on ethnographic observation, interviews, and textual analysis (Bowker & Leigh Star, 2002; Orlikowski, 1992, 2000; Star & Ruhleder, 1994). All these texts deal with the construction of meanings relating to technology – the "hermeneutic elucidation of frames of meaning" to which Giddens refers. In the current research, as outlined in Chapter 2, Orlikowski's concepts of "technology-in-practice" and particularly of "technological frame of reference" are central. These can be investigated either through observation or through interviews.[22] I analyse the interview data from the search engine producers using methods of discourse analysis drawn from social psychology, particularly Potter & Wetherell's (1987) concept of "interpretative repertoires," which I discuss below.

At the beginning of the Section 3.3 I indicated that in addition to the problem of integrating large-scale structural research and small-scale research into everyday practices, there was a practical difficulty in collecting data from the search engine companies. The gathering of the documentary evidence I review in the next section on structures presented little challenge in terms of access as I relied upon publicly available data, primarily gathered online. I collected financial reports and press releases from company archives to assess revenue and the history of mergers and acquisitions from 1993-2005. However, for the qualitative data I had to seek access to the search engine companies, which proved to be quite difficult.

3.4.1 DATA GATHERING

My data gathering in terms of agents and their actions initially had to two strands: participant observation with a search engine marketer and interviews with search engine producers. Unfortunately, shortly after I had completed the participant observation I lost nearly all of the data. The section below therefore first recounts the organization of the participant observation and then discusses how I dealt with that data loss before continuing on to describe the interviews.

---

[22] One review of eight empirical studies using the technological frames of reference concept showed that while observation was occasionally used, interviews were the primary data gathering technique (Davidson & Pai, 2004).

For my participant observation, I wanted a to choose a company which was established and respected, and which, to some extent, represented mainstream practice for search engine marketers. I located a company based in London that had been in operation since 1996, after hearing their managing director speak at a conference. This company was small, with only 5 employees, as are many search engine optimisers, but they had both large and small clients, and their business spanned both natural search engine optimisation and search advertising, as is common.

Access to this company was relatively unproblematic. I approached the managing director via email, and he agreed to meet in person. I explained my study to him and he was keen to be involved. He proposed a trade: I would be able to observe if I aided him in developing a training programme for new recruits into the business, as they were planning on expanding their operations. The training programme was unwritten, and it would be my job, during my observation, to ask questions where I was uncertain and to feed back as to whether the process he had in mind was understandable and feasible for a new recruit. In essence, he wanted to turn the company from a small custom firm to one where anyone could use a process to achieve a good result for a client. We agreed I would come to the office once a week for a period of two months in September and October 2004 (a total of 8 visits). I signed a confidentiality agreement not to reveal any client business, and to anonymise the company and each of the employees.

During each session with the company, I kept what the ethnographers Emerson, Fretz and Shaw (1995) call jottings – brief form notes, where possible, and occasional verbatim comments. Since computers were widespread in the office, I used my own laptop for my jottings or, if necessary, I wrote them on paper. These I wrote up into field notes at the first opportunity. I also collected minutes and supporting documents for meetings I was in both electronically, if I had access to the file, and in paper format otherwise. I was interested in both the technology that was used in the office on an ongoing basis as the means of search engine optimisation and the process itself. As it turned out, this reliance on electronics was unfortunate, as soon after the participant observation was completed, my computer was stolen (as it happens, while interviewing a search engine producer). My jottings, field notes and the electronic documents I had collected were not backed up

elsewhere, so all this information was lost. What remained were the notes from my initial session, the paper documents I had collected, and a few jottings.

Without the detailed field notes I had made, I had to make some difficult choices, as it was impossible to re-do the research. I re-contacted my informants and asked for permission to interview each of the staff. During these interviews I asked about the processes and technologies they used in their jobs, and tried to clarify and check what I had learned from my earlier observations. I recorded and transcribed these interviews just as I using the same process as I had for my interviews of the search engine producers, described below. The interviews with the opotimisation staff and the documents and notes mentioned above then formed the basis for the analysis contained in Chapter 7. In addition, for that chapter I also attended two search engine marketing conferences and did further online documentary research, for example with SEMPO, the search engine marketing professional organization, as detailed in Chapter 7.

In order to proceed with the producer interviews, I developed a list of all the major search engine provision companies, as this was my primary focus. The list was based on my initial analysis of the industry structure and included Google, Yahoo, MSN, and AskJeeves. My intention was to approach each company with a proposal to undertake some observations and, as a fall-back strategy, a series of interviews. I hoped to do a more in-depth case study on a single company. I was particularly interested in Google, Yahoo, and MSN, as these are largest search providers, but I also made an effort to contact smaller providers and alternative providers (such as open-source search engines). However, due to the highly competitive nature of the industry, and the timing – Google was in the middle of its initial public offering (IPO) to the stock market in 2004 and meanwhile Microsoft was launching its own search engine (LiveSearch) – none of the companies would agree to have an observer present in their building. In fact, as one of the people in charge of relations with outside researchers for Google remarked to me, "There are floors in this building where even we can't go."

The denial of access to search engine producers was not entirely surprising. Access is a well-known issue in media production studies. While Tuchman (1973), for example, was able to spend time on a newsroom floor, media researchers have relied heavily upon

interviews in the case of hard-to-access production groups: Deuze (2007) used interviews to elicit production cultures from modern media workers; the Cantors (1992) interviewed Hollywood producers as part of a mixed-method, and Morrison and Tumber (Morrison & Tumber, 1988) based their study of war reporting on interviews with journalists covering the Falklands war. This is also true from a science and technology studies perspective, so while historical methods are common, interviewing was used to elicit data from missile engineers in MacKenzie's (1987) study.

My interviewees could also easily be classed as elites. While the people I interviewed were not of the same status as the hyper-wealthy philanthropists whom Odendahl interviewed (Odendahl & Shaw, 2002) or the CEOs and major institutional investors who formed the bulk of Useem's research participants (Useem, 1995), they are on a par with the lawyers and clergymen of Pierce's and Aldriges's studies (see Hertz & Imber, 1995). My interviewees were very well-educated (often with PhDs); many were wealthy by virtue of their participation in successful technology companies during the dot-com boom; and most had high positions within their company (CEO, head of division, or head of research) and were used to being in command. There was no doubt that I was "studying up," to use anthropologist Laura Nader's phrase (Nader, 1972), and that my interviews could be characterized as elite interviews, with all the attendant methodological difficulties. Odendahl & Shaw divide these into *difficulties of access, difficulties arising in the interview itself,* and *difficulties in analyzing the interview* (Odendahl & Shaw, 2002). Each of these aspects of interviewing becomes uniquely problematic when dealing with situations where the interviewee is of higher status than the interviewer. As I discuss my own strategy below I reflect on Odendahl & Shaw's observations.

Difficulties of access, according to Odendahl & Shaw, include difficulties in locating and contacting respondents who may strive to protect their identities; in organizing meetings with very busy people; and in the extensive preparation required for interviews. In this regard, they emphasise that elite interviews cannot be extensively structured in terms of sample: "Candidacy for elite interviews often cannot be planned in advance of the project; rather, it emerges as part of the fieldwork" (Odendahl & Shaw, 2002, p. 307). This describes well what I found in conducting my interviews.

My first goal was to identify potential interviewees. Although it might be difficult to achieve a comprehensive spectrum of elite interviewees, my goals were to interview at least one person in each major search engine and, ideally, to gain a series of interviews and possibly status as an observer inside one particular organization. I wanted particularly to reach those who were very influential in setting the direction of the technical core of the search engine, as I reasoned that they would be most influential in shaping the technology.

I had three strategies for reaching appropriate interviewees: First, the formal approach. I reviewed company websites in an attempt to establish the identity of the head of the search engineering group in Microsoft, Google, Yahoo and Ask Jeeves. If none was listed, I noted the name of the executive responsible for the division. I began with emails that introduced my project, or, where I could not obtain an email address due to company policies, I faxed a letter describing the project. I then followed up with phone calls, in many cases several phone calls, until I reached the correct person. My second strategy was through my own contacts: I called people that I knew in the business and asked if they could recommend others I might talk to. Finally, I used forums on the internet which are frequented by the technical elites, gathered names, and emailed them. Of my final tally of producer interviews, two were the result of formal letters, two were the result of my personal contact network, four were the result of contacts made from internet research and three were recommendations by another interviewee. Over a dozen approaches for interviews were declined. Unfortunately, at no stage did I possess what anthropologists call a "key informant" or someone who was willing to get behind my project and introduce me to others, and this meant that the process was very slow overall.

It proved impossible to embark on a lengthy programme of interviews or observation within search engine providers. Some of the companies were extremely concerned about commercial security; others said they were simply too busy with product development. While I succeeded in speaking to very high-level engineers and product managers, executives were reluctant to grant access to their staff, even after lengthy nondisclosure agreements had been negotiated and signed. I attribute this to two factors: first, the executives were used to being spokespeople for their division. To allow me to interview subordinates suggested that their version was in some way not comprehensive or not reliable. Second, they were extremely protective of their projects and did not wish to

draw people away from work in process. For these and possibly other reasons, the two agreed interview "series" in major search engines came to a halt after one or two interviews.

The interview setup process was also extremely time-consuming. One interview took seven months from the initial contact to the interview itself, including an initial fax, eight follow-up calls, an initial unrecorded telephone call with agreement for more interviews, a legal agreement, and about 20 emails. While this was an extreme case, most interviews took between one to three months to complete. Other researchers report this extended preparation time as well – Thomas (1995), for example, reports that it took him two years to set up interviews with two executives in a major manufacturing company.

My overall sample, therefore, is comprehensive in that it does indeed contain interviews from producers in each major search engine company that I initially targeted. In addition, I took advantage of the fact that those who no longer worked in the search engine industry were more likely to be forthright and therefore interviewed several people who had been involved in search engines in the late 1990s. Finally, I rounded out the picture by interviewing some people who were involved in other aspects of the search engine industry: optimizers, distributors, and commentators. In some cases they also introduced me to or gave me contact numbers for producers to interview. Overall, I conducted 18 interviews, each approximately one hour long, with staff at search engine producers, distributors, optimizers and commentators, in a range of job functions. Eleven interviews were with search engine producers and seven with distributors, optimizers, and commentators. These are summarised in Table 2. The interview schedule is included in Appendix D.

Table 2: List of interviewees

Interviewee code	Job Function	Organisation	Location
Mr A	Founder and Head of Engineering at open-source search provider. Former senior engineer at major search engine.	Startup search provider	West Coast USA
Mr B	Senior Vice President of Technology	Major search provider	West Coast USA

Mr C	Former Head of Operations at early major search engine. Developer of another early search engine. Currently not working within search industry.	Major search provider	England
Mr D	Chief Scientist	Major search provider	West Coast USA
Mr E	Director of Product Development, Europe	Major search provider	London
Mr F	Managing Director	Major search distributor	Paris
Mr G	Project leader for specialized search	Major search provider	West Coast USA
Mr H	Project engineer for specialized search	Major search provider	New York
Mr I	Program Manager, Search	Major search provider	West Coast USA
Ms J	Development Producer with responsibility for overseeing search	Major search distributor	London
Mr K	Owner and Managing Director	Small search-engine optimiser	London
Mr L	Program Manager, Relevance	Major search provider	London
Mr M	Industry commentator	Independent	West Coast USA
Mr N	Responsible for technical operations	Small search engine optimiser	London
Mr O	Founder of small search engine. Founder and lead developer for early search engine; former fellow with special responsibility for search a large media conglomerate.	Small specialist search provider; formerly several major search engine providers	East Coast USA
Mr P	Responsible for sales	Small search engine optimiser	London
Mr Q	Founder and lead programmer of early search engine	Currently unaffiliated; formerly major search engine provider	West Coast
Mr S	Industry commentator	Independent	UK

The first column of the table shows the anonymous code I assigned to the interviewees. The second shows their place within the functions of the search engine. The third shows their level within the organisation. The fourth shows the organisational classification. Finally I show the location of the interviewee. All interviews outside London were conducted via telephone. In addition, all interviews with the exceptions of those with Mr G, Mr H, and Mr S were recorded. Mssrs G, H and S refused permission to record and

their interviews were written up from notes immediately following the interviews. There is a marked gender bias in the interviews in favour of men (17 out of 18 interviewees). In part, this reflects the small number of women in senior positions in this industry. Despite that, I was able to obtain the names and contact details of several very senior women in the search industry; regrettably, none would agree to speak with me.

Interviewees D and E worked at the same company, as did interviewees C and I, and H and L. All the search engine optimizer interviews came from a single company. All other companies are represented by one interviewee each. The companies that these interviewees worked for included Alta Vista, Aliweb, Ask Jeeves, AOL, Excite, Google, MSN, Webcrawler and Yahoo! among others. Some interviewees had worked for multiple companies.

Confidentiality was a major issue in securing the interviews. Producer and distributor interviewees were concerned that what they said might put their jobs or their company at risk by revealing company trade secrets or other competitive information. Those who were no longer employed at search engine companies were concerned that they might damage their friends' careers for similar reasons. The search engine optimizers worried that their clients' details might be revealed. Industry commentators were concerned for their relationships with sources. For many reasons, therefore, my interviewees requested anonymity.

Each interviewee was provided with a short description of my project (included as Appendix C) either in email or by fax. When I re-contacted them, I reviewed the description verbally, asked if they had any further questions, and provided further information when requested. I then reiterated the information about anonymity and confidentiality and told them that, if at any time they wanted to withdraw from the research, even after they had agreed to it, they were free to do so. None did. I asked for permission to record the interviews and this was granted in most cases. In two cases, I also signed specific legal nondisclosure agreements (NDAs) in addition to giving the participants an informed consent form in advance. In one case, the NDA obligated me in addition to refrain from discussing business pertaining to the clients of the company in question; in the other case, I agreed to refrain from asking questions about trade secrets

and to give any material to the company to see 10 days prior to publication, although without them having any rights to change the content.

The confidentiality agreements meant that I assigned each interviewee an anonymous code, and, in addition, created a set of anonymous designations for the companies they mentioned in order to protect my interviewees. On the occasions when I used transcriptionists, I also signed non-disclosure agreements with them, did not reveal the names of the interviewees, and exchanged all files and transcripts in an encrypted format.

Accounts of elite interviewing emphasise the difficulties arising within the interview itself in terms of setting the location (nearly always in a setting which the interviewee controls) and of ceding control of the interview to the interviewee (Kezar, 2003; Odendahl & Shaw, 2002; Thomas, 1995). Again, my experience bore out their observations. In general, the interviews were held on the telephone since face to face interviews were declined; I had to explain once again the purpose and goals of the research; anonymity and confidentiality had to be assured as discussed above; and, in many cases, I was questioned on my personal status or quizzed about technical matters. My interview schema (included as Appendix D) was semi-structured and, although most topics were included in most interviews, the interviewees also wanted to use the time to comment on the worthiness and purpose of the research.

According to Odendahl & Shaw, in an elite interview:

> "…the interviewer must establish his or her own authority to ensure a productive exchange. There are subtle ways in which an interviewer can communicate expertise either in the field under study or in knowledge of other prominent players within it. These may range from immediately handing over a business card that indicates higher degrees, institutional position, and title to name-dropping around either individuals or projects that can validate the interviewer to the subject."
> (p 311).

In my case, again and again I found that my credibility was established when I discovered that the interviewee and I had acquaintances or friends in common, generally from my time working in search engines in the late 1990s. Sometimes, however, it was the LSE name that secured the interview on my behalf, particularly in the formal approaches.

3.4.2 DATA ANALYSIS

When the interviewees agreed, I recorded my interviews. This was done using a digital recorder (Olympus model DS-330) that connected directly to my computer. I transcribed half the interviews myself, and the rest were sent for transcription. When the interviewees did not agree to be recorded (three cases), I took notes instead and wrote up an account of the conversation the same day.

Once I had the transcripts from the search engine producers I analysed them using a coding scheme derived from my conceptual framework along with some concepts derived from discourse analysis, as I discuss below. The first step in analysing the interviews was to develop a working definition of an "interpretative scheme." Giddens suggests that interpretative schemes "offer accounts" or "give reasons" and Orlikowski & Gash discuss how in technology analyses the technological frame elucidates "assumptions and knowledge" (Orlikowski & Gash, 1994). This suggested that statements of fact and narratives of how things happened and why were the areas in which interpretative schemes could be most reliably uncovered. Therefore, I began the process of coding each of these themes by identifying likely segments where the "interpretative scheme" might be in evidence. As I began my analysis of the quotes, however, I was faced by an immediate difficulty.

Orlikowski and Gash (1994) suggest that different groups – in their study examples included consultants and technologists – each have different technological frames of reference and that the important element lies in the *congruence*, or non-congruence, of those frames. Bijker, similarly, suggests that by definition each relevant social group shares a technological frame or a "shared cognitive frame that defines a relevant social group and constitutes members' common interpretation of an artifact" (Bijker, 1995, pp. 125-126). However, in my work even preliminary analysis of the interviews suggested that multiple competing accounts were present in each group, and that unitary concepts of a technological frame were not a good fit for my data.

A better fit seemed to be apparent when I turned to the discourse analysis tradition to the "interpretative repertoire" concept suggested by Potter and Wetherell (1987). They argue that each speech act is an achievement on the part of the speaker and that different

repertoires have different functions for speakers and will appear in different contexts[23]. For example, Potter & Wetherell review an earlier study by Gilbert and Mulkay (cited in Potter & Wetherell, 1987, p. 147 ff) where scientists talk about the quality of experimental science. In their investigations, Gilbert and Mulkay found two interpretative repertoires which they named the "empiricist" repertoire and the "contingent" repertoire (Gilbert & Mulkay, 1984). The empiricist repertoire stresses the primacy of scientific data in accounting for the acceptance of one theory over another, while the contingent repertoire stresses external factors such as personal characteristics or political affiliations. Each repertoire in their study was deployed in particular circumstances – for example, in formal journal submissions versus informal contexts, or within the same conversation to explain contradictions arising from inconsistencies in a single repertoire.

I therefore grouped the quotes I had identified into a small set of schemes. These schemes were then applied as codes to similar segments in all the interviews, with a further "undefined" scheme serving as a place for quotations that did not match the initial schemes. In order to group and identify the schemes, I looked for strong or unusual verbs or nouns with particular connotations. For example, when one interviewee said "I fought in the search wars" (Interview Q) the use of the verb "fought" and the noun "war" made me confident in assigning that to the previously identified "war" scheme. Most other examples were not as clear-cut. However, when my interviewees talked about measuring, experiments, and discoveries, and referred to themselves as researchers or engineers, I assigned those segments to the "science/technology" scheme. When they talked, on the other hand, about costs, profits, competing and success, identifying themselves as employees or as owners, I assigned that quotation to the "market" scheme. I then examined the elements of the interviews I had identified for the ways in which they characterized the answers to my sub-questions: how the producers conceive of search engine results, how decisions are taken and what "quality" was understood as. I used the qualitative data analysis computer program Atlas.TI version 5.0 to assist me in coding. At the end of the process, two major and two minor schemes were identified. The schemes

---

[23] Discourse analysis has a long tradition and a large literature. Common to most discourse analysts is the idea of language as a "speech act" or practice. Some discourse analysts are primarily interested in how the act is achieved (for example, in the tradition of conversation analysis). Others focus on the use of language in the service of ideology, for example, in reinforcing structures of domination (the tradition of critical discourse analysis). The strand of discourse analysis drawn upon here emphasises the potential strategic quality of discourse without necessarily adopting a political position and may be referred to as a sociological discourse analysis.

are discussed further in Chapter 6. I reflect more fully on the theoretical implications of expanding on the idea of "technological frame" to include multiple discourses or interpretative repertoires in Chapter 8.

By treating the interviews with producers as texts, rather than as 'facts,' I sought to reveal different modes of communication which are themselves partially constitutive of the community of practice in which they operate and which might reveal conflicts and tensions, rather than to discover a 'truth' about how search engines are programmed.

The remaining interviews – those with search engine optimizers, commentators and distributors – were by contrast not treated as discourses. Instead, they were analysed together with the documentary material gathered through participant observation and the remaining notes from the participant observation to develop an outline of search engine optimizer interactions with search engines, presented in Chapter 7. In order to do this, the "modes of interaction" discussed in the conceptual framework in Chapter 2 (sanctions, power, and communication) were linked through the idea of the boundary object, also introduced in the conceptual framework. Interactions between the optmisers and the results in the course of their daily work were noted; also noted were other direct interactions with the search engine companies. These interactions were then categorized as to the mode of interaction and whether these interactions had a conflictual or a co-operative character. Finally, the commentator and distributor interviews formed part of the contextual material used in Chapter 4.

3.5 Level B: Structures

In many ways, the data gathering and analysis for the second level of the methodological strategy, which is concerned with historical and material aspects of the search engine industry, was more straightforward than interviewing the agents. All the documents I needed were available electronically, either stored publicly on the Web in the case of financial filings, or indexed by Lexis/Nexis in the case of press articles.

Like many other analytical traditions, the field of the political economy of media and communications has no canonical methodology. Research methods are conspicuous by their absence in Mosco's comprehensive review of the field (Mosco, 1996). Researchers

working in this tradition sometimes use analytical methods drawn from industrial economics, media and communication studies, organizational theory, sociology and political science. They often take as the starting point of their investigations the economic and material base of media and communication activity, which then forms the basis for their analysis of structural power relations[24]. Data sources may include quantitative data such as financial reports, labour statistics, and network ratings shares, large-scale analysis of media content and qualitative data such as information regarding corporate functioning submitted in court cases and to government bodies (for example, in the case of anti-trust investigations), textual analysis of media texts and interviews.

Two quite different examples illustrate this methodological diversity – one from a political economic study of bias, the other from a political economic study of information technology. Both of these examples have some topical affinities with my own work. First we can take Herman and Chomsky's work on a "propaganda model" of bias in US journalism using a political economic framework (Herman & Chomsky, 1994). Their methodological strategy is twofold. First, by means of institutional analysis including the analysis of ownership, revenue sources, production figures and policy documents, Herman and Chomsky build a picture of a series of complex and interlocking interests of elites who own, fund (through advertising), provide experts to, and criticize the media when necessary. This analysis provides an explanation of the "market forces" (Herman & Chomsky, 1994, p. xii) through which the propagandistic function of the media is said to operate, and constitutes the propaganda model proper. Second, the insights of the model are analysed using content analysis of paired examples: news coverage of "worthy" versus "unworthy" victims, "legitimizing" or "meaningless" elections, or the war in Vietnam versus the wars in Laos and Cambodia. A second method, textual analysis, "consider[s] the spectrum of opinion allowed expression," (Herman and Chomsky cited in Klaehn, 2002) with the aim of discovering the scope of the debate encompassed by the media.

---

[24] Wasko, for example, describes a political economic study of film as follows: "Fundamentally, [it] analyzes motion pictures as commodities produced and distributed within a capitalist industrial structure…Most importantly, the political and ideological implications of these arrangements are relevant, as film must also be placed in an entire social, economic and political context and critiqued in terms of the contribution to maintaining and reproducing structures of power" (Wasko, 2004, p. 132)

As a second example, an institutional economic point of view is offered by Mansell and Javary (2004) and Javary (2004) in analyzing the UK internet service provider (ISP) market. They analyse the role of financial capital and governance processes in the ISP market, using data sources such as financial documents and press releases to reveal common investors, partnership agreements and interlocking directorships, as well as strategic maneuvering on the part of entrepreneurs and established service providers such at British Telecom. This analysis is then contrasted with the stated policy goals, and leading to suggestions for a new legislative framework.

I chose these two examples to represent some of the diversity in political economy studies of media and communication, as they include novel analyses of publicly available information – financial documents, press releases, published interviews, industry reports, etc., often mixing qualitative and quantitative methods. The results are then often contrasted with analysis of policy documents and published statements to critically examine institutional practice. I adopted a similarly multi-method approach in the present study. I used financial statements prepared for stock exchange authorities, press releases, reports in the trade press and interviews to build up a picture of the financial and geographical organization and history of the search engine industry. I supplemented this with information from secondary sources such as ratings data, and interview data. In addition I reviewed technical papers to develop a technical timeline related to the industrial timeline.

In investigating the structural issues surrounding bias, my first goal was to answer general questions about the search engine companies as an industry and the way in which economic and other kinds of power might be concentrating in various firms. I therefore began by reviewing ratings data from Nielsen//NetRatings and ComScore, two commercial online ratings agencies, in order to locate the largest search engine sites in terms of traffic. I also reviewed lists from SearchEngineWatch.com, an online industry commentator, and I used reports from the Internet Advertising Bureau, an industry body, to assess the overall size of the online advertising market. From these sources, I developed a chart that showed the major distributors and suppliers of search engine results at the time. As time went on it became clear that developing a further chart of

mergers and acquisitions would be more helpful as the exchange of results between different search engines became less common (see Figure 5 in Chapter 4).

 I then went to the websites of the major organizations that figured in the initial chart – Microsoft, Google, Ask.com, Yahoo, Terra Lycos, LookSmart and AOL Time Warner – and downloaded their 2003 financial reports, where available. Google at that time was not public, so I used the S-1 form prospectus which has a similar format. Terra Lycos information was only available for 2002. Once I had the financial reports, I reviewed them to understand the role of advertising in the overall business. This step first involved separating the search engine properties from other elements of the business – for example, MSN from Microsoft, AOL from Time Warner's other businesses, and Lycos from Terra Lycos's telephony business. I then read through each report and noted the different types of advertising vehicles they listed and the relative amount of income for each, where available. I also noted the whole income for the division or company. Where significant acquisitions had been made – for example Overture's acquisition by Yahoo! in 2002, I used financial data provided prior to the acquisition to estimate the contribution of the new subsidiary.

These data formed an important part of my initial picture of the industry as one with multiple interconnections where major funding was derived by advertising. Initially, I intended to develop a view of the industry as a snapshot, documenting connections between the major players in 2002 when I undertook the initial research. However, the industry was in a period of transition at that time. As a result, I shifted my focus to an historical analysis of the search engine business model and its context as discussed in Chapter 4. This analysis extends over more than a dozen years, from the start of the Web in 1993 (when necessary making reference to previous history) to 2009. In order to develop this history, I used three major sources. First, I reviewed press releases from the companies themselves. Many of these are available online. For example, I downloaded the stock of press releases from Yahoo!, Google, and AskJeees, in addition to relevant releases from Microsoft. In some cases, press releases were not available – for example, from the major search engine Excite which was acquired by @Home and then by AT&T and has now effectively vanished. In these cases and also to provide different views to the official story, I used press releases captured by the Internet Archive

(http://www.archive.org) and also contemporary press sources[25]. These included general press such as *The New York Times* but also specialist press like *Advertising Age*, *Wired* magazine, *New Media Age* and online publications like *Search Engine Watch* and blogs which concentrate on search engines such as John Battelle's "SearchBlog." I also included local reportage, such as stories from the *San Jose Mercury News*, the local paper of the Silicon Valley area. The traditional publications were accessed via Lexis/Nexis. The rest were gathered through a more time-consuming process of searching through archive and internet listings. The total corpus was well over 1000 documents, including a huge volume of corporate press releases (over 600). I also read, watched, and listened to interviews conducted with key personalities, for example, Mike "Fuzzy" Mauldin (Devlin, 1996), who created the Lycos spider; Sergey Brin and Larry Page, the founders of Google (Correa, 2000); and Matt Cutts, a Google engineer who often speaks publicly (Abondance, 2002; Grehan, 2006).

I made an attempt to divide the revenue of the main companies by geographic location. However, this proved impossible because the figures were not shown in enough detail. For example, advertising revenue was often included with all other revenue in "worldwide" figures for divisions outside the United States (this is a common problem in the analysis of industry revenues where firms operate outside the countries of incorporation).

I undertook an examiniation of the search engine industry in a global context and specifically by tracing the history and provision of search services in four countries as of 2006: China, Japan, Germany, and South Africa. In 2006 China, Japan, and Germany had the largest internet populations and, collectively, accounted for 22.9% of the world's internet population (as measured by ComScore, see Table 7 in Chaper 5); by focusing on these countries and the US the present study examines the search provision for approximately 45% of the total world internet population as of 2006. In addition to these

---

[25] The reconstruction of internet history is a tricky business. Unlike many traditional companies, dot-com businesses by and large do not have official archives. The Internet Archive, as an external project, captures only the public face of the company and only a sample of pages; in any case, it was not in place until 1996, after many of the important developments in search engine history had already taken place. Links between pages are frequently not maintained, with the target of the link either being deleted or shifting to an alternate Web address. I took the precaution in many cases of both printing the page and downloading the entire text of an online blog or article into my referencing software to preserve the reference in case it should be removed.

large countries, the experience of a small country swas included as a basis for comparison and I chose to add South Africa which had the largest internet population of any African country in 2006. Case studies for each country were developed primarily by using press releases and resources found on Lexis/Nexis as well as by drawing on treatments of the overall history of the internet. This aspect of the research was particularly challenging, first, because of language issues and, second, because so little research has been done to examine the accounts which the companies themselves give. This leaves the would-be scholarly chronicler vulnerable to potential misleading claims. The accounts given in Chapter 5 of the global development and current position and strategies of the search engine companies in these countries were therefore been read by colleagues based in these countries with an interest in the internet, to minimize the potential of offering a misleading account of the development of these markets.

My interview data was used in cases where the interviewees commented directly on issues revealed in the financial analysis. I also drew upon my observation of the search engine optimizer company in my interpretation of the data, for example, in developing my argument about the importance of traffic (see Chapter 4). In contrast to the discourse analysis (see Chapter 6) the interviews were treated as offering insights into the analysis in Chapters 4 and 5. That is, the were used as texts but as statements of belief or fact. Other interpretations of this data may be possible, although I endeavored to draw on the data that seemed most representative of the majority opinion amongst my interviewees.

The foregoing outlines the way this current study was conducted using two levels of analysis, the first concerned with agents and their actions, and the second concerned with structural elements of the industry. Interviews, documents and financial data were analysed to build up a picture on these two levels. The analysis in Chapter 8 integrates the two levels. The next section discusses some ways in which I might alternatively have designed and carried out the research.

3.6 Alternative research designs

The chosen research design is, of course, not the only possible design which could be devised address the question of how bias arises into search engine results.

In Herman and Chomsky's work, mentioned above, the methods used include content analysis and textual analysis of news items. A comparable study in this case might have been the study of the search engine results themselves. Herman and Chomsky seek to convince skeptical readers that bias in news exists, what shape it takes, and how it could be related to the "filters" that they propose. I did not deem it appropriate to include an analysis of search engine results in this thesis. The first reason was that studies of this type already exist within computer science as part of the general assessment of the effectiveness of current search engines (Mowshowitz & Kawaguchi, 2002, 2005; Vaughan & Thelwall, 2004). Duplicating these studies would have provided little new insight. The second reason was that these types of studies require large investments of computer resources and skills and this would have distracted me from my focus on the social and economic sources of bias.

Another approach would have been to undertake an assessment of the design of the search results page as a means of discovering of potential bias. I have instead relied on studies that aim to assess search engine skills or "literacy" in order to include an understanding of the some of the implications of search engine page design (Hargittai, 2002; Hölscher & Strube, 2000). An original study of this kind was not included because to assess this properly, panels of users would have had to be convened and the focus of this thesis would have shifted from the potential causes of bias to assessing the nature and even the existence of the bias, for example, as compared to the skills of users. The instability of the search engine 'text' means that it would have been interesting to track the changes and developments of the search engine results over time in response to similar queries (see Bar-Ilan, 2000). However, while the Internet Archive[26] records HTML pages well, it does not keep a record of the dynamic elements of websites: for example, associated advertisements, the effectiveness of the ranking algorithm, etc. The empirical difficulties of conducting this type of analysis are quite substantial.

Both Giddens and Orlikowski recommend or use 'ethnographic' investigations of organizations and these are well-documented in the literature on the social construction of technology. A long period of observation within a search engine company would have been ideal and, indeed, this was the initial research method. However, with so few

---

[26] Available at http://www.archive.org

companies to choose from, the issue of access was critical. I was unable to secure entry into a search engine company in the requisite timeframe. However, while I was unable to spend a significant amount of time on a single case company, I have secured information from a considerable range of search engine sources.

3.7 Conclusion

Every doctoral study is a combination between what would be ideal, what is achievable during a given period of study, and what is feasible in the field. This study is no different.

In this chapter I have developed the research design and methodology for the of investigation. I have outlined a dual strategy of investigation based on the "orienting" perspective of structuration theory set out in the discussion of the conceptual framework in Chapter 2. The first level of investigation concentrates on agents, and the second level of investigation focuses on the industry structure of the search engine companies. The next chapter begins the empirical analysis of search engine bias by presenting the development of the search engine industry.

IV

The History of the Internet Search Engine

Navigational Media and the Rise of the Traffic Commodity

4.1 Introduction

This chapter begins the empirical investigation into search engine bias. Chapter 2 outlined two insights from the political economy of media and communication which can be used to shed light on search engine bias: the processes of *commodification* and *spatialization*. This chapter examines the issues of commodification, while Chapter 5 deals primarily with spatialization. The questions that this chapter addresses were spelled out in the research design in Chapter 3, Section 3.3, and include the following:

- What is the ownership structure of the industry? Is the industry concentrated into certain companies? How has it developed over time?

- What is the role of advertising in the industry? How much does it contribute to revenue? How does it operate?
- How does the search product fit into the overall financial structure of the company? What contribution does it make to revenue?

The answers to these questions, as we shall see below, have changed during the short lifetime of the Web. Companies that began by selling software quickly switched their focus to selling audiences via advertising and finally to selling traffic, which, this chapter argues, is the quintessential new media commodity.

Following from this analysis is the observation that internet search engines are in some respects better analysed as media than as technologies. In 1999, the political economist Dan Schiller, citing examples from Yahoo! and Infoseek among others, argued that "[W]e must locate the internet within the evolving media economy. We must learn to see how it fits within, and how it modifies, an existing forcefield of institutional structures and functions." (D. Schiller, 1999). This chapter also therefore considers in what ways traditional media and communications institutions are involved in the search engine business and vice versa, and thus, how search engines have evolved over time to be part of the media economy.

The chapter builds upon the insight that in order to analyse the search engine industry, we must look at the *supply chain for audiences* rather than for content (e.g., news stories or television productions) as is common in analysis of media (Doyle, 2002). Online, it is relatively simple to produce content – what is considerably more challenging is to attract an audience. With the transformation of the supply chain we can understand the history of search – for example, the otherwise puzzling failure of the large media conglomerates to dominate the search engine industry as they attempted to do – which is based on the creation and exploitation of a new commodity for media: traffic.

The chapter takes the format of a chronology of the search business, which is divided into three periods: first, the creation of the first search engines and the period of technological entrepreneurs in the mid-1990s, resulting in a competitive market of relatively small companies; second, a period of portals and vertical integration in the late 1990s which saw many search engine acquisitions by traditional media and telecom companies; and third, a period from 2002 onwards characterised by the exit of traditional media and telecom

companies and a period of consolidation. Today's search engines are not vertically integrated, but have developed an immense network of alliances, both forward and backward, along the audience supply chain which arguably form a strong, stable and flexible base from which to defend their business position given the rapidity of technical change – a kind of "virtual" integration which nevertheless poses strong barriers to entry.

4.2 The development of the search engine industry

James Curran (Curran & Seaton, 2003:250) argues that the internet from the mid-1990s onwards entered a commercialised phase in which mainstream companies – in particular large media conglomerates such as Bertelsmann, Vivendi, Time Warner, News International, and Disney – began to dominate the Web, owning three-quarters of the most visited news and entertainment sites. The present study, by contrast, found that large media firms are conspicuously absent from the major search engine providers (which are also the most highly visited websites) in 2006, that is to say Google, Yahoo, and Microsoft. In fact, the only large media conglomerate to be represented in the top fifteen properties is Time Warner (most likely its huge ISP and online service provider AOL).

Figure 6 presents in diagrammatic form the development of the major internet search engines of the past dozen years since the invention of the Web. The chart consists of three periods: first, a period of *technical entrepreneurship* from 1994 to late 1997; second, a period which was characterised by *the development of portals and vertical integration* from late 1997 to the end of 2001, in which major media companies and network providers attempted to buy their way into the search arena; and finally a period of *consolidation and "virtual" integration* from 2002 to the present day. While presented as analytically distinct, these three periods of course overlap to a certain degree; for example, it is certainly possible to find technical entrepreneurs in the middle period (Google and Overture are excellent examples), and attempts at consolidation in the early period (e.g., Excite's early acquisition of Magellan and WebCrawler).

The periods into which I have classified the short history of search engines are essentially based on shifts in revenue models and ownership, and give primacy to the economic history of search over its technological history. Clearly technological innovation is also important; and indeed, the shifts in revenue and economics closely coincide with

technological developments and are related to pre-existing structures for capitalising on technology. But a history of technological "successes" is not sufficient to explain the dynamics of the search market, nor can it adequately characterise an industry that generated some $12 billion[27] in 2006.

Of the twenty-one search ventures listed in Figure 6, only six remained independent entities at the time this analysis was prepared, in 2005. Of these, only four produce algorithmic search results of the whole Web: Yahoo, Google, MSN, and Ask. As regards the remaining two, Lycos no longer operated a search engine, but purchased search from Yahoo, and LookSmart no longer operated its own directory, but had transformed into a provider of paid search results.[28]

---

[27] All values are given in US dollar unless indicated otherwise.

[28] In other words, they search an index of advertisements placed by website owners, rather than an independent index of results generated by crawling the web.

Figure 6: Search engine mergers and acquisitions in the three periods of search history.
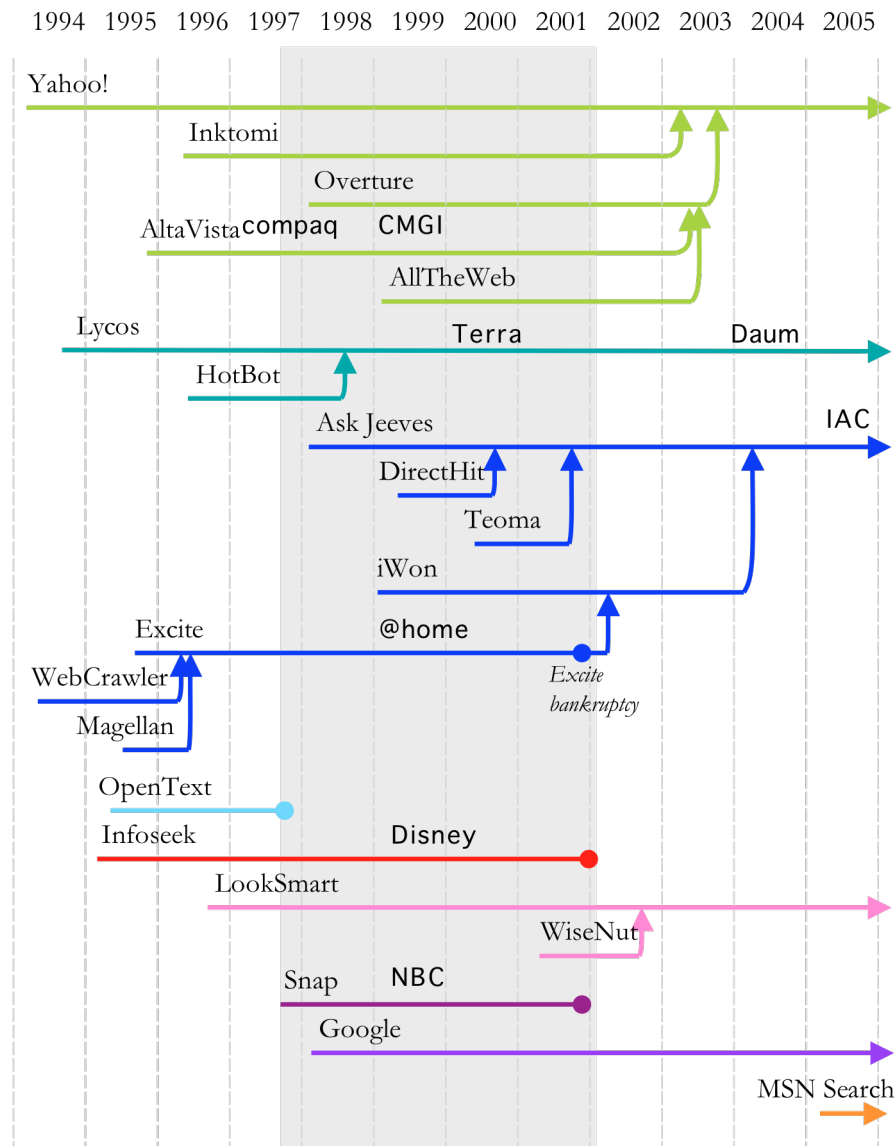
Source: Data from company websites and press reports, compiled by author.

4.3 Technological entrepreneurs (1994-1997)

The history of modern search engines begins in the non-commercial setting of the academy or research institution. Search engine technology primarily developed from the academic discipline of information retrieval, which itself is something of a hybrid between library science (now often called information science) and computer science. From information science, information retrieval draws theories of information categorization

and the human cognitive process in information seeking. From computer science and artificial intelligence springs the desire and the ability to automate catalogue creation and information retrieval from catalogues (see Singhal, 2001 for a short overview of the development of information retrieval as a field). It is no surprise, therefore, that most of the earliest search engines were created in computer science research laboratories, primarily in academic institutions. Table 3 shows the earliest search engines and their locations, organized chronologically[29].

Table 3: Early period search engine dates, institutions, and founders.

Engine/ Directory	Date went live*	Institution (Location)	Developer(s)	Position at time of development
Yahoo (directory)	Feb 94	Stanford University (Palo Alto, CA)	Jerry Yang David Filo	Computer Science (CS) PhD students
WebCrawler (engine)	20 Apr 94	University of Washington (Seattle, WA)	Brian Pinkerton	PhD student in CS
Lycos (engine)	Jul 94	Carnegie Mellon University (Pittsburgh, PA)	Dr Michael Mauldin and Bob Leavitt	Postdoctoral research fellow in CS
Infoseek (engine)	13 Feb 95	n/a (Sunnyvale, CA)	Steve Kirsch	Serial technology entrepreneur – founded Frame Technology and Mouse Systems. BA and MS from MIT.
OpenText (engine)	Apr 95	n/a (Waterloo, Ontario, Canada)	(uncredited, possbly OpenText VP of Information Retrieval Larry Fitzpatrick)	Early provider of search interfaces to products such as Oxford English Dictionary
Magellan (directory)	Aug 95	n/a (Sausalito, CA)	Isabel & Christine Maxwell	Daughters of publishing magnate Richard Maxwell, originally published a print guide to the Web

---

[29] Not included in this chart are Archie, a pre-Web search engine for FTP sites developed by McGill University student Alan Emtage in 1990 and Veronica, a similar engine for Gopher sites, developed at the University of Nevada in 1993. Also excluded are the first two Web search engines, the WWW Wanderer, the first spider to crawl the web, developed by Matthew Gray, a researcher at MIT, in 1993, and Aliweb, developed in 1993 by Martijn Koster while he worked for Nexor in Nottingham, England. Neither of these technologies were commercialised.

			Graham Spence	
Excite (engine)	*29 Sep 95*	Stanford University (Palo Alto, CA)	Graham Spence Joe Krausz Ben Lutch Ryan McIntyre Martin Reinfreid Mark Van Haren	Recent CS graduates (apart from Krausz who graduated in political science)
AltaVista (engine)	15 Dec 95	Digital Equipment PARC (Palo Alto, CA)	Dr Louis Monier	Research fellow
Inktomi (engine)	20 May 96	University of California at Berkeley (Berkeley, CA)	Dr Eric Brewer Paul Gaulthier	Assistant professor of CS and graduate student
LookSmart (directory)	28 Oct 96	Reader's Digest (Melbourne, Victoria, Australia)	(uncredited)	(uncredited – presumably the publishing team acting through ordinary channels?)

*Dates refer to when the search engine became publicly accessible.
Source: Data derived from original press releases and news reports, compiled by author.

In these early search engines, two alternative models of service provision can be seen. First, the *Web directory* provided groups of sites that were categorised, and in some cases rated, by an editorial team. Examples of the directory strategy included Yahoo!, Magellan (who pioneered editorial ratings), and LookSmart. The second model was much more complex technically, and involved used automated technology to browse websites, store them in an electronic index, and automatically retrieve them based on user queries. These were more properly called *engines*. The two main axes of technical competition at this stage were the size of the engine or directory index and the speed of retrieval.

Early search enterprises had three primary sources of revenue: venture capital, product licensing, and advertising. Later, money raised on the stock markets would help to fund the business. In particular, venture capital was absolutely crucial, since during this phase of technological entrepreneurs, no one was exactly sure how the business would be funded – that is, whether the licensing and advertising revenues would prove eocnomically viable.

Just how uncertain the business model of internet search was is emphasised in an interview with the first Chief Financial Officer (CFO) of Lycos, Ted Philip:

> "We didn't have a model to follow," Philip recalled. "There was no such thing as advertising on the internet at that time…We had no business plan. All we had was a piece of technology." (quoted in Gavetti & Rivkin, 2004, p. 15)

Vinod Khosla, the Silicon Valley venture capitalist who gave seed funding to Excite, says the same: "I had to develop a complete business plan. Being a navigation service for the internet wasn't originally on the list of what they wanted to do. " (quoted in O'Brien, 1997). The Yahoo! founders expressed similar sentiments (Battelle, 2005, p. 59). Even those who did have a revenue plan, like Infoseek, weren't able to make it stick. Infoseek's initial $9.95/month subscription plan, which included a hundred free queries and ten cents per query after that (Infoseek, 1995a), quickly crumbled in the face of free services from Lycos, Yahoo, WebCrawler and Magellan.

The business model that most eventually decided on was a mix of advertising and licensing. Webcrawler began taking limited sponsorship on December 1, 1994 (Pinkerton, 2001). On May 22, 1995, a short three months after its debut, Infoseek announced that it was introducing a new free service supported by advertisers[30] in addition to its subscription model (Infoseek, 1995b). It later claimed to have introduced cost-per-thousand (CPM[31]) advertising pricing to the Web[32] (Infoseek, 1997). It certainly was the first in the search market, and it was quickly imitated. Carnegie Mellon announced in June 1995 that Lycos would become a commercial company in partnership with CMGI Ventures (a venture capitalist). It would "offer advertising space on its site and [would] license the catalog as well as key technology components" (Carnegie Mellon University, 1995). Just nine days later, Yahoo! announced that it would, as founder Jerry Yang put it, "make a graceful transition from being a not-for-profit hobby into a professional commercial service" (Yahoo!, 1995). It debuted with five advertisers in a three-month trial. Magellan followed suit in October of 1995.

Thus by the time the second wave of search pioneers – AltaVista, Excite, Inktomi, and LookSmart – launched their services, advertising was already widespread on search

---

[30] Original advertisers were Sun Microsystems, Storage Computer and the Internet Shopping Network.

[31] The "M" in CPM is the Roman numeral for 1000.

[32] CPM pricing essentially charges a fixed cost – say $10 – for every one thousand viewings of an advertisement; sponsorships, on the other hand, are typically paid at a fixed price irrespective of the numbers of people who actually view the advertisement.

engines. However, a second revenue stream was also clearly being developed. OpenText, one of the few companies that preceded the Web, based their plan on primarily on software licensing, as did Inktomi, which launched with a deal from Wired Digital to operate its new "HotBot" search engine.

In fact, licensing was in many ways the preferred model for many of the entrepreneurs: licensing was a known software business model, with predictable, ongoing revenue. Advertising was much more linked to Hollywood than Silicon Valley. Nevertheless, advertising predominated, possibly because the number of companies who wanted to license search engine technology was limited. Advertising, however, was driven by usage (especially after the introduction of cost per thousand, and later cost-per-impression[33] pricing), and the licensing model played a part here as well – many companies quickly understood that by giving or licensing their products to large traffic source – ISPs, for example – they could quickly build up usage. Distribution deals of this type proved critical, and there were no more important sources of traffic in the early days of the Web than Netscape and AOL. These two companies, while never themselves developing search technology, were crucial in the early development of the search and navigation industry. Each of the major players partnered with one or both of these companies and in so doing secured enough viewers to keep their advertising revenue high and the company solvent until their initial offerings on the stock market.

These public offerings, in turn, brought an influx of new cash to the search engines which funded their later expansion. The level of cash generated for such young businesses was unprecedented, as a contemporary account of the Yahoo IPO from the *Financial Times* shows:

> "Definitive proof of the scale of the internet craze comes in the $1.1bn market capitalisation briefly accorded last Friday to Yahoo, and electronic catalogue of the World Wide Web. So egregious is the overvaluation…that it is hard to convey in the FT's sober prose. This is a company with total revenues of around $3m since its launch in March 1995…[it] has achieved an operating profit ($62,000) in only one of its

---

[33] Cost-per-impression or CPI pricing charged a small sum (2¢ to 6¢, according to Yahoo's 1996 Annual Report) for every viewer. This was made possible by the accurate tracking of Internet servers as opposed to the more general audience measurements available for print publications.

four quarters…[and is] run by Jerry Yang and David Filo…[who] have no previous business experience." (Martin, 1996)

Indeed, Yahoo! was one of the defining companies of the internet boom period, to which we now turn. However it is worthwhile noting in passing that despite the 2001 market crash in high-tech stocks, the "internet craze" continued: as of 20 March 2006, Yahoo's market capitalisation was $46.6bn, over forty times its "egregious overvaluation" of a decade earlier.

This first period of search engine history, then, is characterised by technological innovation within research centres followed by commercialisation using advertising and licensing as business models and capitalisation through venture capital and the stock market. The market was competitive, consisting of multiple companies with different technologies.

4.4 Portals and vertical integration (1997-2001)

The middle period of the short history of search engines online comprises the heart of the dot-com boom and bust period, that is to say late 1997 to late 2001. It is characterised by the change in focus from search engines to "portals" and the involvement of traditional media and telecommunications giants in the sector. If the first period of search can be characterised by technological innovation and the establishment of a vibrant, competitive marketplace for search technology, in this second period the search engines become focal points for a struggle to control the internet as a whole on the part of traditional media companies and telecommunicationsd providers.

This period in the history of search is notable for two related dynamics, which sometimes work together, and sometimes in opposition. These are: first, the growing technical opportunities for content integration; and second, the related idea that a proprietary "walled garden,[34]" or secondary internet, could be created which might to be owned by a single company.

---

[34] A "walled garden" refers to an enclosed, protected space, and this metaphor was extended to mean a online area based on internet technology where content was controlled, arguably for the benefit of the visitor.

In order to understand these dynamics, we can use the vertical supply chain as a means of analysis. The vertical supply chain is a tool for analyzing an industry whereby activities are ordered in a sequence, which starts at the early stages of production and works its way through the various intermediaries until arriving eventually at the customer (Doyle, 2002, p. 18). Doyle has recently defined a vertical supply chain for media as consisting of three general phases: production, packaging, and distribution. While generally useful, the supply chain is particularly helpful in understanding the dynamics of search engines at this time – but only if we change its focus, as follows.

The generic media supply chain is based upon taking *content*, that is to say, television broadcasts, news stories, pictures, etc., as the basic unit of analysis. Most traditional media companies have some element of vertical integration along this chain. So, for example, Time Warner owns production companies, networks, and cable television stations. However, it is clear that media companies operate in what is called a dual product market. On the one hand, they sell content to audiences – this is the content supply chain that Doyle describes. On the other hand, however, media companies sell *audiences* to advertisers. On the internet, where audience is extremely fragmented, this turns out to be much more useful supply chain to construct, since the problem is not so much getting content to your audience (a basic Web page being quite easy to construct) but audience to your content. Thus, what we need is not a supply chain for media content, but a supply chain for media audiences.

To construct such a chain, we must begin by considering how audiences get on the internet. First, they must have a computer, and the software to make it run[35]. Hardware manufacturing and software providers are therefore the first two steps in the chain. Second, they must connect to the internet via some kind of an internet service provider whose signal will run over telephone lines (or, possibly, cable lines). The telephone or cable company and the ISP are therefore the third and fourth steps in the audience supply chain. Fourth, they need a browser to access the World Wide Web. In the early days of the internet, the browser was seen as the crucial point for audience aggregation. When Netscape went public, it was this insight that drove its market price sky high. Finally, in

---

[35] Of course, today some audiences access the Internet without having a computer – for example, from mobile phones. However, during this period, the computer was by far the most important means of access.

order for the audience to get to their destination Web site, they may very likely need a search engine, especially if this site is small and has little brand recognition of its own. Figure 7 presents the audience supply chain in diagrammatic form[36].

Figure 7: Supply chain for search engine audiences
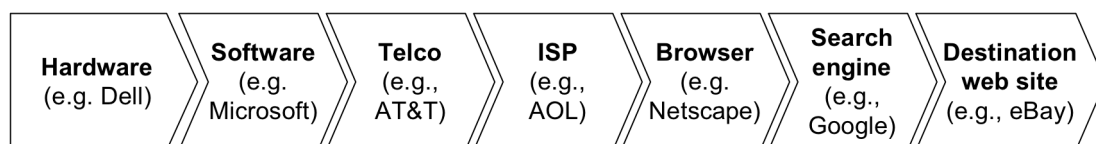
Source: author

This period of search engine history is characterized by attempts at integration – both forwards and backwards – along this audience supply chain. First, we consider attempts by search engines to integrate destination Web sites into their products.

4.4.1 THE DEVELOPMENT OF THE PORTAL

Beginning in 1997 but accelerating in 1998, the "portal" evolved out of the navigational services (both directories and engines) developed in the technology entrepreneur phase. Portals typically had a search engine or directory service at their core, but also had many "channels" which featured content brought in directly from advertisers, including finance, shopping, travel, e-mail, music, etc.

Figures 8 and 9 show the Excite home page from October 1996 and 1997, which illustrates this development clearly. In 1996 the page advertises that the search is "twice the power of the competition" and has content generated by the Excite/WebCrawler team, such as reviews and tours of Web content, below the search. A few services such as travel guides, news, weather, e-mail directory, maps, etc. are also on view, as well as two shopping links – for cars and flights.

---

[36] Although this supply chain is presented horizontally, it is more correctly called a vertical supply chain. Integration along this chain would be vertical integration; backwards integration along the chain is also called downstream integration and integration forwards along the chain can be deemed upstream integration. See section 4.2 for a longer description of vertical integration in the search engine industry.

Figure 8: Excite home page, October 1996

Page retrieved 16 August 2006 from
http://web.archive.org/web/19961022175004/http://www07.excite.com/

In October 1997 the page has been completely redesigned to feature channels, many of which are filled with content from partners.

Figure 9: Excite home page, October 1997
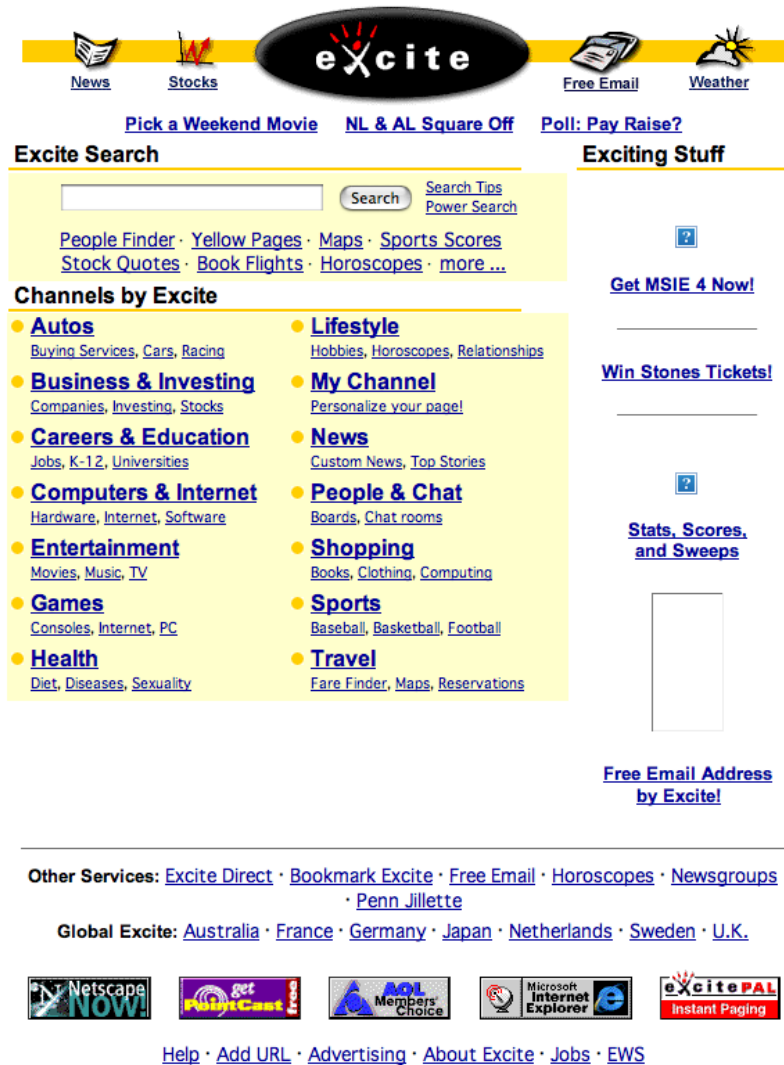
Page retrieved 16 August 2006 from
http://web.archive.org/web/19971012110114/http://www07.excite.com/

These content partnerships are very interesting because they begin to give glimpses of the value that internet traffic is beginning to take online. In an offline network such as a television network, the network pays the production company for rights to distribute the show. However, the online content partnerships were often the other way around – the content producer – for example Preview Travel – would pay Excite to be the main provider of content on its travel page, or "channel", as they began to be called.

This change requires some explanation. In television production, the network pays the production company because they need content attract an audience to sell on to an advertiser. In other words, the network acts as a packager of television content. But although a search engine (or portal, in this era) intuitively seems like the same kind of business, there are key differences. The search engine delivers not just in "impression" or view to the advertiser – although search engine advertisements were sold on a cost-per-thousand-impressions basis, as we have seen – but also, and much more importantly, an interaction – that is to say, an interested person who has actually taken the time to act on the content provided. A growing exploitation of the technical infrastructure of the Web made this change possible. In traditional media it is rarely possible to give advertisers the opportunity to sell directly to customers (apart from newspaper coupons and the like). But it was possible to integrate Preview Travel travel bookings directly into the Excite travel channel, and in effect for Excite to become another avenue of distribution for Preview Travel – and in a sense the Preview Travel website became part of Excite, and vice versa (see Figure 10, below).

Figure 10: Excite Travel Channel, October 1999
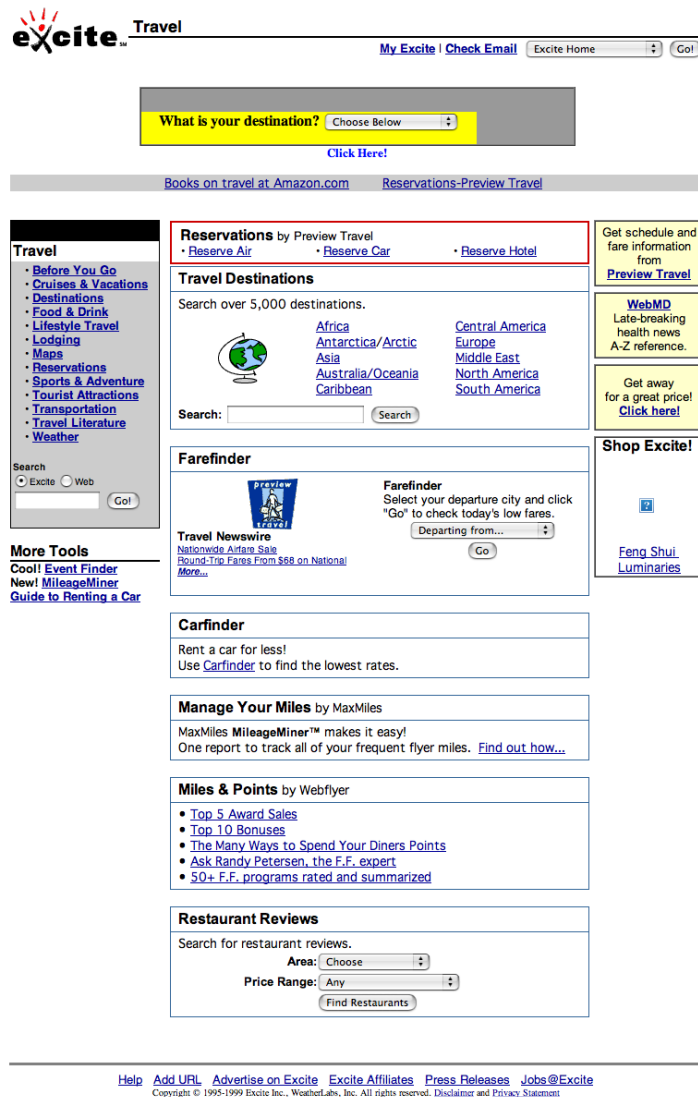
Page retrieved 16 August 2006 from
http://web.archive.org/web/19991008211456/http://www.excite.com/travel. Note: question marks in the figure represent non-archived images which can no longer be displayed.

Thus partnership deals with portals, while they might involve some measure of compensation for content producers, were more typically structured as a mix of direct payments by the content producer (who might now be better understood as an advertiser) and a share of revenues from customers who purchased from a portal website. Here the producer of content becomes the customer, and the traditional value chain gets flipped on its head.

This new revenue based on selling targeted channel impressions to content providers/advertisers and allowing sponsors to sell directly within the portal pages was so successful that channels proliferated and portals became the new face of the search engine. The more channels available, the more high-value sponsorship opportunities could be created, and channels were even specifically created to showcase and sell partner/advertiser products and services. Deals were often long-term (several years) and multi-million dollars – one article in the *Industry Standard* magazine cites a 4-year, $89 million deal and suggests that $2 to $10 million deals were common (Werner & Helft, 2000).[37]

It is important to understand that portals were not examples of vertical integration, in the traditional sense. In general, portals were not buying e-commerce companies, and e-commerce companies were not buying portals. There is no suggestion, for example, that a travel operator like Preview Travel was trying to buy a portal like Excite. But this integration of advertiser and search engine content has important implications, as we shall see later.

4.4.2 VERTICAL INTEGRATION

Nevertheless, during this period many search engines were bought and sold. Dan Schiller argues that with the wide array of cross-media ownership, the increasing transnationalization of media, and the growth of commercial sponsorship as the decisive form of media patronage, the "suitable unit [for analysis] has become the diversified media conglomerate." (D. Schiller, 1999, p. 36). In the second period of search engine history, portals became a natural target for media and telecommunications conglomerates jockeying for position as the internet developed commercially.

Some business strategists hoped that portals could provide a new "window" or viewing opportunity for existing media content, as well as positioning media conglomerates for control of the online operating environment, by controlling the huge audiences that visited the portals. Essentially, the strategy was one of growth through vertical integration in the content supply chain – that is to say, the conglomerates hoped to dominate existing

---

[37] This was also true in Europe. In late 1998, I worked for Jupiter Communications, an market research company specialising in the internet, and documented a $10 million pan-European deal between Lycos and BOL, a book retailer (Van Couvering, 1998).

portals by running their acquisitions more efficiently, exploiting economies of both scale and scope.[38]

Business texts of the time sought to promote this new kind of vertical integration, touting a concept called the "fully-integrated portal" (e.g., Meisel & Sullivan, 2000, p. 484). The vision of the fully-integrated portal was to control the whole user experience online – it was envisaged that users would leave the portal only rarely to visit external sites (see Figure 11). This mega-portal would have three sources of revenue: subscription fees from ISP subscribers, advertising fees, and e-commerce transactions. Economists and business pundits encouraged portals to actively to seek old media partners, develop specialised content, strengthen ties to delivery systems and expand through Europe, Asia, and Latin America.

---

[38] Economies of scale refer to the benefits that accrue for certain types of products when large numbers of them are produced. In media products, the cost of producing the first copy – for example, paying an author to write a manuscript, editing the manuscript, typesetting the book, proofreading the first copy, etc. – often far outweigh the costs of subsequent copies. This is even more true for digital content such as software, where copying and distribution costs are nearly zero. The technical definition is that economies of scale occur when marginal costs (the cost of producing a single copy of the work) are less than average costs – that is to say the average cost declines the more units are produced. Economies of scope refer to the benefits that accrue to companies who can re-use resources to produce a range of products. In media, you might see economies of scope when Harry Potter (the book) is used to provide the basis for Harry Potter (the movie) or Harry Potter (the DVD). Thus economies of scope technically occur when two (or more) products can be produced and sold more cheaply jointly rather than separately. Media industries tend to seek to exploit both economies of scale and economies of scope, and this in turn leads to conglomerates such as Time Warner, Disney, Viacom, News International and Vivendi (Doyle, 2002, pp. 13-15) which have holdings in radio, television, newspapers, cable television, and so on. As digitisation alters the format of media content, these media companies are increasingly also competing with the liberalised telecommunicationss industry.

Figure 11: A fully-integrated portal
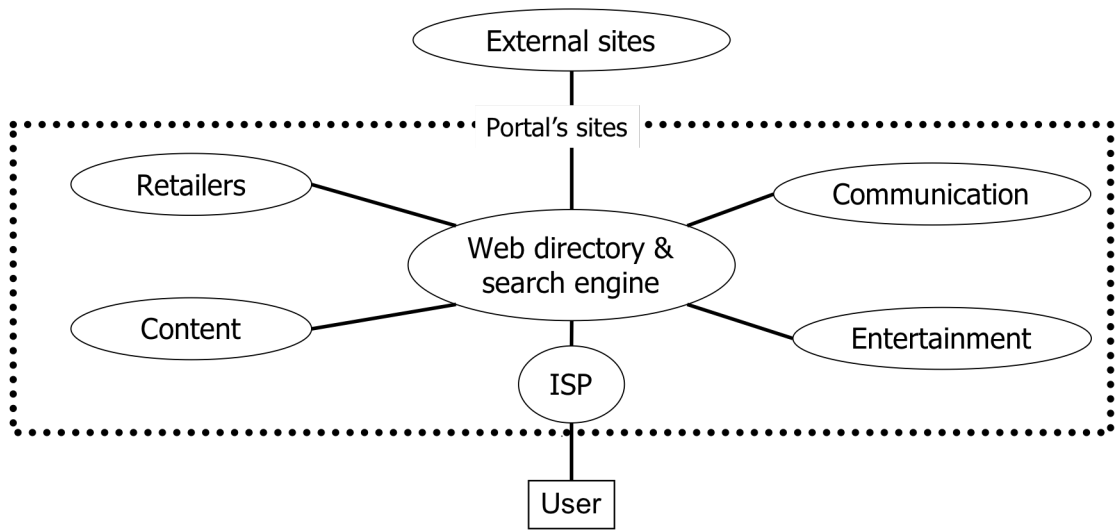
Source: Adapted from Meisel & Sullivan, 2000, p. 480 by author.

Indeed, in 1998 and 1999 the search engine industry witnessed a number of attempts at the creation of these portals by diversified media conglomerates. In mid-1998, Disney acquired 43% of search engine Infoseek for $70 million in cash and $240 million in Starwave stock[39] (CNNMoney, 1998a), acquiring the remainder of the engine in 1999(CNNMoney, 1999). Infoseek was then a popular search engine in its own right, ranked ninth most visited website overall (Harmon, 1998). One week previously, NBC (owned by General Electric) had purchased 19% of C|Net's portal Snap! (CNNMoney, 1998b) Both of these portals had respectable audience, although they were not the market leaders. Nonetheless, both of these high-profile acquisitions both failed and closed in 2001. AltaVista, once the most highly-regarded search engine on the Web, was sold by computer manufacturer Compaq (who had acquired its parent Digital Equipment) to media investment group CMGI (which also owned Lycos) for £2.3billion in June 1999 (Dignan, 1999). In 2003 it was sold to Overture for $140 million , and later vanished into Yahoo! (see the next section, "Syndication and Consolidation").

Nor were media conglomerates the only actors seeking to dominate the online markets. Infrastructure providers, most notably telecommunications providers, also attempted toforward-integrate along the audience value chain and enter the portal space. This was

---

[39] Starwave at the time operated several websites for Disney brands including abcnews.com and espn.com, as well as sites for the NFL, NBA and NASCAR.

part of an overall strategy to engage with media content as digital content made convergence between telecommunications operators and media companies more of a reality. Highly-rated portal Excite was acquired in January of 1999 for $6.7 billion in by broadband internet service provision (ISP) company @Home (a joint venture of AT&T and several cable companies) (Junnarkar, 1999). Similarly, Lycos was purchased for $12.5 billion in May 2000 by Terra Networks (owned by Spanish telecommunications operator Telefónica) (Kopytoff, 2000). These acquisitions were motivated in part by a desire to emulate the enormous success of AOL, whose huge traffic, generated by a loyal base of ISP subscribers, enabled it to make some of the largest portal advertising deals. AOL, the largest ISP in the world at that time, also attempted to forward-integrate by purchasing browser manufacturer Netscape, and its NetCenter portal, in November of 1998 for $4.2b (Clark, 1998).

Yet, none of these acquisitions fared well. Excite@Home went spectacularly bankrupt in 2001 (Wallack, 2001), and Lycos, while still technically in existence in 2006, stopped providing its own search in 1999 and was sold to South Korean online media company Daum Communications in 2002 for $95 million, a fraction of the price Telefónica paid (Reuters, 2004). AOL still operates Netscape's Netcenter, but Netcenter no longer registers as a destination among searchers.

Certainly the nail in the coffin of many of these services was the dot-com crash. To a large extent the growth in sponsorship revenue for all the portals was funded by money from the dot-com boom that was going into start-up internet ventures, which depended on becoming leaders in their respective markets, based on audience numbers that only the search engines could bring them. When the stock market began its crash in spring of 2001, much of this money dried up. But there seem also to have been other factors.

In his research, Blevins analysed the Disney/Infoseek deal in some detail, and accounts for the closure as a failure of "synergy" – put simply, as too much branding by Disney (Blevins, 2004). In my analysis, this relates to a misunderstanding by the media companies about the role of the search engine, alluded to earlier in Section 4.4 when distinguishing between the content supply chain and the audience supply chain. Examining the audience supply chain for the search business as shown previously in

Figure 7, we can see that Web sites are upstream from portals, who act as distributors of audiences for other websites like e-commerce providers. Online, however, there is not much of a distinction between the website of an e-commerce provider or "advertiser," like Ford, and the website of a "content provider" such as ABC. Thus, by adding more Disney content to the Go Network site (as Infoseek eventually became), Disney actually moved the portal away from its position as a distributor and instead it became a destination website. As Blevins describes, its audience immediately began to drop, its traffic dropped, and it lost its paying customers, other advertisers. The problem of "synergy," then, as it relates to big media is as follows: search engines do not represent an opportunity to benefit from economies of scope for media companies. Disney content, as it turns out, cannot be repackaged as a navigational portal. Disney is a destination site, upstream from search. A Disney portal is merely a Disney home page, with little value to audiences not interested in Disney content. In tandem, the Infoseek search engine was put on the back burner. In 2004, in a conference panel discussion on the history of search, Infoseek's founder, Steve Kirsch, said that around 1998 he was the only one pushing developments in search; the business people wanted to focus on the top pages, and management wanted to move towards a portal (Schwartz, 2004).

However this issue of "over-branding", if it may be termed that, seems less pertinent for infrastructure providers who should have little interest in the content of search engine results. Once again it is helpful to examine a particular case. The most high-profile case of failure was the acquisition of Excite by broadband cable provider @Home. This merger of a top-tier portal with an access provider backed by the US giant AT&T seemed certain to succeed and become the "AOL of broadband," but instead failed and went bankrupt within two years. Unfortunately we have no detailed academic study of this case in the way that Blevins studied the Infoseek/Disney case. However, according to press reports at the time of the bankruptcy in 2001, the focus of Excite@Home was on developing a high-speed cable network, at the insistence of its primary shareholders, who were cable company executives. In the meantime, it began to be difficult to justify spending on developing the portal, and particularly on developing the search engine, which was seen as a necessary but unproductive part of the business – in other words, a loss-leader. Later, *Wired* magazine suggested that @Home had simply been a vehicle for

off-book financing of broadband infrastructure, which AT&T bought for $307 million during the disposal of assets (Rose, 2002). If that was in fact the case, the development of the Excite portal would have been irrelevant. In any case, at the time of the sale, the search technology that had built the second-largest search and directory site on the Web was deemed worthless and scrapped, and the domain name was sold for $10 million at the time.

A similar fate seems to have befallen Alta Vista, this time with computer hardware rather than cable at the core of the integration strategy. At AltaVista, too, the emphasis switched from search to portals, and it became impossible to fund the development of the search engine, leading to the departure of the chief engineer and co-founder, Louis Monier, with his team (Battelle, 2005, p. 52).

Thus an important element that characterises this phase of search engine development, in addition to the acquisition of may of the search engines by larger conglomerates, is the downgrading of search within the portal; the search engine itself was no longer seen as a key competitive advantage for a portal, but rather as a simple requirement for doing business. Recall that the vision of the fully-integrated portal was that this mega-website would be so engrossing (or "sticky," as the industry called it) that users would never want to leave. They would arrive through the website of the service provider, browse licensed content, use branded online email, and shop for purchases all within the confines of the portal. But search, of course, is the opposite of "sticky" – the whole point of a search engine is that users search for something and then leave the website. Search seemed like a giant firehose spraying precious audience everywhere on the Web but into the portal.

In Section 4.1 I described the inclusion of partner functionality, such as flight searching from a travel provider, into portal pages. Gradually it became clear that search functionality could be conceived of in the same way. Thus, as part of the movement towards portals, which as described earlier was linked to the integration of content from advertising and technology partners, the search engine market split into those who were intent on developing media properties – for example Go – and those who focused on a more technology-led strategy, through what was called "white-labelling" or licensing of their search technology to third parties. Inktomi was perhaps the best example of this

strategy. In June 2000, for example, Inktomi delivered search results to eight separate portals, including AOL, HotBot, MSN and Snap as well as smaller websites like iWon, LookSmart, GoTo and 4Anything (Sullivan, 2000).

Despite the diminution of the actual search engine from the core of the business to loss-leading commodity, there continued to be new technical innovations in search, and new search companies continued to be funded by venture capital. In 1998, Ask Jeeves debuted with a new interface to the old Magellan idea of editorially-rated sites, by letting users input natural-language questions and organising the results around the most frequently-asked questions. Search aggregation engines such as Dogpile and MetaSearch queried all the other search engines and returned a mix of results. iWon paid its audience directly in the form of a lottery in which each search submitted counted as an entry. Direct Hit began ranking by popularity rather than simply by website content. And also in 1998, Google began a new search engine with a radically new ranking algorithm, backing from significant Silicon Valley venture capitalists, and a key distribution deal with Netscape (for an in-depth history of Google, see Vise & Malseed, 2005).

Important as Google's technical innovations were, equally or perhaps more important for the future of the search engine industry as whole, was the debut of GoTo. GoTo was a search engine with no pretence of searching the whole Web. The GoTo index was instead made up of people who paid to be there, and it allowed these advertisers to buy the search terms they wanted. Thus when searching for "flight to New York" the travel agency or airline which had agreed to pay the highest advertising fee would be listed first. But GoTo knew that advertisers would not pay to be included in an unproven search engine, so Bill Gross, its founder, introduced the policy of charging advertisers not "per impression" as was now common practice, but rather per click. That is to say, the advertiser was only liable for the fee when someone actually clicked the ad – unclicked impressions were given away for free. The importance of this development cannot be overstated. Instead of the multi-million dollar impression and sponsorship deals based on the huge reach of the major portals, GoTo offered small, controllable deals where a few cents would get an advertiser a definite visitor for their site. It was a compelling business model, particularly because at first GoTo deliberately undercut the market (Battelle, 2005, p. 111ff).

But it was more important than simply a brilliant business idea: it was part of a crucial shift in the search engine business. No longer would the *audience* (the traditional media commodity sold to advertisers) be at the core of the search business. Now, the online commodity of choice would be *traffic* or the flow of visitors from one website to another. When audience was the main commodity sold, the key task of online websites was to gather and keep as many audience members as possible, with the ultimate aim being – however unrealisable – to own the whole internet. But as traffic emerged as a key commodity in its own right, sites which had as much traffic as possible – that is to say, as many people coming and going as possible – became the nexus of economic wealth. Search engines were the obvious choices, and the new economic possibilities led to a resurgence of technical competence and the technically complex search product as essential elements of the large online media players we see today.

4.5 Syndication and consolidation (2002-?)

The final period of the short history of search is one of consolidation and concentration. This is due to two interconnected dynamics. First, media and infrastructure corporations have ceded search to technology companies and are content to buy their search from search providers. Second, the revenues generated from pay-per-click search advertising have meant that the large players have been able to buy their rivals, as shown in Figure 6 at the beginning of the chapter – in this period, acquisition activity of search technology is by other search providers – in fact, almost exclusively by Yahoo.

In 2001, during the dot-com crash that marks the end of the second period of search, Disney's CEO, Michael Eisner, accounted the failure of big media online by suggesting "the advertising community has abandoned the internet" (cited in Blevins, 2004, p. 265). Four years after Eisner's quote, the Internet Advertising Bureau had recorded the ninth straight quarter of advertising growth online, bringing 2004 online advertising market in the US to over $9.6 billion and the first half of 2005 to nearly $5.8 billion (the total figure for 2006 was estimated to be over $12 billion). The slump of 2001 has been revealed to be just that: a slump, as Figure 12, below, clearly shows. In fact, the growth in internet advertising has outpaced the growth in television advertising in its first 10 years, according to the Internet Advertising Bureau, which assembles market statistics for the industry.

Figure 12: US quarterly online ad revenue, millions of dollars, 1999-2005.
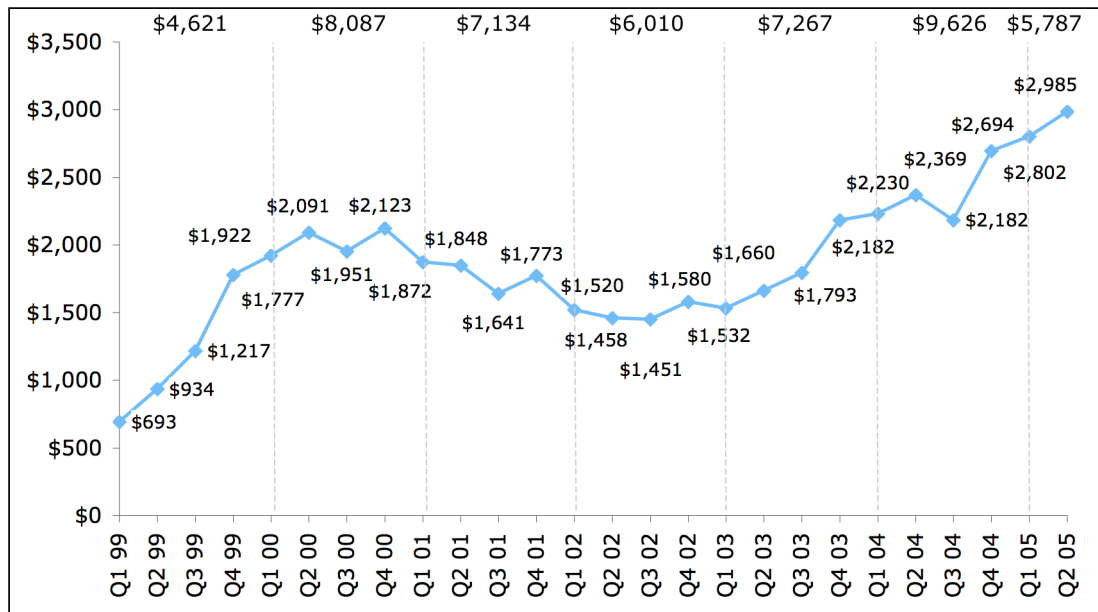
Source: Adapted from Internet Advertising Bureau (PriceWaterhouseCoopers, 2006).

This growing ad market has been increasingly funded by growth in "paid search" advertisements, that is to say the type of cost-per-click advertisements pioneered by GoTo, linked to user traffic, whether on search engine sites or syndicated to other websites. This advertising has three key characteristics: 1) it is priced on a *cost-per click* basis; 2) it is *contextual*, linked either to page content or to the users' search term; 3) it is *syndicated* to other websites on a revenue-sharing basis (ie the fee is split between the owner of the website and the provider of the paid search service).

The market for these ads has been overwhelmingly dominated by Google and Yahoo. In November 2001, Yahoo made a deal with Overture (formerly GoTo) to launch CPC ads alongside their search results, which at that time were being provided by Google on a syndication basis (Yahoo!, 2001). A year later, in December 2002, it began a transformation. Yahoo, originally a directory and always a buyer of syndicated search results, announced it would purchase Inktomi, a pure search engine company specialising in syndicated search results (Yahoo!, 2002). It began serving its own search results in April 2003 (Yahoo!, 2003a). Three months later in July 2003, the company announced it would

acquire Overture for $1.5 billion (Yahoo!, 2003b). At the time, Overture's clients included MSN, ESPN, and CNN, as well as a staggering 88,000 other advertisers.

Meanwhile, Google had introduced its large-scale automated advertising programme, called AdWords, in October 2000 (Google, 2000) – but on a CPM basis. In February 2002 it debuted its own CPC pricing programme (Google, 2002). By March 2003, it announced that it had the largest advertising programme in the world, with over 100,000 advertisers (Google, 2003a). In June 2003 it began to syndicate these CPC ads to partner websites on an automated basis, through a program called AdSense (Google, 2003b). By the end of 2005, the company reported that 44% of its advertising revenue ($2.688 billion of $6.065 billion) had been made on syndicated advertising (Google, 2006a). According to Google's website, by 2006 it had "the largest online advertising network available, reaching over 80% of 30-day US internet users, " (Google, 2006b).

Microsoft and Ask, the two other major providers of search technology, were behind Google and Yahoo in exploiting syndicated advertising. Until 2005/6, both engines simply used the syndicated services of Yahoo (in the case of Ask) or Google (in the case of MSN) (IAC Search & Media, 2005; Newcomb, 2006).

Google and Yahoo also aggressively pursued a syndication strategy with access providers – in Google's case primarily syndication of search results and advertising, but in Yahoo!'s case the provision of co-branded portals including e-mail, chat, news, horoscopes, etc., as well as the technical facilities for integrating partner content and other content through the RSS (Really Simple Syndication) technical standard. One such example is the BT Yahoo! Broadband portal in the UK (see Figure 12, below), available to all BT broadband subscribers in the UK.

Figure 13: BT Yahoo! personalised subscriber portal
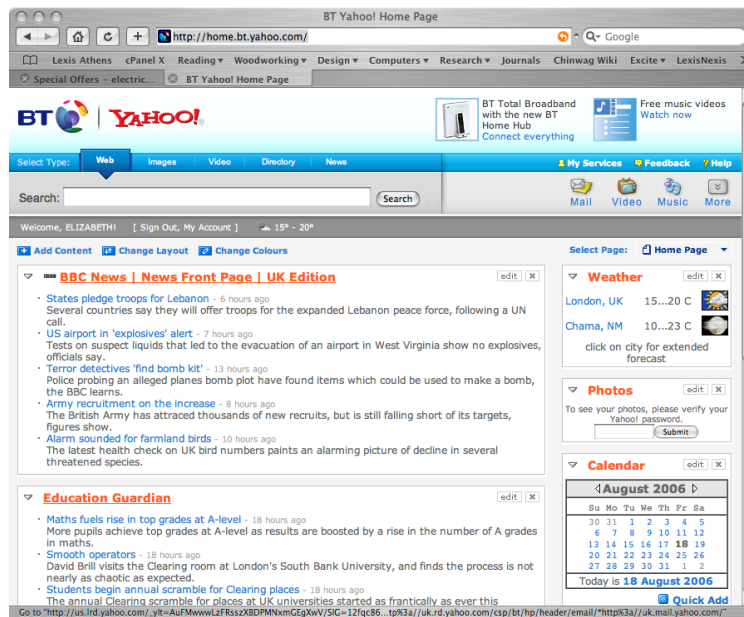
Page retrieved 18 August 2006 from http://home.bt.yahoo.com

While such deals are too numerous to be mapped in their entirety, a review of the US market shows search engine deals on the homepages that ISPs provide to their customers (see Table 4).

Table 4: US ISP search engine affiliations by rank and provider

Rank	ISP	Subscribers (millions)	Subscriber homepage	Search results provider
	All others	22.3		
1	AOL	18.6	aol.com	Google
2	Comcast	9	comcast.net	Google
3	SBC (AT&T)	7.4	sbc.yahoo.com	Yahoo
4	Verizon	5.7	varies	Yahoo OR MSN Premium
5	Road Runner (TWC)	5.4	www.rr.com/publicpass/	Google
6	Earthlink	5.3	my.earthlink.net	Google
7	Cox	3.1	www.cox.net	Google
8	BellSouth	3.1	home.bellsouth.net	Google
9	United online	2.8	my.juno.com	Yahoo
10	Charter	2.3	www.charter.net	Google
11	Cablevision	1.8	www.optonline.net	Infospace
12	Qwest	1.7	qwest.msn.com	MSN

13	Sprint	0.78	my.sprint.earthlink.net	Google
14	Insight BB	0.51	www.insightbb.com	Infospace
15	Mediacom	0.5	e.g., suncity.mediacomtoday.com	Infospace
16	Covad	0.48	b2b lines only	not determined
17	ALLTEL	0.44	www.alltel.net	Infospace
18	Citizens	0.33	frontier.myway.com	Ask
19	CenturyTel	0.29	www.centurytel.net	Google
20	LocalNet	0.26	start.localnet.com	Google
21	Hughes DIRECWAY	0.26	hughesnet.myway.com	Ask
22	Cincinnati Bell	0.17	broadband.zoomtown.com	Google

Source: Data on ISP rank and subscriber numbers from Goldman (2006) and reflect Q1 2006 status; other data compiled by author.

It is clear from this table that Google in particular has been very effective in distributing its search engine backwards to ISPs[40]. Figure 14 shows that if these figures are aggregated, Google is distributed on the home pages of ISPs that account for 55.6% of the internet subscribers in the United States.

---

[40] Infospace, which figures several times in this table, is a provider of paid search results only – in effect, a modern GoTo.

Figure 14: Search affiliations of US ISPs
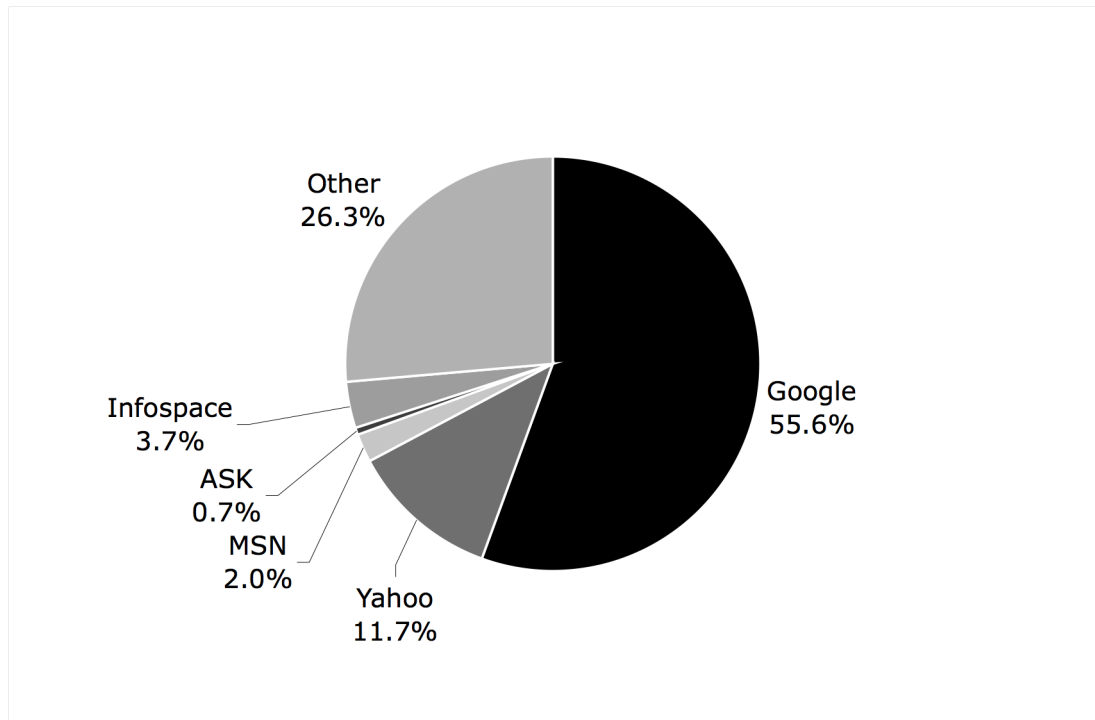
Source: Data on ISP rank and subscriber numbers from Goldman (2006) and reflect Q1 2006 status; other data compiled by author.

What these very successful syndication efforts have meant is that, effectively, Google and Yahoo have achieved a situation where, without needing to purchase companies, their advertising is carried across the Web through syndicated advertising and audience is directed to them though syndicated search engine functionality.

In his book reviewing the state of research on political economy of communication, Mosco argues for an analysis of market concentration in media markets which focuses on something more than ownership. He suggests that "networks of corporate power" might need to be investigated through "forms of corporate interaction that build powerful relationships without actually merging businesses. These forms encompass a range of 'teaming arrangements,' including *corporate partnerships* and *strategic alliances…*" (Mosco, 1996, p. 189, italics original).

This analysis of the search market seams to suggest that earlier efforts at vertical integration have been replaced by what we might term a "virtual" integration along the

audience supply chain. In contrast to the fully-integrated portal, the new model might be conceived as a *syndicated portal*, as in Figure 14 below.

Figure 15: The syndicated portal
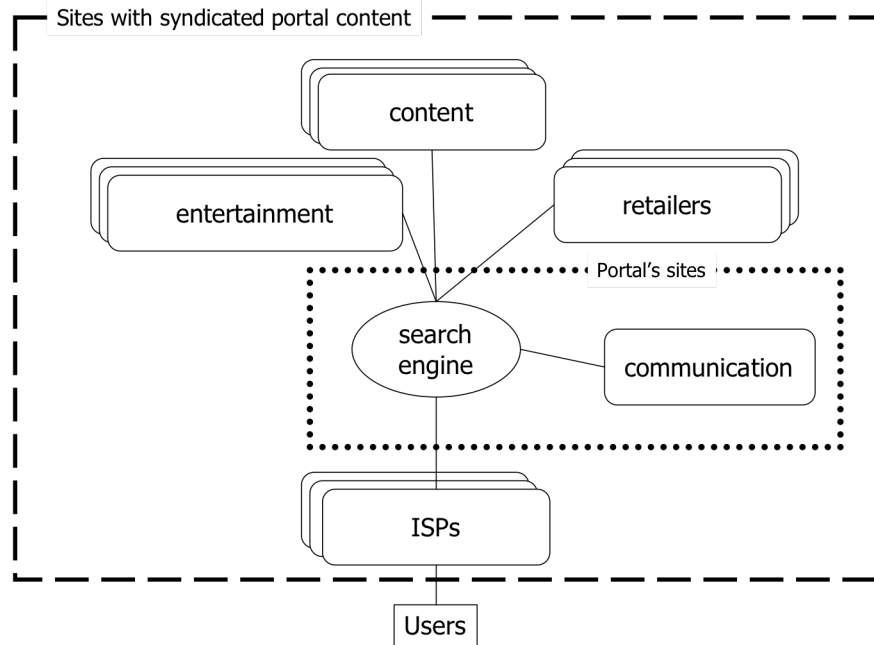
Source: author.

The difference between the syndicated portal discussed above and the fully-integrated portal imagined by dot-com boom enthusiasts consist not merely of the qualitative difference between ownership and partnership, but also in the quantitative differences of having multiple ISPs, multiple content providers, multiple entertainment venues and multiple retailers attached to the portal. The lines between the search engine and its partners are lines of both traffic and money.

By using syndication both into advertisers and also into partners who are further up the supply chain such as ISPs, the new giants of search have developed a network that extends across the internet. No longer is it necessary to "own" the internet, as those who dreamed of controlling a fully integrated portal did. Rather, by means of "virtual" integration using technology to achieve syndication, Google and Yahoo, and to a lesser

extent Ask and MSN are able to stretch their ability to monetise (or commoditise) traffic across the Web, without the need for ownership[41].

Beginning from about the middle of the first decade of the 21st century[42], social media sites such as Facebook, MySpace, Twitter, Blogger, LiveJournal, Flickr and YouTube began to draw considerable traffic and audience engagement (measured as time spent) to their websites. By the end of 2009, Facebook traffic was reported to have exceeded that to Google's search engine in the UK (although just for a few days) (Schwartz, 2009). Given the large market share of search engines and their central place in the web infrastructure as discussed above, this is a fairly astonishing statistic, and the relation of these websites to search engines deserves some consideration. The rise of social media networks has implications for search engines but it is not clear whether search engines will end up competing with, co-ooperating with, or co-opting social networks.

Social media sites are diverse, but have in common an infrastructure whereby users create their own content within a defined technical framework, and also use the supplied framework to link directionally to other users, for example by "following" them (Twitter) or "friending" them (Facebook)[43]. The content allowed can be widely varied, ranging from personal web pages (MySpace, Facebook), to dated updates (Blogger, LiveJournal), to small snippets including links (Twitter), to pictures and video (Flickr and YouTube). In many cases there is some overlap (updates via Facebook, for example), and third-party services allowed people to aggregate connections via various platforms; indeed, some websites published programming interfaces (API's) to facilitate this connection. In some cases, for example, Facebook, technically savvy users could also create their own small applications using these APIs and distribute them to their contacts through the interface.

---

[41] It is also worth noting that although emphasis in the industry has shifted to paid search, Yahoo and MSN also retain more traditional "portals" with channels filled by advertiser content.

[42] Social network sites had been launched earlier, as boyd and Ellison's (2008) chronology indicates, but only began to be "mainstream" around 2003-4, with much activity happening in 2006, including the launch of Twitter.

[43] This definition differs only slightly from that offered by boyd and Ellison: "We define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site." (boyd & Ellison, 2008, p. 211) The definition in the text above is slightly broader in that it includes sites where the "profile" mostly consists of visual information, such as Flickr and YouTube.

These sites are not strictly search engines, but many of them incorporated search facilities as an essential element to facilitate initial connections, and to enable users to follow topics as well as connect with other people. They also had other connections to the search industry. Some were owned by search engine companies: YouTube and Blogger, for example, are owned by Google, and Flickr is owned by Yahoo; but others were independent (Facebook[44], Twitter), while yet others have affiliations with traditional media (MySpace is owned by News, Inc.). While they may not be search engines, there are some areas where these sites seem to share features of navigational media, as I discuss below.

First, these sites share with the search engines the characteristic reliance on outsourced and distributed content providers. Each of the sites listed above provided a technical interface but did not, per se, provide media content – texts, pictures, videos and even small applications – although these are what its users relied on it for. Instead, rather than rely on indexing technologies, as search engines do, users are positioned as active content creators and it is their content that formed the base upon which the social media site operated. Some of this content, unlike the content upon which the ordinary search services are based, is *exclusive* to the service in question, and this provides the service owners with an enormous asset. For example, a post on LiveJournal may only be available to other members indicated as "friends," depending upon the user settings; and a full profile will only be available to "connections" on LinkedIn. Social media sites are quite heterogeneous, so there is at least one important caveat to this: blog content (such as that hosted by Blogger or WordPress), unlike the content produced on other social networks, tends not to be as restricted to other members of the service, but is more freely available.

Second, the reliance on a network of distributed content creators gives social network providers a huge amount of frequently changing content, which is of great value to users, and thus a huge amount of traffic, which is of great value to advertisers. It cannot be stressed enough that most social networks have an *internal* traffic source (their members' content and links) and source of new members (members soliciting friends for new members) and are not as dependent on search engines for traffic as other websites. Like search engines, these services are free to users, and like search engines they typically make

---

[44] Facebook is not entirely independent, as Microsoft purchased a small stake in 2007.

most of their income from advertising, although some, like LiveJournal, also charge for "premium" accounts with additional features. Advertising could be charged on a cost-per-click basis; but unlike search engines they could also target ads to a range of demographic or personal details, which had the potential to make each click more valuable. They also had a possible additional funding model, which was the sale of access to their user-created content, including profiles and demographics, to other companies for further commercialisation. Having said this, at the end of 2009 there was no uniform business model.

Three examples help to clarify this: the cases of Facebook and Twitter, two independent providers of highly-successful social-networking websites, and Technorati, a search engine based on blogs. Facebook's funding model was to provide demographic- and interest-based targeted advertising on either a CPC or CPM basis. Although you could search for people or groups on Facebook, there was no generic search – ads were purchased by targeting, just as is the case with traditional media, but unlike the case of search ads, which are based on user behaviour. Facebook's revenue stream was additionally heavily supported by Microsoft, which took a stake in the company in 2007. Microsoft's Bing search was integrated into the Facebook search pages, and only Bing was able to access information stored in Facebook's public profile pages. Twitter, by contrast, appeared to have no revenue stream at all other than investment. Speculation abounded as to how it would turn its popular service into a sustainable business, with advertising being widely tipped (Tartakoff, 2010). In the latter part of 2009, Twitter took a step towards sustainable revenue by licensing the content its users create to Google, Microsoft, and Yahoo for indexing, for rumoured tens of millions of dollars per year. Technorati belonged to the class of businesses that are not social media but are made possible by social media. Technorati indexed only blog sites, and implemented special search features appropriate to the format, such as ranking by date order (see Thelwall & Hasler, 2006). Technorati's funding model was also based on advertising and it announced in 2008 that it would begin selling syndicated advertising to its network of blogs (Arrington, 2008).

These three examples represent quite different ways of creating and monetising social media. With Facebook, the user's profile and interests are spelled out and this forms a large part of the content which users and advertisers value. Twitter takes the form of a

newsfeed nearly devoid of other personal information apart from linkages between followers and followees. And yet, each of these two websites, in their different ways, has taken advantage of its proprietary network to include search on their own terms by charging for access to the content their users create, including content about themselves, which in turn generates traffic. Meanwhile, Technorati has followed in the footsteps of the major search engines, capitalising on the much more loosely defined and freeform social network of bloggers. In each case, the currency of traffic remains central, and central to understanding the business strategies of social media providers.

Finally, along with the value of traffic, as we have seen in other navigational media, comes the incentive for some users to try to manipulate the system for profit; thus there exist Facebook, Twitter and blog optimisation services, and Facebook and Twitter spam, as well as straighforward Facebook and blog advertisers, and hundreds if not thousands of profiles, Twitter feeds, and blogs and blog commenters which might be called spam, all dedicated to driving traffic to private interests.

The relation between the large search engines and these social media sites is complex. Since they have their own internal source of traffic, social networking sites can form a large proprietary traffic network with personal data that is not easily available to the search engine. They can also have search-like functionality in terms of driving traffic – many Twitter updates contain URLs, for example, so Twitter functions as a source of traffic to a range of sites, as do blogs, which often contain links and references to other sites. Social networks are clearly valuable properties: the connection of large traffic volumes with personal data is irresistible to advertisers. But they cannot remain wholly separate from the Internet and increasing the size of the network and the volume of content and traffic must be of paramount importance to network owners. Here the search engines take on their role of connecting traffic through a range of disparate technical infrastructures. Zimmer (2008) calls this mixture of personal data and search technology "Search 2.0" and raises concerns about the clear privacy implications, implications that may well prompt governments to act to restrict it. Thus while search engines already owned many important social networking properties, an uneasy and slightly competitive relationship between the independent networks and the dominant search engines was in place at the end of 2009.

4.6 Conclusion

Using a conceptual framework derived from studies in the political economy of media and communications tradition, this chapter has analysed the historical development of the search engine industry. Search engines, it has argued, are the purveyors of a new media form – we can call it navigational media – that have taken advantage of a fragmented media market to establish their power as distributors of traffic via the creation of flexible and stable networks. By 2006, we had a situation where the large search engines overwhelmingly dominated the search market, as Figure 16 shows.

Figure 16: Share of U.S. searches, November 2005.
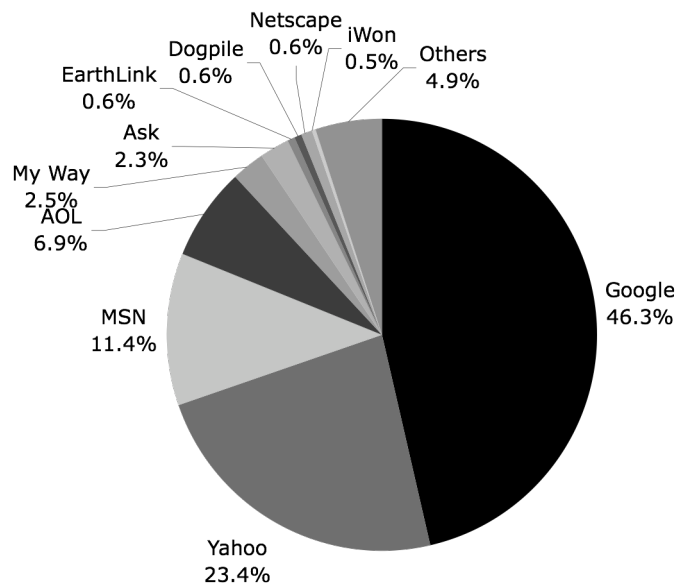
Source: Nielsen/NetRatings for SearchEngineWatch

Other smaller search engines existed, such as Nutch (www.nutch.com) and Gigablast (www.gigablast.com); and there were also a range of small vertical search engines. However, Figure 16 shows that Google, Yahoo, and MSN accounted for 81.2% of all searches in the US market as measured by Nielsen Net/Ratings. Further, all the named others on the chart had search results provided by one of these companies or by Ask (formerly Ask Jeeves). These smaller search engines, therefore, were disregarded for the purposes of the present analysis.

As a result of the growth of paid search versus all other types of online advertising, we can also see a surprising result: the same four companies also accounted for nearly 70% of the *total* online advertising market in the US in 2005, as Figure 17 shows. In the first half of 2005, the total online advertising market, according to the Internet Advertising Bureau (IAB), was $5.8 billion . For the first half of 2005, Google reported a US advertising income of $1.591 billion, Yahoo of $1.475 billion, MSN of $517 million, and AOL of $445 million, leaving $1.772 billion to be divided amongst all other online advertisers.

Figure 17: Share of the total US online advertising market for the first half 2005.
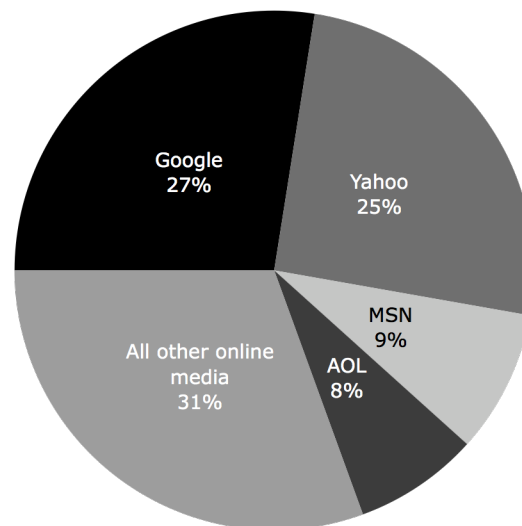
Source: Internet Advertising Bureau, company quarterly SEC filings, author's analysis.

This chapter has divided the history of the search engine into three periods in order to examine this growing concentration. In the first period, many new technologies were created, and venture capital systems helped to launch the new companies into the emerging industries created out of the development of the internet. The new companies turned to both advertising and licensing for revenue generation, and succeeded in gathering large audiences at least in part through significant strategic alliances with the internet service provider AOL and the browser manufacturer Netscape, and the market was competitive, with multiple companies providing multiple search engines..

In the second period, search engines developed specialised content "channels" created of advertiser content where lucrative sponsorship deals became possible through the segmentation of their audiences. They were the focus of acquisition activities by both traditional media companies and telecommunications and cable companies who sought to acquire these portals with the hope of owning a large slice of the Web. However, during this period the technology of search was neglected in favour of developing channel content. There were three exceptions: first, those entrepreneurs with new technology for search who continued to be funded by venture capital in hopes of capitalising on the booming market for internet stocks, such as Google and AskJeeves; second, those who developed and licensed search to other websites, such as Inktomi; and third, those who sought to develop alternative models of payment, such as iWon and GoTo. This chapter argues that the cost-per-click model that the latter engine pioneered helped redefine the online media commodity from audience to traffic.

The third period saw the emphasis on traffic and the sale of traffic give a massive boost to search engine revenues, particularly for the early movers Overture (formerly GoTo and acquired by Yahoo during this period) and Google. Instead of seeking to acquire and control content, the engines concentrated their attention on distributing their traffic-based advertising throughout the Web. As a result, they have developed a diversified and flexible revenue base which includes hundreds of thousands of advertisers, tens of thousands of websites on which their ads are distributed, and distribution of their search engines on most major ISPs. Microsoft, the only significant new entrant of the latter period, has so far been unable to match this "virtually-integrated" network. The rise of social media networks has implications for search engines but it is not clear whether search engines will end up competing with, co-operating with, or co-opting social network sites.

The current situation, therefore, is one of oligopoly. This applies not only in the United States, upon which this chapter has focused, but in many, if not all, parts of the world. The geographic dynamics of the search engine industry are the subject of the next chapter. This chapter, however, raises the very serious issue of whether or not we can now rely on competition in the marketplace, as some have urged (e.g., Goldman, 2006), to

assure that the provision of search to the public remains at a high quality and the deficiencies already present in search engines are remedied.

V

Finding the Centre
Search engines in Japan, China, Germany and South Africa

5.1 Introduction

This chapter examines the spatial aspect of search engine bias, addressing the research sub-question in Chapter 2, Section 2.6: *What aspects and elements of the capitalist processes of spatialisation can be linked to search engine bias?* and the related empirical questions set out in Chapter 3, Section 3.3, as follows:

- Is the search engine industry concentrated into certain companies and/or certain geographies?
- What is the relationship between global search provision and local search provision? Can we distinguish a global centre or periphery in current search engine operations?

Mosco suggests that spatialisation is particularly relevant to the political economy of media and communications because the wider process of spatialisation, whose dynamics include "reconfigurations" of space, uses communications tools and infrastructure to achieve these reconfigurations (Mosco, 1996, p. 173). One example of such a reconfiguration is the apparently effortless way in which search engines appear to be able to reach content anywhere in the world. Although spatialisation and, in particular, the

"annihilation of space" that even Marx referred to (cited in Mosco, 1996, p. 172) can seem to have an ineluctable force, still, local physical dynamics may conflict with spatialisation processes. This chapter is an examination of the dynamics of search both as a local and a global service.

As discussed in Chapter 1, Section 1.2, search engine bias is expected to have a geographic dimension. The major search engines such as Google, Yahoo and MSN offer many different language versions. However, given that one of Google's one hundred and sixteen language services in 2006 was Klingon, a fictitious language created for the television series *Star Trek*, and others included "Elmer Fudd" (from the Warner Brothers cartoons), "Bork, bork, bork!" (from the Swedish Chef character on the television series *The Muppet Show*), "Hacker" and "Pig Latin," it seems reasonable to question whether the existence of a language service can be a proxy for commitment to crawling or displaying results from the language community served through this interface. Studies of US-based search engines have shown that national context can make a difference in how accessible a Web page can be. One possibility is that local language features may result in poor search algorithm performance and indeed some studies have reported failures of major search engines when confronted with non-English languages (Bar-Ilan & Gutman, 2005; Choros, 2005). However, Vaughan & Thelwall also reported American sites being favoured independently of language of origin in a cross-national comparison of results for China, Taiwan, Singapore, and the US (Vaughan & Thelwall, 2004).

In Chapter 2, Section 2.4.1, it was suggested that the political economy dynamics of traditional media play a part in the development of global capitalism. First, they may act as an instrument of foreign policy (through institutions such as the Voice of America, for example). Second, media industries were said to help to further capitalism itself, essentially through the establishment of advertising and actions to help weaken public broadcasting in the pursuit of new markets. Third, media companies help to establish American representations and American cultural values, creating, in effect, an American centre and a global periphery when it comes to cultural representation in the media. These arguments have been were made with regard to representational media such as television, radio and the press, and need careful reassessment when it comes to search engines.

While acknowledging that search engines represent a different media form, we can suggest that it is possible that the global dynamics of the search industry may be similar to the global dynamics of other media markets. On the internet, there are millions of websites and, although many of the most highly visited are US-based, there are many examples of highly visited websites around the world. In particular, the issue of representation in the case of search engines is not so much about whether they produce and export American content, or about positive representations of the United States, but whether or not the people around the world are able to obtain material which is relevant to their national context, or make accessible their own material to a wider world in a fair and equitable manner.

Research on search engines, particularly on the search engine industry, becomes more difficult when looking outside the borders of the United States. Industry figures suggest that US search engine providers dominate worldwide traffic (see Table 5)[45]. However most scholarship on search engines has been written from a United States context, with the notable exception of work done on the German market (e.g., Machill, Neuberger, & Schindler, 2003; Machill, Neuberger, Schweiger, & Wirth, 2004), and this work unfortunately does not compare Germany to with other national contexts.

Table 5: Top 15 online properties worldwide by number of unique visitors, March 2006

Property Name	Unique Visitors (000)*	Global Reach
Worldwide total	694,260	n/a
MSN-Microsoft Sites	538,578	77.6%
Google Sites	495,788	71.4%
Yahoo! Sites	480,228	69.2%
eBay	269,690	38.8%
Time Warner Network	241,525	34.8%
Amazon Sites	154,640	22.3%
Wikipedia Sites	131,949	19.0%
Ask Network	127,377	18.3%
Adobe Sites	115,774	16.7%
Lycos, Inc.	109,394	15.8%
CNET Networks	107,589	15.5%
Apple Computer, Inc.	98,622	14,2%
Real.com Network	78,104	11.2%

---

[45] According to ComScore, the measurements in Table 6 derive from a massive panel of two million users recruited online which 'has active representation from countries that comprise 99 percent of the global internet population' (comScore Networks, 2006).

Monster Worldwide	74,152	10.7%
Wanadoo Sites	73,446	10.6%

* Those aged 15+ who have used the internet during the month. Excludes traffic from public computers such as internet cafe and, access from mobile phones or PDAs.

Source: adapted from comScore World Metrix (comScore Networks, 2006) by author.

The figures in Table 5 indicate that some 77.6% of the total global audience used an MSN or Microsoft site during March 2006, 71.4% used a Google website, and 69.2% used a Yahoo! site – that is to say, nearly half a billion people viewed the websites of each of these companies. These figures do not consider the search provision by these firms to major ISPs such as AOL (Time Warner) and other websites such as Lycos. If they did, the reach of Google and Yahoo would almost certainly be greater. However, these data do provide a first indication that the concentration of online search provision may not be a US phenomenon but, rather, a global one. It appears at first glance that these three companies – Google, Yahoo, and Microsoft – were mediating access to online information for a great proportion of the online public worldwide. Nevertheless, these figures may mask local differences and, in any case, they give little insight into the dynamics of the relation between the global and local markets for search.

In order to shed light on this relationship, this chapter examines four national contexts (China, Japan, Germany and South Africa) to highlight some of the contours of the global system of search provision as it had developed in 2006.

5.2 The four country cases

In 2006, internet penetration outside the United States was continuing to grow, with studies showing that the United States's 152 million internet users made up 21.9% of the number of active internet users globally (see Table 6).

Table 6: Active internet users by country, top 15 countries, March 2006

Country	Unique Visitors (000)*	% of Worldwide Unique Visitors	Cumulative % of Worldwide Unique Visitors
World Total	694,260	100%	n/a
United States	152,046	21.9%	21.9%
China	**74,727**	**10.8%**	**32.7%**
Japan	**52,100**	**7.5%**	**40.2%**
Germany	**31,813**	**4.6%**	**44.8%**
United Kingdom	30,190	4.3%	49.1%
South Korea	24,645	3.5%	52.6%
France	23,884	3.4%	56.1%
Canada	18,996	2.7%	58.8%
Italy	16,834	2.4%	61.3%
India	16,713	2.4%	63.7%
Brazil	13,186	1.9%	65.6%
Spain	12,452	1.8%	67.4%
Netherlands	10,969	1.6%	68.9%
Russia	10,833	1.6%	70.5%
Australia	9,735	1.4%	71.9%
South Africa**	**2,552**	**n/a**	**n/a**

* Those aged 15+ who have used the internet during the month. Excludes traffic from public computers such as internet cafe and, access from mobile phones or PDAs.

Source: adapted from comScore WorldMetrix (comScore Networks, 2006). **Data for South Africa from March 2006 from Nielsen NetRatings for the South African Online Publishers Association (OPA) (Nielsen//NetRatings, 2006b).

China, in particular, was forecast in 2006 to overtake the United States in the coming decade due to the very high growth of internet usage in the country; the US, by contrast, had a very low rate of growth in internet users[46]. Japan and Germany were also growing slowly. South Africa was adding users at a more rapid rate, but was expected to have a relatively small online population for the foreseeable future.

These four countries were chosen because they represent countries of different sizes and in different stages of development vis-à-vis the internet. China, Japan and Germany compose the group of large countries: taken together, these three had the three largest populations of active internet users outside the United States and together with the US comprised nearly half of the global population of internet users at the time the research was conducted. Examining how the market was developing in these countries gives us a

---

[46] By 2009 the expected had taken place: according to comScore, in January 2009 China had 17.8% of the worldwide internet audience, with the US in second place at 16.2%.

good basis for suggesting of what might have been happening for most internet users of the time. By contrast, while South Africa had the largest online population of any African country it was tiny in comparison to the United States or China. Many countries in the world were in a position more equivalent to South Africa, so although data about the internet and search engines is difficult to collect for most of the smaller countries, the South African case can help provide us with some insight into what was happening at the time in 2006 outside the larger internet markets.

5.3 Search Engines in Japan

This analysis of search provision outside the United States begins with Japan, the first foreign country where US-based search engines began operating[47]. In January 1996, Yahoo announced that it was forming a joint venture to create Yahoo Japan with Softbank, a Japanese investment firm which was one of Yahoo's original investors. In December of 1996 the *Financial Times* reported that Softbank at that time owed 36.33% of Yahoo (Nakamoto, 1996). In April 1996, Yahoo Japan was officially launched to provide an 'online guide' to Japanese internet users (Yahoo!, 1996).

Other US portals followed Yahoo's entry into Japan. Infoseek came to Japan soon after as a joint venture with Kanematsu Corporation, a *sogo shosha* or general trading company, in April 1996 (Infoseek, 1996). Like Softbank with Yahoo before it, Kanematsu also took a small stake in Infoseek. NTT, Japan's national telecommunications company, launched its search engine "goo" powered by Inktomi's white-label search product in March 1997 (Inktomi Corporation, 1997). Excite Japan followed in July 1997 as a joint venture with Itochu Techno-Science Corp., a major trading company, and Dai Nippon Printing (Martyn Williams, 1997). Lycos Japan launched in April 1998 as a joint venture with Internet Initiative Inc (a Japanese backbone provider) and Sumitomo Corp., one of the big six Japanese *keiretsu* (M. Williams, 1998).

---

[47] The challenges of researching specific situations in a country where one does not speak the language are substantial. This was the case for me with China and Japan. While material for the search-specific elements of the country-specific sections was gathered primarily through the use of resources such as Lexis-Nexis and the helpful archives of Search Engine Watch, for an overall view of the Japanese adoption of the internet and the general regulatory situation, I have primarily used Coates and Holroyd's 2003 book on the development of the internet in Japan. For a fascinating overview and historical treatment of the internet in China, I refer the reader to Zhou (2006).

Thus, while in the early stages of the Web in Japan there were many different search engines, there does not seem to have been independent Japanese development of these engines. Rather, search on the Japanese internet was provided by US companies in partnership with Japanese companies in joint ventures, in the tradition of the Japanese *keiretsu* relations.

Gradually, as their US parent companies closed or were acquired, as discussed in Chapter 4, the Japanese joint ventures of Infoseek, Lycos and Excite were left behind and their search capabilities vanished. As portals, they continued to live on their advertising revenues and began to buy in their search technology from others, primarily Yahoo-owned Inktomi (via NTT's goo website).

According to one general history of the internet in Japan, the Japanese government was slow to embrace the internet and, in particular, slow to reform NTT, whose high connection costs were seen as standing in the way of generalized internet adoption amongst the population (Coates & Holroyd, 2003). However, in 2000 the political situation and the economic value of the internet (then at the height of the dot-com boom) prodded the Mori government into action and, in 2000-2001, a series of government initiatives and grants was targeted towards creating a vibrant "Japanese IT Society." Amongst these were policies aimed at easing foreign firms' entry into Japan (Coates & Holroyd, 2003, p. 55). Possibly for this reason, Google's entry to the Japanese market was not in the form of a joint venture, as nearly all its rivals had chosen previously. Rather, Google entered Japan in April 2001 with three large search distribution deals: Nifty, Biglobe, and Yahoo Japan. They debuted their own Google Japan site at the same time. By June 2004, when Yahoo Japan began using its own technology, Google was estimated to have the largest share of search engine traffic in Japan, largely because of its agreement with Yahoo Japan. Perennially the smallest service, Ask Jeeves came late to Japan, debuting ask.jp in August of 2004, despite having negotiated a joint venture in 2000 with transcosmos, an IT-focused Japanese company specialising in outsourcing and marketing (Ask Jeeves, 2004).

By 2006, Yahoo Japan dwarfed every other website in Japan both in terms of reach (number of total users who visit within the month) and page views (see Table 7). It has

been suggested that, in many respects, Yahoo *was* the internet in Japan – it not only dominated search engines, but also had become the top e-commerce venue through its auction site and the second most popular provider of broadband internet through Yahoo BB. The next most-visited site for Japanese users had 54% of Yahoo's visitors and only 11% of its traffic. Yahoo Japan was a colossus of a website.

Table 7: Top 10 Japanese domains for February 2006

Domain Name	Users (000)*	% Reach	Page Views (000)
Yahoo.co.jp	33,316	84.52	19,284,776
Rakuten.co.jp	18,029	45.73	2,114,540
Nifty.com	17,374	44.07	745,850
Infoseek.co.jp	15,989	40.56	955,479
Amazon.co.jp	15,506	39.34	455,698
Biglobe.ne.jp	13,955	35.40	534,608
Geocities.jp	13,894	35.25	906,680
Goo.ne.jp	13,894	35.25	906,680
Microsoft.com	13,713	34.79	122,047
Ocn.ne.jp	13,536	34.34	333,655

*Users within the month aged 2+, access from home PCs

Source: Nielsen//Netratings Japan.

Moving down the list of most-visited Japanese websites from Yahoo, there are perhaps some unfamiliar names on the list alongside the US mega-websites Yahoo, Amazon, Yahoo-owned Geocities and Microsoft: Rakuten was a Japanese online shopping mall which later also bought Infoseek Japan (an offshoot of the former US search engine); Nifty, Biglobe, and OCN were internet service providers.

Neither MSN nor Google appeared on the Japanese top ten list in their own right. MSN was often mentioned as a top portal in the press, however – a June 2005 report of Nielsen figures cited in the press says it had 12.8 million users and a 41.48% reach (Asia Pulse, 2005). As of mid-2005 it supplied its own search results. The same report cited 5.5 million users and a 17.98% reach for Google's own Japanese domain, google.co.jp. In addition, Google supplied search results to six of the ten sites on this list: Rakuten (although it used another provider for its shopping results), Infoseek, Nifty, Biglobe, Goo, and OCN.

5.4 Search Engines in Germany

Historically, Germany had several important local search engine companies, such as Fireball and Web.de. These search engine names remained in 2006, but as the figures in Table 8 show, the German market primarily used search engines manufactured in the United States. Table 8 presents the top German websites as of July 2006, according to Nielsen NetRatings.

Table 8: Top websites in Germany by parent company, July 2006

Property Name	Users (000)	% Reach
Google	24,339	68.64
Microsoft	21,784	61.44
eBay	19,348	54.57
United Internet	15,815	44.60
T-Online	15,084	42.54
Time Warner	14,431	40.70
Bertelsmann	10,827	30.53
Yahoo!	10,283	29.00
Amazon	9,243	26.07
Wikipedia	8,708	24.56

Source: Nielsen NetRatings Home/Work Panel (Nielsen//NetRatings, 2006a)

In addition to Google's top spot at the head of the German website list at this time, it is worth noting that its search engine also powered the portals for T-Online and AOL Germany (Time Warner). In addition to the data in Table 8, in Germany there was also some evidence as to the share of *search queries* that each search engine was able to command. According to a presentation by the managing director of comScore Europe (Ivins, 2006), Google was overwhelmingly the choice of German searchers with 82% of all searches in Germany in March 2006 being conducted through Google.

Despite its prominence, Google was not the only search provider in Germany. In addition to Microsoft and Yahoo, two companies listed in Table 8 provided home-grown German search technology: United Internet and media giant Bertelsmann. United Internet was the parent of the internet portal Web.de. Originally a directory, in August 2004 Web.de launched its own SmartSearch technology which it had been developing since then (Stuttgarter Nachrichten, 2004), launching (among other services) picture searching and a full integration with the German version of Wikipedia. The second

German engine on the list is Lycos Europe (which was jointly owned by Bertelsmann, Telefónica of Spain and public shareholders and is a different company than the US-based Lycos, Inc.). The Lycos Europe websearch (Lycos IQ) was powered by Fireball search technology, a local company owned by Bertelsmann and merged with Lycos Europe in 2000.

In fact, local search engines had made something of a comeback in Germany by 2006, with the realisation that pay-per-click advertising was a viable model for search engine businesses. For example, a new company, Seekport, had been created by former employees of Infoseek and had launched in Germany, France, and the United Kingdom. These companies seemed to be looking to succeed European competition rather than in global competition. According to the Seekport website, for example, "it is the declared aim of Seekport to become one of Europe's leading search engines," (Seekport, 2006). Lycos Europe's CEO Christoph Mohn said, "With our expansion we can get back to a point where we rival, say, Yahoo in Europe, but I don't think we're now going to rival it globally. My aim, though, is to build a European service that can rival the big players in Europe," (quoted in New Media Age, 2006)[48].

5.5 Search Engines in China

In 2006, China had the world's second-largest online population. However, this represented only approximately 5% of the total Chinese population. In most respects China's internet population at this time resembled the early-adopter profile of the US market in the mid-to-late 1990s: overwhelmingly young, male, urban, and affluent (China Internet Network Information Center, 2006). According to CNNIC, which produces the official internet statistics in China, about 80% of people were using use search engines (roughly the same percentage as in the US) and searching was the second most popular application after email (again, mirroring US figures).

The search business, too, resembled the US in the late 1990s (discussed in Chapter 4), although with some differences. Three large portals dominated the Chinese online landscape: Sohu, Sina, and NetEase, all public companies traded on the US NASDAQ

---

[48] In November 2008, Lycos Europe announced it was to shut down and sell its remaining assets.

exchange. The market structure of these portals, however, has been markedly different than in the US. At this time, they were making their money by selling short messages delivered by mobile phone, by pay-per-use online games purchased via top-up cards at local stores and by gambling. Advertising was at best a secondary revenue stream. Until about 2005, therefore, these companies were content to purchase their search services from other search providers. However, in 2005 the Chinese government restricted the messaging services which formed the bulk of the portals new revenue, following complaints from users about the amount of unwanted messages they were receiving (and being charged for). The major portals then began casting around for new revenue streams, and competition in the search market began to increase.

Although neither of the major online ratings agencies (Nielsen//NetRatings and comScore) had released search figures for China, two other reports on searching in China were published in this period. The first consisted of official statistics from CNNIC (Lu, 2005) and was for major metropolitan areas only (Beijing, Guangzhou, and Shanghai); the second set of statistics was available from a commercial research company, iResearch (2005), and covered the whole of China. Both of these datasets were based on surveys rather than on the user tracking devices installed in home PCs that Nielsen//NetRatings and comScore use, although iResearch also uses Alexa data which reports popularity via a toolbar installed in the users' browser. The market share figures from these two reports have been amalgamated in Figure 17 for the period up to late 2005. The picture is remarkably consistent: Baidu (a local search engine) is highest, followed by Google, Yahoo/Yisou (Yahoo's Chinese search engine) and Sohu/Sogou (another local search engine).

Figure 18: Search market share in China***
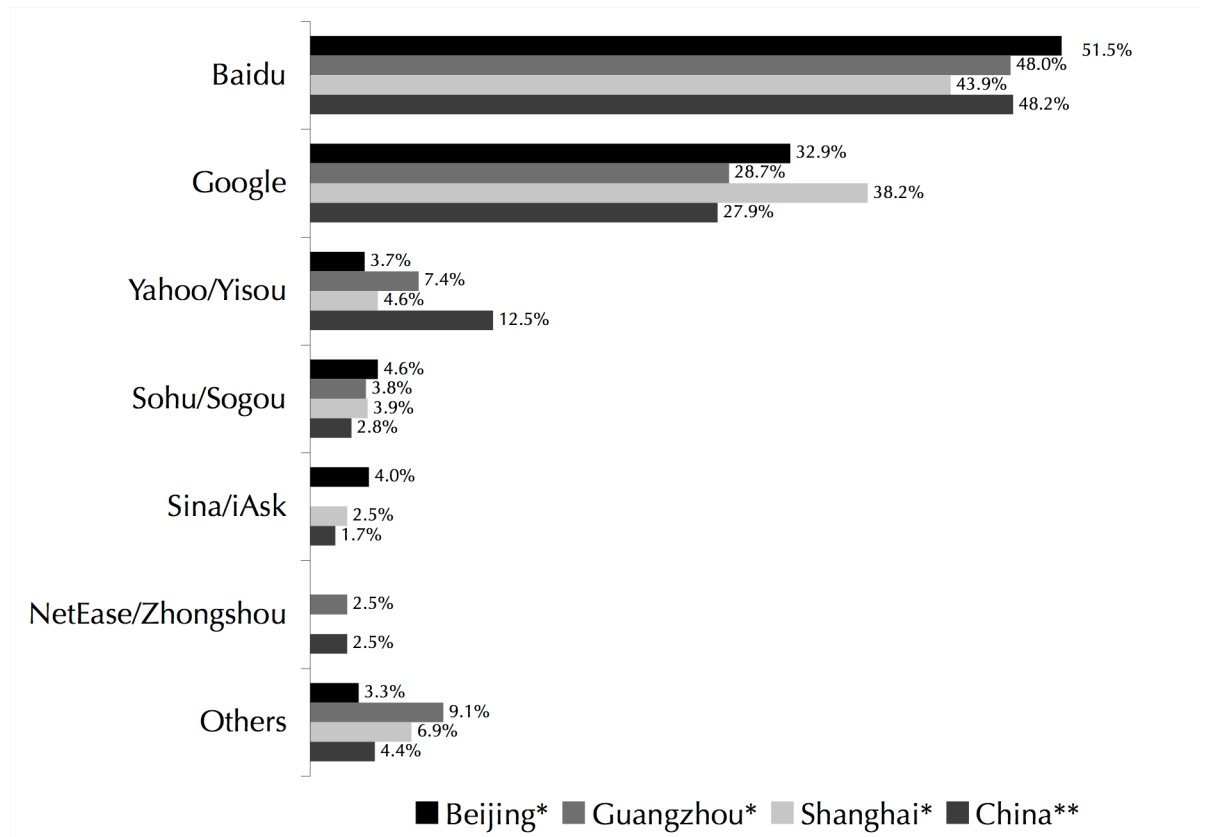
Source: *Lu (2005), **iResearch (2006).

***Market share for the Lu/CNNNIC study is defined as number of people saying the search engine is the primary or only search service divided by number of total users; iResearch uses the term "traffic share" but does not define it precisely.

It should be noted that the percentages in Figure 17 represent *market share* which will be a smaller number than the *reach* figure cited for other markets, since market share measures only the main or primary search engine, whereas reach considers *all* the search engines a user visits. Interpretation of the figures above is also complicated by the fact that in China a very common use of search engines is to find MP3 music files. Baidu, particularly, was used for searching for downloadable music, as if Google and iTunes were offering a single integrated service. According to iResearch, MP3 search accounted for 14.5% of all search traffic at this time.

While there is evidence of concentration in Chinese search resultsin this period, the companies that were providing search services to China were not necessarily the same companies as those providing search services in the United States. In particular, the local

search engine Baidu was very popular; there were also two other significant Chinese search engines called Zhangsou and SoGou.

Zhangsou (formerly Huicong) developed from searching databases of corporate information (similar to early Canadian search engine), and began to provide search to a range of important Chinese websites, developing what it called the "China Searching Alliance" of some 500-600 websites (SinoCast China IT Watch, 2003). Initially focusing on business information, it began broadening its remit and by 2006 was providing Sina and NetEase with their Web search. The fact that they had been chosen to provide search on major government sites also perhaps indicates their closeness to the Chinese state. Zhangsou, given the services that it provided and its commercial model, based on licensing its technology to other sites, is perhaps best thought of as the Chinese Inktomi.

Baidu, on the other hand, had acquired the reputation of being the Chinese Google. Baidu was actively modelled on Google. Its founder, Robin (Yanhong) Li, was a Chinese engineer working in Silicon Valley in 1999 who saw the great potential of providing a modern search engine using link-based algorithms to the Chinese market. Li completed two rounds of venture funding in the Valley in 1999 and 2000, raising $11.2 million and began to build up Baidu's technology and index in Beijing (SinoCast China IT Watch, 2004). In 2004, Google helped Li lead a second round of financing estimated to have been worth $25 million, in which eight investors participated including Google, which spent $4 million to purchased a small stake (about 4%) in Baidu (Francisco, 2004). In 2005, Baidu made an initial public offering on the NASDAQ market, following Google's own spectacular debut and, overnight, became the most valuable Chinese internet company, briefly worth $4 billion (Economist, 2005). Finally, SoGou was a return by Sohu to providing its own search technology, following the collapse in its messaging revenues (Total Telecom, 2004).

Google first developed a Chinese version of its search engine in the US, but after difficulties in operating through China's content firewall – and also reportedly because the Chinese government refused to let it report revenue received in China without a physical office – it announced in May 2005 (a month before Baidu's IPO) that it would be opening

an office in China after having secured an agreement with the Chinese government (Kopytoff, 2005).

Until 2003/2004, Yahoo maintained a token presence in China, purchasing its search results from Baidu or Google and maintained a portal that paled in terms of traffic in comparison with local rivals Sina, Sohu and NetEase. However, in late 2003 it purchased 3721, a local firm which enabled Chinese characters to be input into the address bar of a brower and, in June 2004, it debuted its own search technology, Yisou (AFX News, 2004). This decision was apparently motivated by competitive positioning rather than forecasted earnings – David Lu, the deputy managing director and VP of search and marketplace for Yahoo North Asia, told one reporter "We have no plans for monetization. All we care about is technology and user experience. We have to look at market maturity. Although we have a good business model in the US, we don't know if it can be implemented in China," (quoted in AFX News, 2004).

The quote from the Yahoo representative indicates just how speculative search investments in China were at this time due to the fact that the paid-search business model was not nearly as effective there. Most of the population in China is rural and agricultural. In particular, a very large majority of them is without a credit or debit card. Electronic payment fuels the consumer internet economy – without it, small businesses are unable to process online orders. Indeed, most online orders in China (about 50%) were completed at a bank or post office, another 25% were COD (cash on delivery), and only 25% were completed online. Chinese consumers, in addition, did not have the same types of protection from fraud or shoddy goods and were reputed to be extremely concerned lest the quality of online goods not prove to be up to standard (New Media Age, 2005). Only mobile phones had developed payment systems through standard billing (going some way to explaining the popularity of SMS services for the portals). Otherwise, online games and other pay-per-use services were being funded by top-up cards purchased from local retailers.

The effect of the uncertainty and difficulty of online payment was to severely dampen consumer e-commerce activity in China. In the US, it is small e-commerce shops rather than major brand advertisers, registering and purchasing ads online that form the bulk of

contextual search advertising – and this type of advertiser was much rarer in China than in the wealthy economies (Marshall, 2005) at the time of Marshall's research.

5.6 Search Engines in South Africa

South Africa is the smallest country in terms of internet population that is examined in this chapter, and relatively little information was available on search services in South Africa or on the internet market as a whole in the country (for an interesting history of the conflicts which have shaped the modern South African internet, see Lewis, 2005). The best information came from the Online Publishers Association (OPA), a confederation of South Africa's largest online website publishers, who measure traffic to members' websites in association with Nielsen//NetRatings. Unfortunately, the OPA's figures only count websites originating in South Africa. Table 9 shows the top 20 of these websites and calculates reach based on the number of internet users for July 2006 from the Nielsen NetRatings/OPA figures.

Table 9: Top 20 websites originating in South Africa

Website	Q2 2006 Unique Users*	Q2 2006 Page Views*	Reach**
www.mweb.co.za	608,882	14,682,977	23.5%
news24.com	588,832	14,163,114	22.8%
iol.co.za	412,688	6,744,698	16.0%
ananzi.co.za	294,912	2,485,603	11.4%
fin24.co.za	225,596	3,126,550	8.7%
iafrica.com	202,645	7,342,343	7.8%
careerjunction.co.za	179,172	8,723,442	6.9%
health24.co.za	174,523	3,081,222	6.7%
mg.co.za	148,900	1,533,296	5.8%
wheels24.co.za	132,597	1,730,291	5.1%
women24.com	132,916	1,233,263	5.1%
yellowpages.co.za	126,000	2,343,244	4.9%
www.privateproperty.co.za	119,949	4,433,166	4.6%
Property24.com	122,664	3,017,867	4.7%
sundaytimes.co.za	107,547	1,534,867	4.2%
Bizcommunity.com	91,130	2,191,196	3.5%
tonight.co.za	98,779	599,508	3.8%
ioljobs.co.za	99,221	2,760,961	3.8%
Total for SA July 2006	2,586,312	127,317,739	

*Figures count only users and page views originating from within South Africa.

** Reach calculated as unique users for a website divided by total users.

Source: Nielsen//NetRatings for OPA (Nielsen//NetRatings, 2006b), author's analysis.

An alternative source was data collected from the Alexa toolbar, which gives rankings within South Africa but no further data. The top 20 sites in South Africa according to Alexa as of September 2006 are shown in Table 10:

Table 10: Top websites viewed in South Africa

Rank	Website name
1	Google.co.za
2	Google.com
3	Yahoo.com
4	msn.com
5	Wikipedia.org
6	Mweb.co.za
7	News24.com
8	Microsoft.com
9	Bbc.co.uk
10	Blogger.com
11	Live.com
12	Iol.co.za
13	Standardbank.co.za
14	Absa.co.za
15	Myspace.com
16	Amazon.com
17	Iafrica.com
18	msn.co.za
19	Ebay.com
20	Za.net

Source: Alexa rankings (Alexa, 2006)

Unfortunately, this deficiency of data leaves us in some doubt as to how important US based search engines were to the South African online public in 2006, although from the Alexa data we can suggest that they were playing a major role. This was borne out by a qualitative examination of the South African market.

The top two local websites in South Africa as identified both by the OPA and Alexa were mweb.co.za and news24.com. These were both portals and were both owned by newspaper group Naspers (which also owned the ISP Tiscali South Africa). MWeb.co.za used Google's South African search engine. News24 also appeared to use Google's engine to offer search to its users; however, it had also said that it plans to 'make searching local content easier and will be "tweaking" 24.com's search engine to improve the relevance of local search result. [CEO Kim Reid] said it aims to be "better than Google on the local front."' (Burbridge, 2006). Naspers had also integrated a Web development business, including a shopping channel, into its own digital arm, so it seemed that it might be developing this local search itself.

There were also two other local search engines of note. Ananzi was founded in 1996 and offered both a local directory and search facility which was based on the Verity engine (an enterprise search product most often found in corporate websites rather than on Web search sites). Its competitor, Funnel, was linked to Naspers main competition, the IOL portal from the rival Independent newspapers group (Sunday Times (South Africa), 2005). Funnel was founded in May 2005 and its tagline was 'Proudly searching ONLY South Africa.' According to news reports, Funnel introduced the pay-per-click advertising method to South Africa (Liquid Africa, 2005), although 24.com also offered pay-per-click ads in 2006.

Yahoo had no local South African operation in 2006 after the collapse of local representation in 2002 (ITWeb, 2005). Google launched its google.co.za domain in January of 2005, including interfaces in the local languages of Afrikaans, Sotho, Xhosa and Zulu as well as English (ITWeb, 2005). The managing director of Ananzi, Mark Buwalda, indicted at the time that he did not consider Google to be a threat because he believed it operated in different markets: 'Large numbers of South African searches still come to us, whereas the international searching goes to Google and Yahoo, more so to Google.' (quoted in ITWeb, 2005). In 2006, however, Google had been hiring new staff in preparation for opening a South African office. Buwalda was more cautious when he spoke to the press at the time: 'We always have to be careful but we operate in a global environment. If the largest player in the world is opening up a full-scale office in SA, it's good for everyone' (quoted in Business Day, 2006).

The direction in which South African search engine development would take place was unclear. In 2006, there appeared to be a two-tier system in which international (primarily American) websites were used most often and South African sites were used to supplement searches locally when the US based search engines were unable to deliver relevant local content. If, on the other hand, Google would be able to improve its local search results with its imminent local presence, it might be that local South African search would be a thing of the past.

5.7 International search engine roduction

Having considered the four case countries, we can summarise the results as follows: local search engines with their own technology existed in all the case countries apart from Japan. Nonetheless, the US-based engines of Google, Yahoo, and Microsoft were present in each country, vying with each other for reach and advertising spend, while search engines from other parts of the world had, at best, a regional strategy for expansion. Neither Baidu nor Lycos Europe, the largest local providers in the countries examined here (and bearing in mind that these countries with the US represented just under half of the total worldwide internet population), had announced any intention to compete in the United States, which represented two-thirds of the world's online advertising market. The South African model, with people differentiating between 'international' search (for which they used the American versions of the search engines) and 'local' search (for which other services like Ananzi or Funnel might be used) may have been replicated in other markets.

Based on the cases developed here, therefore, we can characterise the global search market as a two-tier system at the time the research was carried out, with US-based providers competing in what were perceived as core markets and national providers competing in national or regional contexts. This analysis of the situation provides a basis for a further investigation of the factors that contributed to the creation of this two-tier system. Following the conceptual framework (as developed in Chapter 2) the history, economic structure and regulation of search may is likely to provide insight. This next section examines, first, the operational structure of the US search engines; second, the market for online advertising in various countries; third the overall level of technical expertise; and, finally, the political and regulatory environments of search, with special attention to the situation in the case countries.

5.7.1 INTERNATIONAL SEARCH ENGINE OPERATIONS

The US-based search engine companies have treated each case country differently. Yahoo, in particular, operated through joint ventures in some countries, as discussed with Yahoo Japan. Even when there is not a joint venture, a physical presence may be required; as we saw, Google was compelled to open local operations with offices on the ground in China in order to be able to operate in that market. These operations may, in practice, be more or less autonomous of the parent company, particularly in the case of Yahoo. Light

can be shed on this two-tier system by examining the international operational structure of the US-based search engines.

Google had registered services in one hundred and fifty-seven countries in 2006, according to its websites but, operationally, it listed offices in twenty-four countries[49]. Eighteen offices were designated as research or engineering offices. Of these, eleven were inside the United States (in Mountain View, CA; Irvine, CA; Santa Monica, CA; Seattle, OR; Phoenix, AZ; Boulder, CO; Chicago, IL, Pittsburgh, PA; New York, NY; Atlanta, GA; and Cambridge, MA); and seven were in other countries (London, UK; Trondheim, Norway; Zurich, Switzerland; Bangalore, India; Tokyo, Japan; Seoul, South Korea; and Sydney, Australia).

Yahoo, at the time of this research, did not disclose its office locations and they were not available on its corporate information website. According to a representative from its press office, this was to protect the security of its employees. Nevertheless, a review of its human resources website showed office locations in twenty-five countries which corresponded with the number listed in its 2006 Annual Report[50]. Engineering jobs were advertised in the following offices: Sunnyvale, CA; Santa Monica, CA; Burbank, CA; San Diego, CA; Dallas, TX; Grenoble, France; Frankfurt, Germany; Munich, Germany; London, UK, and Bangalore, India). The distribution of Yahoo offices in 2006 is shown in Figure 19:

---

[49] Google had offices in: the United States, Canada, Mexico, Argentina, Brazil, the United Kingdom, France, Germany, the Netherlands, Spain, Switzerland, Denmark, Norway, Sweden, Finland, Turkey, mainland China, Taiwan, Hong Kong, Singapore, Japan, Korea, India and Australia.

[50] Yahoo had offices in: the United States, Canada, Mexico, Argentina, Brazil, the United Kingdom, Ireland, France, Germany, the Netherlands, Spain, Switzerland, Italy, Denmark, Norway, Sweden, mainland China, Taiwan, Hong Kong, Singapore, Japan, Korea, India, Australia and New Zealand.

Figure 19: Yahoo! offices worldwide

Map key: Yellow marks: headquarters; green marks: engineering offices; blue marks: other offices.

Source: Map created using Google Maps, by author. Data from Yahoo! websites and author's analysis; office locations as of 2006.

The Yahoo! office map is instructive because one can see the clustering of engineering offices on the West Coast of the US and in western Europe, with sales and marketing offices much more widely spread. A map of Google offices would be expected to have a similar distribution. MSN offices were not disclosed separately from Microsoft offices. Microsoft itself operated in over 90 separate countries, while the MSN service targeted 39 countries, although it almost certainly had staff in fewer.[51] The Microsoft human resources website listed engineering "Development Centres" in Redmond, WA; Cambridge, UK; Dublin, Ireland; Copenhagen, Denmark; Aachen, Germany; Haifa,

---

[51] The MSN service targeted: the United States, Canada, Mexico, Argentina, Brazil, Chile, Latin America (unspecified), the United Kingdom, Ireland, France, Germany, the Netherlands, Belgium, Spain, Italy, Switzerland, Austria, Denmark, Norway, Sweden, Finland, Turkey, Israel, Arabia (unspecified), Russia, mainland China, Taiwan, Hong Kong, Singapore, Japan, Korea, India, Indonesia, Malaysia, the Philippines, Thailand, Australia and New Zealand. It is unlikely to have staff or production resources in all these countries but unfortunately this cannot be determined.

Israel; Hyderabad, India and Beijing, China which, once again, is similar to the Yahoo and Google patterns.

Thus, the picture of activity from a search engine operations point of view at this time was split into producing and receiving countries. Countries that were producing search engine technology included the United States (especially the west coast); western Europe (particularly the UK, Ireland, Germany, France, Switzerland and the Scandinavian countries); India; and East Asia (especially mainland China, Japan, Singapore and South Korea). Other countries and regions, including South America and Latin America, eastern Europe, Africa, the Middle East and Australia and New Zealand were receivers of search engine services with few exceptions. This group of receivers was further split between those who received search services specialised or dedicated to their country (e.g., Finland) and those who did not (e.g., Ghana).

5.7.2 ADVERTISING MARKET

Since the economics of search engine provision in the United States are based on advertising and, particularly, on keyword-linked cost-per-click advertising as discussed in Chapter 4, the level of search provision in different national settings may reflect the *existing or potential market* for that advertising. In other words, it might be related to the online population of the country or to the amount spent on online advertising. It might also be related to the volume of Web pages produced by that country, onto which syndicated advertising from search engines might be distributed.

The online advertising market gives a crude measure of the potential economic value to search engines of entering any particular market, bearing in mind the large share of the online advertising market that search engines command in the United States. Globally, the online advertising market was very uneven at the time of this research, with two-thirds of expenditure in the United States. Japan and Germany both also had large advertising markets, though they were small in comparison to the United States. The advertising markets of China and South Africa were tiny by comparison, representing only 0.2% and 0.1%, respectively, of the total world online advertising market (see Table 11).

Table 11: Online advertising expenditure by country, 2005

Rank	Country	2005 Online Advertising Expenditure, US$ millions	% of 2005 World Online Advertising Expenditure
	World total	14,767.53	100.0%
1	USA	9,896.35	67.0%
2	**Japan**	**1,962.68**	**13.3%**
3	United Kingdom	514.57	3.5%
4	**Germany**	**478.47**	**3.2%**
5	Australia	349.17	2.4%
6	Canada	280.2	1.9%
7	Sweden	199.12	1.3%
8	Spain	143.66	1.0%
9	Denmark	131.82	0.9%
10	Italy	127.40	0.9%
11	France	111.04	0.8%
12	Brazil	75.63	0.5%
13	Norway	61.19	0.4%
14	Netherlands	58.70	0.4%
15	Austria	45.36	0.3%
16	Russia	42.32	0.3%
17	Belgium	35.42	0.2%
18	Switzerland	33.29	0.2%
19	Finland	29.99	0.2%
20	**Hong Kong, China**	**27.38**	**0.2%**
28	**South Africa**	**7.72**	**0.1%**

Source: Euromonitor International from World Association of Newspapers / Jupiter Research (Euromonitor, 2005)

The figures in Table 11 allow us to categorize the case countries: China, Japan, and Germany had the largest online populations outside the US and so represented potentially large markets for online retailers that advertise on search services. Japan and Germany's online advertising markets were, in fact, already the largest outside the US, with the exception of the United Kingdom. China and South Africa, on the other hand, had little advertising online at the time.

In addition to the advertising market at a given time, another important indicator of the state of the market is local Web page production which can serve as a proxy for technical expertise and also provide an indication of the venues available for online advertising within a particular national context – that is to say, the *potential* online advertising market. Reliable figures about the number of Web pages produced in each country were extremely difficult to obtain. One method which gives an approximate comparison is the number

of domain names. Each country in the world has a top-level domain associated (ccTLD) with it (for example, .jp for Japan or .de for Germany). However, organizations throughout the world may also register a generic top-level domains (gTLD) such as .com or .net, complicating the situation. Fortunately, Zook (2005) has done an analysis which enables the comparison of the global online production of Web pages (see Table 12).

Table 12: International distribution of domains, July 2003

Country	Country code	gTLD	ccTLD	Total	Per 1000 population*	Percent of world's domains
World total		28,121,541	22,121,541	50,250,943	8.4	100.0%
USA	us					
		16,111,005	597,984	16,708,989	59.0	33.3%
Germany	**de**	**1,498,239**	**6,491,981**	**7,990,220**	**97.4**	**15.9%**
UK	uk	2,234,532	4,327,511	6,562,043	110.5	13.1%
Canada	ca	1,070,259	352,800	1,423,059	46.2	2.8%
China**	**cn/hk**	**982,665**	**318,181**	**1,300,846**	**1.0**	**2.6%**
South Korea	kr	749,786	548,486	1,298,272	27.8	2.6%
Italy	it	436,145	818,874	1,255,019	21.8	2.5%
Netherlands	nl	333,224	904,011	1,237,235	77.8	2.5%
Japan	**jp**	**409,750**	**519,653**	**929,403**	**7.3**	**1.9%**
France	fr	735,497	168,538	904,035	15.3	1.8%
Argentina	ar	51,189	750,000	801,189	21.7	1.6%
Australia	au	353,500	347,576	701,076	36.7	1.4%
Switzerland	ch	165,924	530,838	696,762	96.8	1.4%
Brazil	br	79,118	488,295	567,413	3.3	1.1%
Denmark	dk	54,822	438,863	493,685	93.2	1.0%
Spain	es	402,291	48,933	451,224	11.3	0.9%
Austria	at	90,313	287,194	377,507	46.6	0.8%
Belgium	be	98,910	263,997	362,907	35.6	0.7%
Sweden	se	187,467	172,953	360,420	41.0	0.7%
Taiwan	tw	41,800	226,551	268,351	11.9	0.5%

* Population figures from 2000, ** Figures for China include Hong Kong

Source: Adapted from Zook (2005, p. 19)

Table 12 shows Germany at the top of the list after the US, and China, somewhat surprisingly, several places ahead of Japan at numbers five and nine respectively. Unfortunately, comparable data for South Africa proved impossible to obtain. Using Web pages as a proxy for technical expertise indicates that Germany had the greatest level of expertise outside the United States. China, by the same measure, also scored well, perhaps surprisingly leading its more developed neighbour Japan. If those same Web pages were also potential venues for syndicated online advertising for search engines, then certainly

China represented the market with the most potential for growth for search engines based on this analysis.

It was initially expected that US search engines would be the most active in countries with the most developed online advertising markets and would be at least visible in countries with a potentially high advertising market. Further, it seemed logical that local search engine activity would be strongest where local technical capacity was highest. These initial expectations were borne out in broad outline in this analysis. Yahoo and Google overwhelmingly dominated the large online advertising markets of Japan and Germany, respectively, in terms of search provision. Both companies had entered China, with its potentially huge market, investing in technology without being confident about whether their advertising model would be replicable and despite the considerable negative press they received because of the requirements of the Chinese state to censor their content and disclose user data on request (e.g. Amnesty International, 2006). South Africa, on the other hand, had received relatively little attention from the US search services, although Google had launched a South African service from the US. The countries with the highest number of Web pages and, therefore, potentially the greatest technical capabilities, Germany and China, also had the most active local search engines, with Germany supporting several engines with European ambitionsand local Chinese engine, Baidu, taking the top spot from Google in China.

5.7.3 POLITICAL AND REGULATORY ACTIVITY

The development of search engines around the world is affected by local political and regulatory activity. From their earliest history, search engines have been involved in a range of legal actions, ranging from individual suits brought against them by individuals and companies to major legal campaigns that have pitted them against governments around the world. The major regulatory context is the United States, where although several laws pertain to search engine activity no search engine "policy" has been articulated. This section examines the emergent legal and regulatory position of search engines in the United States. Political and regulatory activity in the case countries is then considered.

McQuail (1994, pp. 171-173) draws attention to the three contrasting models of regulation of different media systems: the free press model, the broadcasting model and the common carrier model, each derived from a different perception of how the idea of the 'public good' can apply to media, which are recognised to have a public function in society in addition to a commercial function. The free press model aims to ensure diversity of content by preventing interference by the state in matters of content and distribution. The broadcast model aims at achieving high quality and representativeness in a restricted medium through public regulation of content and distribution. The common carrier model aims to achieve universal participation, normally in a peer-to-peer medium such as the post or telephone.

Different elements of the internet system are regulated in ways which correspond to different elements of McQuail's typology. For example, there has been an increasing tendency in both Europe and the United States for internet service to be considered in universal service provisions. The FCC, for example, while not directly regulating internet access, supports schools and libraries in getting internet access with the Universal Service Fund. Again, while internet access is not mentioned in the EU's Universal Service Directive (2002/22/EC), the European Commission's eEurope 2005 plan (2002/263/COM) has "e-inclusion" or "providing access for everyone in order to combat social exclusion" as a general aim. More recently, net neutrality advocates have worked toward the right to have all data bits carried equally by internet service providers (ISPs). Thus in infrastructure, McQuail's common carrier model seems to be the most applicable. With regard to online content, McQuail's "free press" model is prevalent, and search engines most typically are considered within this framework (see the discussion on case law, below). McQuail's "broadcast" model, including elements of public-service programming, is noticeable by its absence.

Through the development of case law in the United States search engines have gained a series of protections that enable them to operate more or less independently from outside scrutiny[52]. As a result of decisions in a range of lawsuits, by 2006 search engines enjoyed

---

[52] For the development of this section, I am indebted to the several comprehensive reviews of case law as it applies to search (Gasser, 2006; Grimmelman, 2007; Lastowka, 2007), to which the interested reader is directed for a fuller presentation of the issues involved.

broad protection in their core businesses of generating search engine results and selling. First, the courts had accepted the argument that search engine results were subjective opinions and therefore deserved the protection of the First Amendment, which guarantees the freedom to express an opinion (see Lastowka, 2007, pp. 23-25), which protects the search engines against suits from aggrieved parties who contend that their poor rankings were the result of neglect or malice by the search engine companies. Second, the courts had also accepted that spidering (where the search engine visits pages to add their content to its index) did not constitute trespass upon the spidered servers (Grimmelman, 2007, p. 25) (despite this, in practice, however, a robots exclusion file was widely used to block unwanted spidering). Third, courts had held that the creation of a copy of any Web page to go in the search engine index, and the subsequent display of a snippet, cached copy, or even thumbnail picture, fell within the "fair use" provision of copyright (Grimmelmann, 2007, p. 27). Fourth, courts had often held that the use of trademarked terms is permissible in targeting keyword advertising, although in some cases trademark owners' claims had been upheld (for details, see Lastowka, 2007). Fifth, search engines had been generally held free of liability for the content that they index in regard to suits for defamation or loss of privacy (Grimmelmann, 2007, p. 40). Despite these protections, search engine companies had been less successful in resisting government calls for access to user data and/or search data and have also acceded to requests for censorship from a range of governments, including China and Germany among others (Grimmelmann, 2007, p. 43).

Search engines have also become involved in wider legal debates regarding the future regulation of internet infrastructure. Google, for example, has been very active in supporting so-called "net neutrality" provisions, which argue against distinguishing between different types of traffic on the internet. Google's argument is that the tiering of traffic would harm the internet as a whole (see Cerf, 2006). However, their economic interest in this is also plain: should traffic discrimination become the norm, Google and other search engines might become liable to charging by ISPs or telecommunications providers on their key revenue source. There has also been government discussion of search engines in the US as they relate to other countries with regard to protecting the "free flow of information and ideas." This has been especially of concern with regard to China. The US government promotes search as part of the "free flow of information"

online, a continuation of its earlier media policies, and just as clearly opposes the censorship of search engines by other governments. In February 2006 two subcommittees of the US Congress, the House Subcommittee on Africa, Global Human Rights and International Operations and the House Subcommittee on Asia and the Pacific jointly convened to investigate the operations of Yahoo, Google, MSN and Cisco in enabling the Chinese government to establish tight control over use of the internet in China, including censorship of search results, particularly by Google.

In the US, legal scholars have begun to urge a consideration of search engines in the round, arguing that the different principles which apply to search engine results may, in some cases, work against each other, making balanced law and policy more difficult, and that the legitimate interests of different parties are often in conflict. For example, the users' interest in privacy and the search engines' interest in competition or innovation may be opposed. The objective, according to Grimmelmann, is to create a balanced framework in which each party's interests are properly considered. An unbalanced framework, by contrast, has the power to do great harm because of the central position of search engines: "Search engines do not generally cause harms out of inherent malice. They cause harms in the process of serving their other constituencies. It is precisely the fact that search engines create enormous value that gives them such power to cause enormous harms" (Grimmelman, 2007, p. 15).

Eric Goldman (2006), in considering search engine bias, reaches the conclusion that regulation by the government is un-necessary and potentially harmful, suggesting that technological innovation and market competitiveness will render the problem obsolete. His view may be said to represent one extreme end of the spectrum of views on search engine policy, where other scholars contend that at least some clarification of the law and ideally some further co-ordinated regulation may be necessary. Gasser (2006), Grimmemann (2007) and Pasquale and Bracha (2008) have argued for the consideration of search engine law holistically and have suggested different avenues and potential frameworks for policymakers. Gasser (2006) and Pasquale and Bracha (2008), evoke a media analogy and suggest that public interest or the public good are at stake.

Gasser (2006), who also considers European policy contexts, proposes a framework more specifically targeted at shaping future regulation and policy towards search engines, based on "core values of a democratic information ecosystem from which one might derive normative criteria for the assessment of search engine governance proposals" (Gasser, 2006, p. 126). He also emphasises the need to craft a holistic approach to search engine policy which avoids premature intervention and rests, rather, on the issues which search engines seem to embody and the principles derived from those issues. He suggests guiding principles that should be observed when crafting search policy (Gasser, 2006, p. 154-157). The first principle is access to search (both in terms of availability of search services to users and the ability of users to have their content included in indices and ranked fairly). Second, he suggests a principle of informational self-determination in which data collection about users is "otpimised or, preferably, minimized" (2006, p. 155). Finally, transparency is, he believes, a third key principle, although he admits its implementation may be difficult.

Pasquale and Bracha (2008) discuss search engine regulation within the context of other media regulation, saying "The spectre of control by a handful of powerful gatekeepers over a critical bottleneck of informational flow threatens the openness and diversity of the internet as a system of public expression," (2008, p. 13). They focus on "speech" as a key legal term, finding that the contention that search engine results should be considered "opinions" insupportable and that the protection of free speech should not blanket the production of search engine results. The key regulatory challenge, they believe, is the "proper balance between secrecy and transparency..." (2008, p. 29)

In contexts outside the US, other regulatory mechanisms have appeared. In Germany, concerns about search engine practice led major search engine providers to institute a form of self-regulation by signing a voluntary code of conduct (Freiwillige Sebstkontrolle Multimedia-Dienstabieter, 2004), under threat of government regulation. This code of conduct provides for transparency of operation and of exclusion criteria, that advertisements be clearly labelled as such, that search engines take precautionary measures to protect children and young people from unsuitable content, and that "data economy" in the collection and holding of personal data will apply. Special provisions apply to ensure that search engines comply with German law banning certain types of speech such

as Nazi propaganda, Holocaust denial, child pornography or content that glorifies war or degrades human dignity. In China, the government has taken an active interest in search engines, and it both monitored and censored search results (Johnson, 2005). The Chinese government also seemed to informally support some local Chinese companies such and Zhongshou, and to allow other companies, such as Baidu, to operate without undue interference. In Japan there was little government attention to search engines, but the Japanese government announced that they were creating a consortium of universities and businesses to run project to create a Japanese-built search engine; a similar government-sponsored search engine was announced by the German and French governments working in cooperation (Associated Press, 2005; Litterick, 2005).

5.8 Conclusions

The analysis in this chapter has considered the spatialisation dynamics operating in the search engine industry. Search engines may appear at face value to "annihilate" space, so that content from anywhere in the world can be found by anyone in the world. However, upon closer examination, search neither search engine operations nor search engine coverage appear to be evenly distributed.

In all four countries we reviewed, which, with the US, represented nearly half of the global online audience in 2006 (including examples in Asia, Europe, and Africa) the US-based search engine companies, Google, Yahoo and MSN, are either in fierce competition with local providers (China) or dominant in the marketplace (Japan, Germany, and South Africa). This is perhaps not surprising, since in terms of technological expertise, internet population, and especially online advertising expenditure, the United States in 2006 was an internet superpower dwarfing its closest rivals; the US-based search engines therefore had many local advantages.

The production centre of the search engine market is the west coast of the U.S., followed by Europe, with outposts in China and India. Dedicated search provision, as opposed to search production, was closely related to the existing and potential online advertising markets. The periphery of receiving nations is divided into active receivers of search services with larger advertising markets, for whom search services are tailored and incidental receivers with small advertising markets that may have quite inadequate service.

One of the spatial dynamics of search engines, therefore, seems to be a flow of search services from richer nations that receive dedicated services, to poorer nations whose service may be compromised. In this way the dynamics of the search industry are resonant with the one-way flow model that Wildman (Wildman, 1994) suggested operates in the newspaper market.

However, the evidence from the search engine industry in 2006 does not show a simple one-way flow. In China, Baidu was able to adapt the search business model to cater for the culturally specific content tastes and local payment methods, as well as to tap into financial capital available from the US stock markets, and thereby compete successfully with Google. In Germany, local concerns prompted the US engines to sign up to a self-regulation agreement guaranteeing greater transparency. Efforts to compete with US-based search engines were also underway, sometimes with government subsidies (in Japan and Germany). China, Germany, and Japan were all host to search engine development offices. The local search industry in these nations participated, although to a lesser degree than search companies based in the US, in the flow of ideas, technology and money that comprises the global search engine industry. The situation in South Africa was less connected to these flows: to supplement global services not tailored to the market, local providers used less technology-intensive services to provide access to local content, and these were not strongly linked to the global search companies.

The legal and regulatory contexts in which the search engine companies operated, moreover, affected the degree and kind of their participation in the industry. In the US, for example, few constraints exist since legal decisions had given broad protection to their operations. In China, by contrast, operations were restrained by censorship or, for US engines, by requiring physical presence in China, thus encouraging compliance with local law and resisting the "virtuality" of the organisation. Similarly, threats of regulation and local laws constrained the free operation of search engines in Germany, while access to government support for development was made available in Germany and Japan. There are, however, no guarantees that these legal positions are fixed, as proposals for search engine "policy" begin to be debated.

The analysis in this chapter makes visible two aspects of the spatialisation process at work in the search engine industry. First, it shows the rapid spread of US-based search companies, both in terms of physical offices, development activities, virtual services offered and economic activity. From this evidence, the search industry can appear to be a truly global business which, nevertheless, incorporates counter-flows from local activity, for example, German self-regulation or Chinese censorship, while still maintaining a centre on the west coast of the United States. Second, this apparently global activity still marginalises much of the globe, so that in smaller or poorer nations a supplementary service, such as that offered by South Africa's Ananzi, may be required to access local content which has less economic value, while a majority of traffic and business are still funnelled towards websites based in countries with better search connections. In this way the search industry dynamics can be said to contribute to an uneven access to internet content that bears a similarity to centre-periphery structure that have been documented in other contexts of media development.

VI

Is Relevance Relevant?

Discourses of Search Engine Quality

6.1 Introduction

The previous chapters provided an analysis of on the development and spread of the search engine industry. This chapter is the first to focus on the level of agents and their practices outlined in the research design (see Chapter 3, Section 3.3). It addresses the research sub-questions detailed there: "How can we understand search engine results creation as a practice, in which members of different communities participate, and from which the search result emerge as a reified object? Where does bias stand in relation to that practice?" The empirical questions addressed in this chapter are:

- How do search engine producers conceive of search engine results?

- How do they make decisions about where to allocate resources and how to make changes to the search engines?

- What notions of quality exist, and with what consequences for the
  search results?

This chapter goes inside the search engine, metaphorically, applying discourse analysis to the interview data elicited from senior search engine producers (see Chapter 3, Sections 3.4.1 and 3.4.2) to investigate how search engine quality is understood. Quality is an important normative issue strongly linked with bias, but despite its importance it is difficult to study empirically. This chapter highlights accepted and contested views of search engine quality within the community of practice in which search engines are produced.

The discourse of search engine producers is examined in this chapter in order to provide insight into how they articulate their assumptions about quality. A key concept for this chapter is the idea of the *technological schema*, which I define as a discursive formation through which technology is given meaning. The term is chosen for its resonance both with the "technological frame" construct employed by Pinch and Bijker and the "technological frame of reference" employed by Orlikowski and Gash (both discussed in Chapter 2, Section 2.3) and the "interpretative scheme" used by Giddens, (discussed in Chapter 2, Section 2.3.1). These technological schemas were discovered through discourse analysis (see Chapter 3, Section 3.4.2). These discourses should be seen to be a form of practice, a practice through which producers in participate in their community and from which in part the search engine results emerge as reified forms (see the discussion of communities of practice in Chapter 2, Section 2.3.1). The schemas in evidence, based on my analysis of interview data from search engine producers, primarily high-level engineering staff (see Table 2 in Chapter 3)[53] are introduced in this chapter (Sections 6.2-6.4). Each of the schemas identified ascribes meaning to search engine technology in a different way, using different definitions of quality. In Section 6.5, I examine how the different definitions of quality in the technological schemas appear to be used strategically by the search engine producers to control the development of search engine technology. Finally, in Section 6.6 I consider how these technological schemas

---

[53] The analysis for this chapter was conducted only on the interviews with search engine producers, not optimizers, distributors or commentators. This included eleven interviews: Mr A, Mr B, Mr C, Mr D, Mr E, Mr G, Mr H, Mr I, Mr L, Mr O and Mr Q.

may constrain both the possible interpretations of quality and the mobilization of resources around alternate definitions.

The results of the discourse analysis suggest that there are two major schemas in evidence that appear to influence the development of search engine technology. The first I have chosen to call the *market* schema, because the discourse within this schema refers mainly to business-related issues: costs, revenues, and competition. The second major schema, I call the *science and technology* schema, where the discourse is dominated by reference to experiments, measures, proof, and utility. There were also minor schemas, including *war* schema where the discourse includes references to enemies and combat. Though the schemas are analytically distinct, in practice they were not mutually exclusive. One of the striking elements of the analysis of the interview data is the ways in which interviewees negotiate between the discursive schemas. Each of the schemas is described in turn in the following sections.

6.2 The market schema

During the interviews, each interviewee was asked to describe a time when there was a modification to the search engine he was working on, what the rationale for that modification was, and who was involved in the modification. This was intended specifically to bring out typical accounts and rationales for change and, implicitly, quality. Many interviewees described both general processes and specific incidents. The most common descriptions and justifications were with reference to business issues including competition, revenues and costs. The interviewees used language that referred to the search engine as a commercial service in a marketplace that is highly competitive. In the market schema, decisions to alter the functionality or the display of the search engine were related generally in the interviewees's discourse to revenues, costs, or competitive goals.

The market schema links revenue to high quality through the discourse of "customer satisfaction." This makes sense when considering that companies exist to create profit and wealth, so an increase in wealth is easily understood as a quality business outcome. The belief that more customers lead to more revenue was unquestioned, and measures of customer satisfaction were said to be based on the idea that satisfied customers will

recommend other satisfied customers, leading to increased revenue, whereas dissatisfied customers will both leave and tell their friends, leading to a revenue decrease.

As discussed above, discourse that invokes the market schema stresses the business rationale for changes. An example is given by the following quote from a senior engineer who relates the way in which changes to the search engine are discussed and developed before being implemented. In this quote, changes to the search engine are articulated within a discursive framework related to commercial motives and goals:

> *"Well, I mean basically if a change is suggested there needs to be some kind of motive for it… [gives examples of specific motives including leveraging assets, market distribution, and market share…] and these things all drive towards market share, which of course is the ultimate goal, which leads to revenue, etc." (Interview B)*

Despite the fact that the discourse of the market predominated in the interviews, the interviewees seemed generally hostile to the interference of other parts of the company in the search product and, in particular, to the demands of advertising, as the quotes below illustrate:

> *"Product managers come up with completely irrelevant types of features they want to see implemented. So, for instance, instead of focusing on core technology, they ask you to put in yet another link or yet another space for ads in the interface." (Interview C)*

> *"I had to sacrifice a portion of the homepage to promoting their stuff. Which was pathetic." (Interview A)*

> *"…it was clear to us that if we started to give too much weight to the advertisers in terms of our index, we would dilute the value of our product." (Interview D)*

"Irrelevant" and "pathetic" it may be, but advertising is central to the search business, as Chapters 4 and 5 illustrated, and a reduction in advertising implies an immediate *reduction* in revenue, whatever the increase in customer satisfaction. Nevertheless, in the years leading up to when the research interviews took place in 2005, many companies decided to move away from search results which returned many results from the parent company or that included hidden advertisements. Some of the interviewees were involved in arguing for these changes and mentioned their role in them with pride. Here is how they argued for, and secured, the reduction in advertising, using a quality argument based on customer satisfaction as their justification.

First, one interviewee articulates the early strategy behind showing many results from the parent company:

> *"[D]on't send users, our customers away [from the search engine]…[instead] send them into our portal so we can help to monetize them again and all that good stuff." (Interview E)*

The business rationale for advertising is clear: keep the users in the portal and show them as many ads as possible. He then goes on to say why his company no longer follows that strategy:

> *"[W]e are moving away from that, as policy, because fundamentally, it doesn't work. As it turns out, if the portal has what you are looking for, you'll go there. You will, right? If it doesn't, then sending you there just pisses you off and you look like a shill, because you are." (Interview E)*

By "monetizing" the customers in accordance with business demands – that is, by sending them to other products so they can look at more ads – he argues that the search engine will "piss you off" and damage its own reputation by "looking like a shill" or conman. The perceived risk to customer satisfaction is clear. Another interviewee takes up the tale:

> *"It was a tough decision to make, because it meant a big revenue impact in the short term. But taking the long-term view, we knew that if we didn't do that, we'd probably have dissatisfied customers who would not want to use our service or not recommend it to friends, or maybe even switch to another service. So we were sort of taking the long-term view." (Interview F)*

The short-term revenue impact is justified by arguing for long-term revenue ("sort of" as he says). A third interviewee continues in the same vein, by saying that immediate monetization – "controlling where people go" – can be dangerous because "the users will stop":

> *"You have to be subtle in controlling where people go. You can't just only show them your own content. You can't hit them over the head. But you can certainly influence…It's tough, right? [There are] editorial concerns as to where you drive people. But it can obviously only be done without affecting perceived quality. If the user doesn't think they are getting the results they want, that won't fly. So you can't stick inferior products on the top of better ones, just because they are your products. The users will stop. They will object." (Interview G)*

From these quotes we can see that within this framewok, the discourse relating to quality seems to be linked to long-term satisfaction on the part of customers and that, in turn, is

linked to revenue, the positive goal or norm of the market schema. The next important observation is that a discursive anlysis suggests that it is the engineering production teams who define search quality (in terms of "relevance"), as is discussed in the next section.

6.3 The science-technology schema

The second major schema identified as a result of the analysis of the discursive strategies of the interviewees, the science and technology schema, is characterized by discourse that includes experimentation, measurement and proof (the more scientific aspect) and also usefulness, feasibility, and design qualities like "state-of-the-art" (the more technological aspect). The science constructed by this discursive schema is a positivist, experimental science that has objectivity as an essential norm. Technology is seen as the application of this science to the problems of search and is focused on solutions and progress. Thus, in the discourse of the science and technology schema, the search engine is both interesting in itself as a research object and also as a potential solution to people's needs.

The science-technology schema is exemplified by quotes like the one below, where the interviewee, who is Chief Scientist for a major search engine, described the procedure for making changes in their search engine algorithm:

> *"Yeah, well we're constantly making changes. The key thing to understand is that search and indeed basically all internet business is highly data-driven. One of the key components of what we do here is to develop a deep array of metrics with which we measure what is going on in the service. These are quality metrics. So a lot of the decision-making is really focused around observing deficiencies in some particular metric relative to where we'd like to be or relative to the competition, considering changes that would improve it, and often provedly improve it, because you can do a test and see what is the impact on the number of hits, and then know what would happen. So a lot of the work that I do is meant to be driven pretty objectively, and we tend to do that for most of what we do."* (Interview B).

Here, the interviewee gives as his reason for making changes some elements of positive science: measures, observation, proof and objectivity. As indicated, this is the more scientific discourse of the schema. It should be noted that the interviewee also speaks of measuring "relative to the competition." Later, I will examine the relationship between the science-technology schema and the market schema.

The next quote exemplifies a more technological discourse. Here the interviewee, the founder of a search engine that was successful in the late 1990s, is talking about how he came to develop the technology for that search engine:

> *"The common thread is big problems…I was working on all sorts of interesting things. One of the things I did from the very start was to try to make [Engine 1] useful to everybody, so that was a big effort…Making sure that everybody can access it …the analogy I was using at the time was – think of it as a pencil. You don't want your pencil to be some big complicated contraption that starts singing at you every time you pick it up." (Interview A)*

The emphasis on "big problems" and things that are interesting for their own sake is characteristic of the discourse of scientific research; but when the interviewee also goes on to talk about things being useful, accessible, practical, etc., the discourse becomes more applied or technological than scientific. It is worth pointing out that the way in which science and technology are constructed within this discourse is quite specific. The strong impression is that science deals with measurable (if complex) 'facts' that are causally linked and that the goal of technology is to use the knowledge of the causal links to enable the user of the technology to act on the world effectively and efficiently. Both science and technology are, or should be, progressing. Within this discursive frame, if x, then y. If only we can do $x+$, we should achieve $y+$.

As discussed above with reference to the market schema, many interviewees using that schema's discourse equate search engine quality with "customer satisfaction." However, there is a second major way to discuss quality that relates to the science and technology schema. This is the concept of *relevance*. As interviewee B says above, changes to the search engine in this schema are "meant to be driven pretty objectively." The implication is that things are not always objective and the term relevance, explored below, is a form of discourse that seems to encapsulate this contradiction.

The term "relevance" is drawn into the discourse of search engine producers from information science, where it forms the bedrock of several traditional measures of information retrieval quality, including, for example, recall and precision. *Recall* in information retrieval refers to the proportion of relevant documents retrieved from the database. *Precision* measures the proportion of retrieved documents that are relevant (Singhal, 2001). These terms were developed for relatively small, relatively high-quality

databases of documents – for example, news articles contained in Lexis-Nexis (see Chapter 1, Section 1.3). In those cases, one does not want to miss any relevant documents (that is, to have high recall) nor to retrieve very many irrelevant documents (that is, to have high precision). Metrics such as precision and recall are still part of quality-testing search engines, but what is remains ambiguous within this discourse is the ability to categorise documents into relevant and irrelevant.

Discursively, relevance takes central stage when the interviewees are talking about changes related to technical quality, as the senior vice president for technology of a major search engine told me when I asked what motivated changes in his search engine:

> *"Relevance. So relevance, freshness – I mean you can almost lump everything under relevance, but that's such a big umbrella." (Interview H)*

So, what is relevance? Again, in a small, well-defined database, it is relatively easy to sort 'relevant' from 'irrelevant' documents. On the Web, this is not necessarily as simple. One interviewee said that the standards or criteria for assessing relevance have changed from when he began to work with information retrieval systems:

> *"[W]here the systems used to only be the Dialogs and the Lexis-Nexises, you know, I think they strove for a more academic standard of relevance, where you define relevance as the relationship between the subject that is in the document with what the user is asking about. So it is sort of topical relevance. Whereas in the practical world where the search engines are reaching today, something being useful to the user and something where the user grabs the information and continues, has become, I think, more important and there is less emphasis on say, getting the best document." (Interview G)*

In other words, as he says elsewhere, it is about "satisfying users." Thus the discourse used by this interviewee suggests that the concept of relevance has changed from some type of topical relevance based on an applied classification something more subjective. Most of my interviewees defined a 'relevant document' as a document that answered the user's question or was what they wanted:

> *"Really, it is the standard definition, which is, we are trying to answer people's questions. Period. Relevance is when we actually return something that answers their question." (Interview E)*

From a technical standpoint, then, the definition of a quality search engine is simple: if the search engine gives you results that answer your question, then it is relevant and the results are high-quality, desirable results.

6.4 The war schema

The war schema refers to discourse that was characterized by terms referring to fighting, guarding, war, defenses and the like. This schema appeared to be a relatively minor schema – the words or phrases which characterize it occured in sections of the interviews that were related primarily to the market schema or the science-technology schema. For example, an interviewee would talk about an "arms race" between spammers and search engine producers, refer to competition as a "tough battle," etc. Nonetheless, military words and expressions occurred with enough frequency and sufficient clarity to warrant their inclusion as a separate schema. For example, one interviewee said, characterizing his time developing search engines: "I fought in the search wars." (Interview A).

The war schema offers little insight into the search engine 'quality' discourse. In contrast to the other schemas, the war schema was very focused on others, particularly, the enemy (and, conversely, on the identity of the speaker). In this context, the enemy discussed as twofold. First, the competition (other companies) was characterized as the enemy:

> *"[W]e're not trying to beat [Engine 2] and [Engine 3] at their game directly, I think that's a very tough battle, they've got lots of bright people, very well-paid, working on this stuff. And to try to go head-to-head, say on search quality, is a very difficult thing. (Interview I)*

This is commonplace usage – "battling" other companies, going "head-to-head," etc. The animosity visible in to the war schema seemed to be quite impersonal when the discourse is referring to respected opponents.

However, discourse referring towards the second class of enemy, the guerilla fighters of spamming and hacking, seemed to contain more direct animosity. Here an interviewee describes them as trying to "get at you:"

> *"There is also an adversarial aspect to it in that you have hackers and spammers trying to get at you." (Interview F)*

Sometimes the enemy can be trying to "get at you" by threatening your revenue, and sometimes by threatening your technology.

The talk within this schema is about 'beating' the opponent or enemy. In other words, decision-making seemed to be characterized not by any kind of appeal to hierarchy, consensus, or objective measure of quality but rather to who can "win" – even though several interviewees likened it to an "arms race" in which no one was likely to come out on top. This particular metaphor, the "arms race," was not used to refer to competing with other companies. Spammers were also likened to criminals, particularly fraudsters or conmen, and specifically contrasted to "honest" people.

The war schema appeared not only to be important in defining relationships with other communities, it also provided an important reflection on the identity of the producers, as their discourse suggests that they assume the role of the guardian or protector of something precious, in this case, access to the Web, as Interviewee D does explicitly in this quote:

> *"We considered search to be important, we considered it to be a service that people needed and wanted and it was up to us guardians to make sure that we gave them the best experience possible." (Interview D)*

The discourse of the war schema is important because it frames much of the discussion about people *outside* the search engine organization. In this schema, the guardians of search appear to defend against the incursions of the other, whether those others be honoured competitors or fraudulent spammers.

In summary, therefore, the search engine producers were found employ discourse that suggested two primary technological schemas used to ascribe meaning to search technology. In the market schema, the search technology appears as part of a business; in the technology schema it appears as a piece of engineering work. Each schema seems to have a concomitant definition of quality: either as "customer satisfaction" or as "relevance." The minor war schema characterizes search technology as a defence, either against the competition (market schema) or against those who would affect the results for their own ends (technology schema). Quality issues were not specifically referred to, but

one can draw the tentative inference from the analysis that secrecy (from the competition) or robustness (against spammers) might be indications of high quality search.

The discourses represented by these technological schemas seem to be not only a method of accounting for and explaining technology, but also to have consequences for search development as they function as a device to mobilize other resources, e.g., cash, office space, extra personnel, etc. Which schema to use, at which point, to mobilize what resource, is therefore a strategic question for the actors, with many implications. Having discussed these discursive schemas and some of their implications for definitions of quality, the next section examines these implications in greater detail.

6.5 The strategic use of technological schemas

This section analyses the strategic aspects of the schemas evident in the interviews in line with the methodology outlined in Chapter 3, section 3.4.2. First, it investigates the way in which, discursively, producers construct their own identity and agency as they talk about their work with search engines. This has important implications since each schema seems to construct the speaker in a distinctive way with specific abilities. Second, it focuses on the recursive relationship between 'relevance' and 'customer satisfaction', and the construction of these two terms in such a way as to potentially empower the producers. Finally, it examines how these major schemas appear to constrain the expression of alternative quality schemas.

6.5.1 IDENTITY AND AGENCY

As interviewees discussed their work and accounted for their actions, they used language which either implicitly or explicitly appeared to reflect their senses of identity and agency, that is, the ability to act. When framing actions within the market schema interviewees often discursively constructed themselves as significantly constrained, in marked contrast to the rather more empowered constructions of the science-technology schema.

The grammatical structure of the interview texts suggests that interviewees overwhelmingly referred to actions and descriptions as part of a collective corporate "we," typically using the pronoun "I" only when discussing personal matters or when they were unsure of the agreed corporate version.

Here, for example, is one interviewee discussing what led to the decision for his company to build its own search technology instead of purchasing listings from a third party as had been their previous strategy. In the following quote, where an interviewee accounted for a change to the search engine that began before he joined the company, note the use of "I" where the interviewee is uncertain, in contrast to the corporate "we" as he returned to more familiar ground:

> *"I don't have a lot of background on that, but I would imagine, personally, just observing the explosion of information online, partially comments from our own customers and business partners, partially an observation of customer dissatisfaction with our search experience. We measure that on a very regular basis, and we care a lot about what the end user tells us, and we knew that customers weren't happy with various aspects of the service. (Interview F)*

The interviewee also described the way in which the decision was finally taken, which involved the whole management chain – he later clarified that it went all the way to the CEO of the company. The interviewee went on to say that "while I had responsibility for designing a service that we could operate to bring in revenue, I don't have final say on a decision that's going to have significant revenue impact." (Interview F) His perceived sphere of action – the changes he could and could not make to the search engine – were positioned as relative not to his technical competence nor his ability as a leader but, rather, relative to business factors. In this case, he accounted for his own agency through the market schema: changes that have no or little revenue impact were within his sphere, those with significant impact were outside it.

While none of the search engine producers identified himself explicitly as a "Microsoft man" or as a "Google man," there were explicit professional identifications as engineers and researchers in the interview texts. One early search engine developer talked about why he began to work on search engines in the mid 1990s:

> *[T]here was a need, the need wasn't being met, our collective internet experience was less as a result of it, and we wanted to fix that! You know, we're engineers, we fix things. (Interview D)*

This is a quote in which the interviewee identified specifically as an engineer. He very emotionally expressed why he belonged – because engineers fix things so that our collective experience can be greater.

It is notable in this analysis based on both notes taken during the interviews and the interview texts that, while the market schema was more the pervasive one, the interviewees were most animated and excited when expressing themselves using the language of the science-technology schema. Their voices rose with excitement, they spoke more quickly, they engaged more with the interviewer and almost on the role of an educator. In short, they seemed to be more comfortable with this way of expressing themselves and seemed to identify more with this schema. In the discourse of the science and technology schema they spoke as "experts", fully comfortable with their agency or ability to act, in contrast to the market schema where even the very senior personnel expressed the limits of their actions.

6.5.2 QUALITY CONTROL

One implication of this discussion of identity and agency is that within the dominant market schema interviewees construct their ability to act as being significantly constrained, as described above, whereas within the science and technology schema, they perceive themselves as being rhetorically empowered. Strategically, then, it is of benefit to these producers to be able to use the science and technology quality construction of "relevance." They do this, first, by constructing a rhetorical "customer" who does not correspond to the actual customer. This rhetorical customer is satisfied by greater relevance, unlike the actual customer. Second, the slippery subjective concept of relevance is quantified, reified, and discussed in highly technical language that makes it inacessible to other actors within the search engine company, for example, advertising salespeople.

A customer ordinarily means someone who buys products from a company. In that case, greater customer satisfaction would, in most cases, lead to more frequent purchases and have a positive impact on the revenue of the company. In the discourse of search engine producers as analysed here, however, "customers" are equated with users. Users – the people who type in queries and click on results – are not customers of search engines in an ordinary sense, because they do not purchase products from search engine companies. In fact, the customers of search engines (in the ordinary sense) are the hundreds of thousands of businesses that purchase advertising and other services. The rhetorical customer/user of the search engineer appears to serve the function of creating "customer

satisfaction" though greater relevance and thus rendering relevance a key benchmark of quality in the market schema as well as the science and technology schema. None of the interviewees mentioned changing the search engine to make it friendlier to advertisers and, indeed, many were openly hostile to advertising.

Relevance, therefore, appears to be the linchpin of producer discursive strategies towards search engine quality. Recall that in the most basic terms, the relevant search engine result provides answers to the users' questions. But providing this is not a simple operation, since what the user wants is understood as a subjective thing, as the following interviewee pointed out:

> *"..[I]t is completely subjective based on the customer's frame of mind. So we are of course trying to develop models so we can figure out what that subjectivity is and therefore get our customers the best thing that we can do. But it is, you know, all…completely whatever they want! And it changes – after they look at the first result, it can completely change. It's a little bit of Heisenberg. Really that is what it is." (Interview E)*

The user/customer conflation is clear in this quote. The interviewee used the language of the science and technology schema when he discussed "developing models" to "figure out what that subjectivity is." What the interviewees seem to be engaged in as they "develop models" is a process of making an objective, causal and factual experiment out of the uncertain "Heisenberg" process of answering an often not-very-specific question on the part of the user. For example, what is the correct result for the query "new york apple" (a type of apple or a reference to the city)? How about "Napa" (both an auto-parts chain and a wine-growing region in California)? How about "abortion" (medical advice, addresses of clinics or political issue)? The results generated from this process must be seen to be objective and, indeed, Google specifically defends its results from accusations of inappropriateness on this basis: "Our search results are generated completely objectively and are independent of the beliefs and preferences of those who work at Google" (Google, 2004a).

In addition to being objective, the science and technology discourse suggests that results must be replicable and subject to improvement. This process of reifying an intensely subjective choice is a difficult moment for producers (see Chapter 2, Section 2.3.1 on reification as a process). The Chief Scientist of another large search engine was working

on creating a new type of search algorithm for specialist queries at the time of the interview. He indicated that one of the most important parts of this work was creating a measurable, improvable model out of subjective preferences: "[I]n the case where something is brand new, we work very hard to understand how we would go about measuring it." (Interview B).

However, once the model is completed and the subjective is perceived to have become objective, most of the day-to-day work begins. According to this interviewee, about 80% of the work they do is incremental improvements to existing technology, also called "tuning" the search. Interviewee E is particularly enlightening on this topic as he is the Program Manager for search relevance at a major search engine. He went on to describe in highly technical language, using specialized terms, how quality as relevance is constructed:

> *"We have to use things like precision, recall at* n *or* r *is like 15. We have to use all of that good stuff. That's a secondary…Primarily what we are trying to do is, we are trying to figure out what is a model. Once we have a model, that is when you start to use things like precision at* n. *You have a belief that these documents are good for a given query. Now I can actually crunch some numbers and improve on that." (Interview E)*

This mathematical language, with the use of *n* and *r* (measures of precision in information retrieval) echoes that of Interviewee B quoted earlier discussing how he can "provedly improve" search quality "because you can do a test and see what is the impact." Thus, the account of the day-to-day process of "tuning" the search engine running appears to be one where the discourse of the science and technology schema comes to the fore:

> *"We assume that we have got…we call them relevance judgments, right, or test sets, or whatever you want to call them…So, you have been given this test set. Tuning is simply a matter of optimizing the test set. Effectively it's a classification problem. You know, dividing documents from your entire corpus into good and bad." (Interview E)*

Again, the tuning process is referred to via the specialized language of information science, discussing test sets, optimization, classifications and the document corpus, which are known, identified elements of information retrieval technologies. Thus, the analysis of this discourse suggests that, although the model for relevance is established by interviewing groups of users to see "what really is good and bad" (Interview E), the day-

to-day management is really "pretty straightforward and relatively boring scientific number crunching" (Interview E). This "objective" relevance judgment is then re-incorporated into the world of the market by being used as a competitive measure to judge the quality of other search engines, as the quote from Interview B earlier in this section indicated. Here again, relevance is a measure, albeit a discursive one, which works well within the organization, but encounters some strain from its subjective roots:

> *"We have people whose job is based around relevance. And that is both relevance just looking at the result, it's also relevance looking at the competition's result and so on. Because just because it's different does not mean it's not relevant. Because sometimes a query, sometimes it's subjective too - if you don't know enough from a query, you don't know what a user's looking for." (Interview H)*

In this quote, relevance appears to be used as a competitive yardstick even while its validity is being questioned. This was the case in more than one interview. One interviewee who formerly worked in a major media conglomerate as head of search was still shaking his head in bemusement years later over why people's relevance judgments were not exactly in line with their assessment of search engine quality:

> *"We did some studies internally when we were doing some stuff at [Engine 4], comparing how [Engine 4's] results were compared to [Engine 3's] and I remember at one point that this experiment showed that our results were actually better but were being rated by the users as not as good." (Interview G)*

The difficulty here is squaring results that are "actually" better – from the objective point of view of the experiment – with the subjective ratings of the users. It is vital for the producers that the reified and quantified "relevance" not be moved back towards subjectivity, as the following interviewee articulated:

> *"There are two features as far as relevance, right? One is, given a query, produce the most relevant things. That task in and of itself is – to find that level, it becomes purely technical. The business side says "give us the most relevant thing" that is all they care about. The related task, and this is where you start to see other aspects of the company come into play is when you are defining the language, for example you want to allow the customer to say what relevance means and give them some options. That is where various – the business and the marketing side will start to say 'ooh, it would be really great to have* xyz *feature.'" (Interview E)*

As long as relevance is the discursively agreed "purely technical" measure, the "business side" just wants it to be better. As soon as the definition is put up for grabs, or you

"allow the customer to say what relevance means," then sales and marketing begin to intervene in the development of the search product. Thus the elision of relevance, quality, and customer satisfaction in the discourse of these interviewees as elucidated above and in the section on quality in the market schema appears to be necessary for the producers to retain a perception of influence over the search engine code.

6.6 The difficulty of articulating the public good

The quality metrics of customer satisfaction and, particularly, relevance appear to serve discursively the strategic ends of producers, helping them to overcome the limits on their agency which the market schema implies. Nonetheless, the focus on relevance seems to constrain the articulation of other quality goals. For example, in journalism, objectivity, fairness, diversity and representation are typical examples of quality goals. In the course of this research, interviewees mentioned many everyday practices in search engine programming that could be considered censorship of search results and have the potential to lead to biases in search. These included blacklisting, or the exclusion of certain sites or site owners; whitelisting, or the automatic inclusion of certain sites or site owners; weighting content according to whether sources were considered to be authoritative or not; and adjusting results based on pressure from executives to respond, for example, to current news events. None of these practices were considered problematic because all were linked to obtaining greater relevance in search engine results.

Not all of my interviewees were entirely happy about this state of affairs. In two interviews (D and I) there were hints of another minor schema – what we might call a "public service" schema, emphasizing equal access and fairness. But even Interviewee I who said that the goal of his search engine is to be "more transparent," had a difficult time discussing an alternative quality criterion such as bias. He introduced the topic into the interview, but also rejected it in the same sentence, which suggests that bias is conflicting with other ideas about search engines:

> *"I don't think that major search engines today are horribly biased. … but they're also not objective, and the more you operate secretively, it makes it harder for people to see where you're subjective. Maybe subjective is a better word than biased, although it pretty much means the same thing." (Interview I)*

Here the word bias seems to be dissonant with the discursive "objectivity" contained within science-technology schema's discussion of the day-to-day operation of search engines.

Another interviewee was asked what he thought about people who said that search results were becoming too commercialized. He refused to believe it was an issue:

> *"This is not an example of the commercialization of search, but of the commercialization of documents available on the Web…I mean if there are only commercial documents on the Web, then even a noncommercial search engine will come up with a list of commercial documents." (Interview C)*

However, if it were an issue, he goes on to suggest that it is a technical, infrastructural issue:

> *"I see the bigger problem being centralization. The crucial part of your infrastructure is centralized, that is, you introduce a single point of failure." (Interview C)*

The idea of a technical solution was echoed by another interviewee, whose utopian vision of the solution to a potential problem of bias or the over-commercialization of search included not only 'perfect' technology but also education and literacy, distributed throughout the world:

> *"What we need to do is to create a world of educated, thinking people. Then the ads will have no effect, and we don't need to be concerned with what the corporations do. You need to look at it in a longer timespan – advertising is just a feature of a particular stage in evolution. Technology will get to a point where there is a big space, and so you need to make the profit on the difference and the quality, and yes there will be popular things. But the technology will distribute everything to everyone who needs it, and everyone will be able to find what they like. That's much better than regulation." (Interview J)*

More common was be the view of a third engineer who simply suggested that if search engines were censoring their results, they were doing it for the good of the population:

> *"When you get into this logic, unfortunately that's the tragedy of the commons, right, you have this free resource, people with this mindset will actually go and use it all, and destroy the value, so it's the same for the Web. I think search engines are totally justified to not be kind on those spammers. Sorry, I'm getting a little excited here". (Interview A)*

The "guardian" of the war schema seems to defend the Web in this quote. It would be reasonable to suggest that quality schemas including ideas of full disclosure, representativeness, or diversity operate at a tangent to the way in which producers primarily frame their work in their discourse.

6.7 Conclusions

The evidence from the discourse contained in the interviews analysed for in this chapter suggests that search engine producers appear to conceive of quality in two separate, but interrelated, ways. First, a high quality search engine, from the producer's perspective, has high *customer satisfaction*. This definition of search quality is embedded in a larger technological schema that I have called the "market" schema, in which search engines are primarily conceived as businesses. Second, a high quality search engine produces very *relevant* responses to queries. Again, this definition of quality is related to the technological schema of that I have characterized as "science and technology". Search engines are characterized within the discourse of the science and technology schema as, primarily, engineering artefacts.

The implications of these conceptions of quality are far-reaching precisely because they are embedded in larger cultural schemas. The discussion in Chapters 2 and 3 emphasized how interpretative schemas and their associated norms may guide the allocation of resources. In this chapter the related concept of the "technological schema" has been used to focus on the analysis of the discourse of search engine producers. The foregoing analysis suggests that in the case of search engines, several schemas are at work simultaneously. The schemas in the ascendant – the dominant market schema and the science and technology schema – encompass little if any of the discourse that might indicate that this community is concerned with issues of public interest, fairness, or bias. Instead, the discourses emphasize profit, in the case of the market schema, progress and efficiency, in the case of the science-technology schema, and defence, in the case of the minor war schema. The analysis of the interview texts suggests that the minor public service schema offers a potential alternative to the market and science and technology schemas. However, the low incidence with which occurs, and the difficulty its exponents have in articulating it, as well as the opposition of others to hearing its articulation,

indicates how problematic it might be to build up an alternative discourse within the context which search engines producers foster their community of practice.

VII

Ranking Highly

Advertising, Optimisation and Spam

7.1 Introduction

In Chapter 6 the focus was on the technological schemas of the search engine producers, examining how they articulated and gave meaning to search results. This chapter moves away from the search engine producers to investigate another important group or community of search engine influencers: those professionals who call themselves "search engine marketers" (SEMs) or "search engine optimisers" (SEOs). SEMs and SEOs attempt to manipulate the search engine rankings, either on their own behalf or on behalf of clients, and this has led to some being labelled as "fraudsters" or "spammer". As this chapter will suggest, the relation of SEMs to the search engine companies is much more subtle and complex than these labels initially suggest.

A technology-in-practice, as discussed in Chapter 2, Section 2.3.1, includes both social and technological forms. This chapter examines the technological forms of search advertising and then its social forms based on a thematic analysis of interviews with SEMs and the results of my participant observation conducted at a well-established London-based search engine marketing firm, as detailed in the methodology discussion in Chapter 3, Section 3.4.

In Chapter 4, Section 4.5 the importance of syndicated advertising overall to the business model of search engine companies was emphasised. This chapter examines the advertising forms specifically and, in particular, the forms of *non-advertising* promotion that companies use to increase their ranking in the "natural" search results; that is, those that exist outside the areas labelled "sponsored" content. This area is important because the actions of paid professionals in manipulating search results are, in part, a contributor to the bias of natural search towards commercial results that is evident in the studies discussed in Chapter 1, Section 1.2. This chapter is concerned with the research question: "How can we understand search engine results creation as a practices, in which members of different communities participate and from which the search engine results emerge as a reified object?" (See Chapter 2, Section 2.6). The empirical questions examined are set out in Table 1 in Chapter 3:

- How do search engine optimisers interact with search engine results?
- How do search engine results act as a boundary object between optimisers and producers? How else do these communities interact?

This chapter examines the relation between SEMs and search engine producers like Google, particularly with regard to the production of search engine results, the way in which SEMs were found to interact with search engine technology and with search engine personnel, and how SEMs seem to define 'quality' in their work. SEMs and search engine producers may be in direct, overt conflict, or they may be involved in a tacit negotiation, or some combination of the two. The elements and terms of their interaction, as well as its resolution, are of interest in this chapter.

7.1.1 SEARCH ENGINE MARKETERS AND SEARCH ENGINE ADVERTISING

Search engine marketers, broadly speaking, are people who actively use search engines to promote companies, products or services. According to the Search Engine Marketers Professional Organization (SEMPO), most search engine marketing is done in-house by employees promoting the companies for which they work (SEMPO, 2007). However, a substantial amount of marketing is done by specialist agencies or by ordinary ad agencies or digital agencies with specialist search marketing personnel.

Companies that market their products and services on search engines range from giants like AOL (which generated approximately 9% of Google's total income in 2005) (Google, 2006a) to tiny single-person businesses and blogs. In between these extremes, large to medium-sized companies are the most likely clients of SEM agencies. Most SEMs are themselves small companies – one estimate gives the average number of employees as 25 people (SEMPO, 2007) – and are local rather than multinational.

The structure of the chapter is as follows: Section 7.2 examines ways in which search engine producers and SEMs work together in formal arrangements, primarily in the paid search advertising arena; Section 7.3 analyses the area of search engine optimisation (SEO) in which SEMs work independently from search engine production companies to influence search results, without necessarily coming into conflict with them; and finally, Section 7.4 examines search engine spam which, as discussed in Chapter 6, seems to be framed by search engine producers as a highly conflictual "arms race" or war with SEMs.

7.2 Working together: Search engine advertising

This section examines the relationship between SEMs and search engine companies as they work together on search engine advertising. In the search advertising arena, the SEM is a client of the search engine, much as advertising agencies in any medium are the clients of publishers. Search agencies are an important source of revenue for search engine producers, since perhaps between 10 and 20% of advertisers' total search advertising budget is spent through agencies (SEMPO, 2006). Thus, in this capacity, the SEM is a client of the search engine company and the search engine companies provide preferential services to them. In this capacity the SEMs can also hold the search engine to account, for example, on the issue of click fraud as discussed below. Their market power is limited, however, with the advertising market being dominated by a few powerful suppliers (Google, Yahoo!, and MSN) and many relatively weak buyers.

While some advertising on search engine sites is similar to advertising in other mass media, particularly print media, a large and growing category of *search-specific* advertising is not. Essentially, there are two types of search-specific advertisements, which are summarised in Table 13: keyword advertising (also called paid performance advertising) and paid inclusion services. In 2004, a team of researchers reviewed eight search engines

using a matrix of queries of different subject and query lengths to understand the mix of advertising-based versus 'natural' results available and concluded that on "general Web search" engines (Google, Yahoo!, and MSN) 44% of the first screen and 18% of the first Web page were sponsored results – that is, advertising (Nicholson et al., 2006).

Table 13: Search-specific advertising formats

Format	Description	Pricing
Paid placement or "keyword"	An advertisement linked to a search term – for example, the 'sponsored result' for Expedia that might appear when you type 'travel'.	Run on an auction basis to determine cost-per-click.
Paid inclusion	A fee paid to the search engine to include the site in the search index.	Per item.

In contrast to most online media, search engines can offer advertisers more than a generic visitor identified by demographics. Because users of a search engine site are actively looking for certain information, search engines are able with confidence to sell specific ads to advertisers. In this respect, search-specific ads are reminiscent of ads placed in the Yellow Pages. By linking the search query term that the user types in with the advertisements that are displayed, search engine companies have found a way to help advertisers break through the 'clutter' of the mass market and target precisely who they are looking for. This type of advertising is known as *keyword* or *paid performance* advertising.

Paid-performance ads caused a huge controversy when they were introduced (Rogers, 2000), because they called sharply into question the assumption of search engine objectivity by including paid results in the same list as free results. Paid results are now shown on every major search engine results page, although they have become more visible and less controversial due to labelling by search engine providers – they are now usually, though not always, indicated by a term such as 'sponsored results'. See Figures 20 and 21 for examples of paid-performance ads from Google and AskJeeves.

Figure 20: Paid performance ads on Google
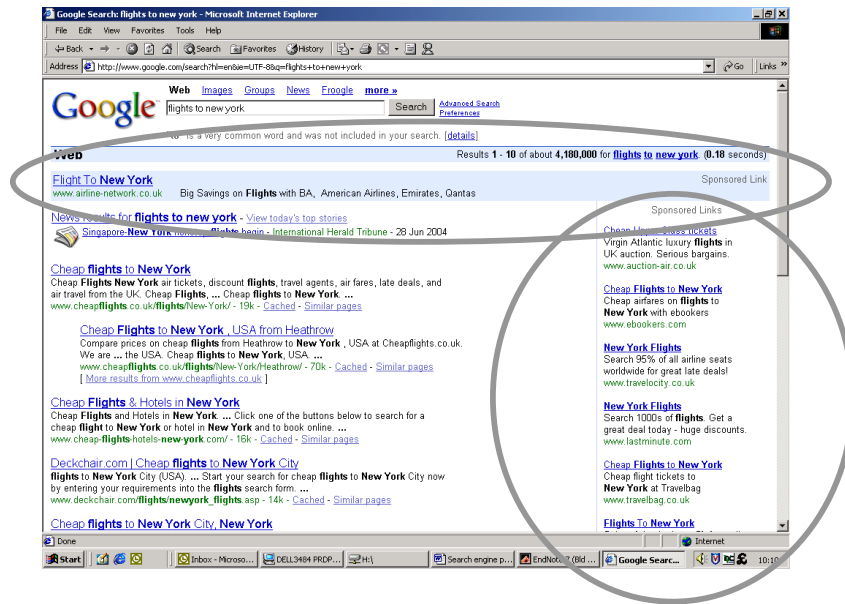
Paid performance ads (indicated by 'sponsored link') are included at the top and right hand of the page. From http://www.google.com, last accessed 1 July 2004.

Figure 21 : Paid performance ads from Ask Jeeves
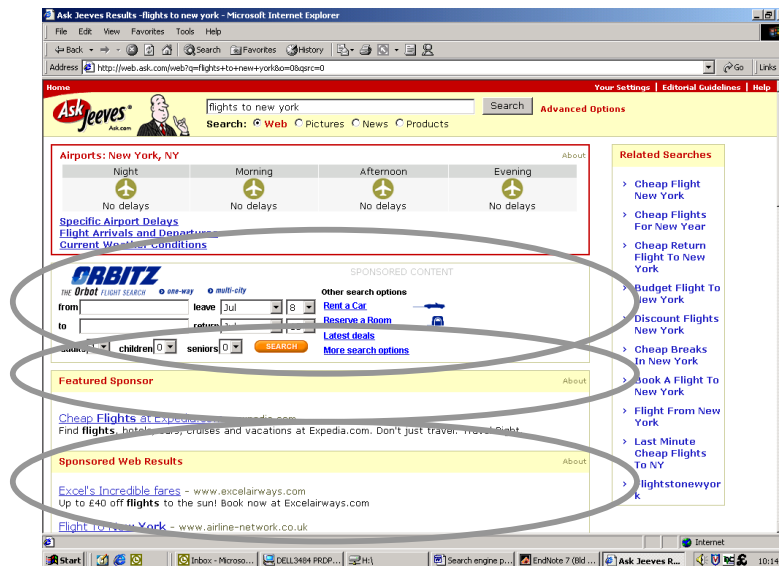
Paid performance ads indicated by 'sponsored content' on the Orbitz ad and 'featured sponsor' and 'sponsored results' on the two bottom sections. From www.ask.com, last accessed 1 July 2004.

The second type of search-specific advertising is *paid inclusion* advertising which is invisible to the end user. Paid inclusion advertising requires a short technical explanation. When search results are displayed on a page, they are generated by the search algorithm program. The search algorithm does not search the actual Web itself; instead, it searches a

previously collected database of sites called the index. The index is generated by another computer program, called a spider or crawler that automatically follows links on the Web and records details about the pages it finds. The Web is so huge that no spider can crawl its entirety. Thus, some sites and some pages may never be included in the index. Some search engines therefore offer a service that guarantees that the spider will crawl a site an advertiser submits and include it in the index. Since this index entry looks exactly like an index entry from a normal spider crawl, and since search results are generated from it in the same way, it is invisible to the customer. It is not clear from company reports how much revenue is generated by paid-inclusion advertising[54]. The screenshot from Yahoo!, Figure 22, illustrates how the company markets the whole of its search page to advertisers either via sponsored keyword search or via paid inclusion programs.

Figure 22: Promotional page for Yahoo! paid inclusion programme
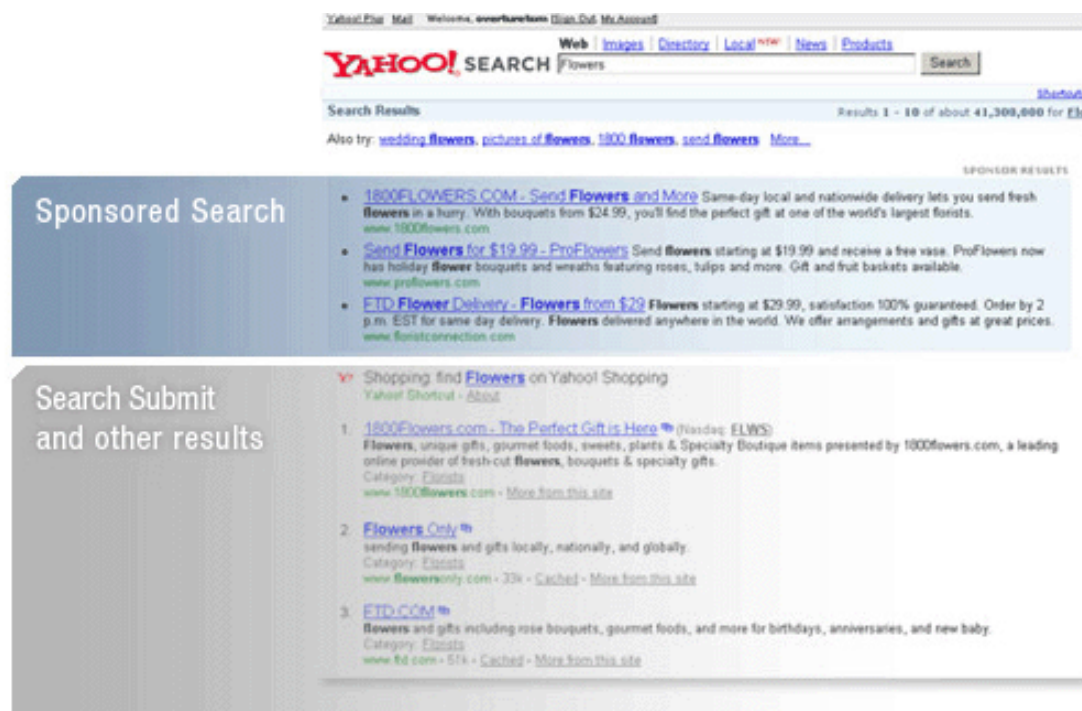
From http://searchmarketing.yahoo.com/srchsb/ssp.php, last accessed 31 October 2007

In North America alone in 2007 the amount of money spent marketing via search engines was estimated to be \$12.2 billion[55], of which approximately 87% was paid placement

---

[54] Google does not accept paid-inclusion ads as a matter of policy.

[55] All monetary amounts in this chapter are in US dollars unless indicated otherwise.

advertising and 1% was paid inclusion advertising (SEMPO, 2008). The balance was spent on search engine optimisation (11%) and technology investment (1%).

As large advertising clients of the search engine companies, SEMs receive some preferential treatment, mostly in the form of access to some technical tools, but also in the form of human contact with named account representatives at the search engines, invitations to various officially hosted events, and small gifts and marketing samples – for example, Google lamps or Yahoo search t-shirts. Nevertheless, despite friendly official relationships and appointed human representatives, the business of search engine advertising is a heavily automated one and even large advertisers are subject to automated advertising approval services and face frustrations when trying to assess how they have fallen foul of the search engines' multiple advertising restrictions[56].

During my period of observation with the London-based SEM firm (from September to October 2004) I repeatedly observed that customer service representatives were often not available and did not return calls or emails. Much of the communication between SEM and search engine provider was carried out via the technology provided by the search engines specifically for advertisers. Indeed, the advertising activity is on such a scale that this automated process may be the only practicable one: one SEM at the Search Engine Strategies conference in London in February 2005 reported managing 25,000 to 30,000 paid keywords every month for his clients; at the SEM I observed, the number per client per month was more typically 500-1000.

The tools and interfaces directly provided by the search engine companies are critical to managing these volumes of keywords. Available freely or at a low cost are all the tools

---

[56] Google, for example, had 57 separate named policies for text ads using its AdWords service, as follows (Google, 2007a): "Editorial & Format" policies for text advertisements include accurate ad text, capitalization, character limits, competitive claims, grammar & spelling, implied affiliation, inappropriate language, prices, discounts & free offers, proper names, punctuation and symbols, repetition, superlatives, target-specific keywords, trademarks, and finally unacceptable phrases. "Content" policies include those on academic aids, aids to pass drug tests, alcohol, 'anti' and violence, automated ad clicking, bulk marketing, copyright, counterfeit designer goods, data entry affiliates, dialers, drugs and drug paraphernalia, e-Gold, fake documents, fireworks and pyrotechnic devices, gambling, hacking and cracking, miracle cures, mobile subscription services, prescription drugs and related content, prostitution, scams/phishing for personal information, sexual and adult content, solicitation of funds, template sites for ad networks, tobacco and cigarettes, traffic devices, weapons, webmaster guideline violations; "Link" policies cover affiliate URLs, the back button, destination URLs, display URLs, landing page and site quality guidelines, pop-ups, site security, software principles and webmaster guidelines; and, finally, a series of miscellaneous policies covers double serving of ads, differing terms and conditions, advertising in China, invalid clicks and Google's privacy policy. Still other policies cover local business ads, video ads, audio ads, print ads, mobile ads and image ads.

necessary to bid on keywords, enter ads, and track performance of individual campaigns. Many of these tools target the individual website publisher and the small business or individual client. For example, Microsoft offers "AdCenter"; a keyword campaign management tool for advertisers of any size.

Some services and tools, however, are preferential services specialised for the agency handling multiple clients. Yahoo's "Search Submit Pro" service, for example, allows qualifying clients to feed individual Web pages in bulk into Yahoo's index. It is available to those who spend over $5,000 per month on search advertising or submit more than 1000 pages per month to the search engine. Google's services for SEMs are very robust. They were offering a full package through their "Google Advertising Professionals" programme, including professional accreditation, a training programme, marketing credits, promotional tools and a special multi-client interface (the My Client Centre or MCC) available for third parties who spend more than $100,000 in any 90-day period (Google, 2005). The MCC is shown below in Figure 22.

Figure 23: Google MCC (My Client Center) interface for SEMs handling multiple clients

From http://adwords.blogspot.com/2006/10/managing-multiple-client-accounts.html, last accessed 20 March 2008.

When considering keyword advertising then, the search engine companies can be seen to treat SEMs as valuable clients, extending privileged technology and information to them.

At the SEM office I was observing, there were also many small items – T-shirts, lamps, pens, mugs, flash drives, etc. – branded with Google and Yahoo company logos which had been given as gifts to loyal customers.

This official relationship, while clearly supported by the search engine companies as well as by the SEMs, is not always amicable. One particularly long-standing dispute concerns so-called "click fraud." My informants told me that often when they began bidding on competitive terms, they would receive calls or emails offering to click on competitive ads. This, in turn, at least wastes the competitors' budget on fraudulent clicks or, at most, exhausts their budget so that competitive ads are no longer shown. In 2006, Google and Yahoo settled a class-action suit in the United States out of court on this issue, where plaintiffs claimed that the search engines had failed to halt these fraudulent actions because they in part profited from them (Kopytoff, 2006; Sherman, 2006). The accusations continue with one research house reporting a click fraud rate of 16.6%, while Google claims it is 0.02% (Sherman, 2008).

7.3 Working apart: Search engine optimisation

In addition to paying for search advertisements, many SEMs also provide a search promotion service to their clients for which search engines receive no direct compensation, called search engine optimisation or SEO. Although search advertising is big business, not all clicks are for sale: the natural search results on most search engines are not part of advertising inventory (other than via the paid inclusion programmes discussed in the previous section). Given, first, that search engines are the online media properties with the highest traffic and, therefore, the most potential clicks, and second, that most searchers do not go beyond the first or second page of search results (Spink, Jansen, Wolfram, & Saracevic, 2002), there is an economic incentive for businesses, in particular, to attain a high place in the natural search rankings of a relevant search query. From this it follows that placement within the results rankings assumes importance financially. This is so whether the business is making money directly from commerce conducted on its website, or from advertisements or subscriptions.

Just as companies that seek good press coverage may turn to a public relations agency, the online company looking to increase its ranking on major search engines for relevant

queries may turn to a search engine marketer. The SEM's function is, in fact, very similar to that of a public relations firm in a traditional media context. In 2006, the Search Engine Marketing Professional Association (SEMPO) estimated the total amount spent on "organic" search engine optimisation (as opposed to paid advertising) to be $1.1 billion (SEMPO, 2007).

Stated another way, rather than purchasing on a pay-per-click basis an obvious advertisement linked to keyword search, the client purchases, via the SEO, a long-term placement in the main search listing which they believe to be more stable and more valuable in the eyes of users, as recent marketing research has shown that people are increasingly ignoring some online advertising altogether (Drèze & Husherr, 2003). The search engine optimisation process is more risky and often more expensive than keyword advertising, but if done well it may result in a much greater flow of traffic, especially in the long run. Indeed, the SEO business is predicated on the fact that search engine optimisation *does* work to increase traffic and, in the hands of a skilled practitioner, works very well. It is particularly cost-effective on terms that would command a high price if purchased through ordinary keyword advertising.

From the perspective of the SEM, the search engine results pages do not appear as a given. Rather, they seem to be the result of a series of practices, each susceptible to influence which leads to the production of a particular results page. They can be, and are, actively managed. How does search engine optimisation work? Here is what one of my interviewees, the head of a London-based SEM, had to say:

> *"It's not like…paid search, where you know it's either live or it's not, it's not a banner, it's not a* buy*…There is no contract and so therefore there is no guarantee, you can't pick up the phone and shout at Google…It's just the way that SEO is, because you're dealing with 3rd party technology [the search engines] which is unknown, because there is no actual open information about that technology…It's a bit like being a scientist, that's one way I look at it…SEO is some weird form of information science." (Interview K)*

The "science" of SEO, according to my informants, is based on observation and deduction. Undertaking an SEO project is a very active process, typically involving at least four areas: 1) changes to the content of the target website, including the title and text of pages; 2) changes to the information structure of the website, that is, the linkages

between pages; 3) changes to the technical structure of the website, particularly the domain names, directory structure, and filenames of key pages; and 4) the development of linkage networks from other websites. At each stage of the process, specialised software may be employed; once again, the search engines themselves provide some of it.

The first phase of an optimisation project is to establish the language searchers are using in seeking sites like yours. Google and Yahoo both maintain keyword suggestion tools (see Figure 24). These tools are intended to assist advertisers in purchasing keyword-linked advertisements but also work for SEMs that use them to establish the language of searchers. In Figure 24, for example, the phrase "communications PhD" entered into the Google search tool reveals that many searchers search instead for "communications courses" or "communications studies." As a result of this, the SEM would likely advise their university clients to include these terms on their communications department pages.

Figure 24: Google Keyword Suggestion Tool
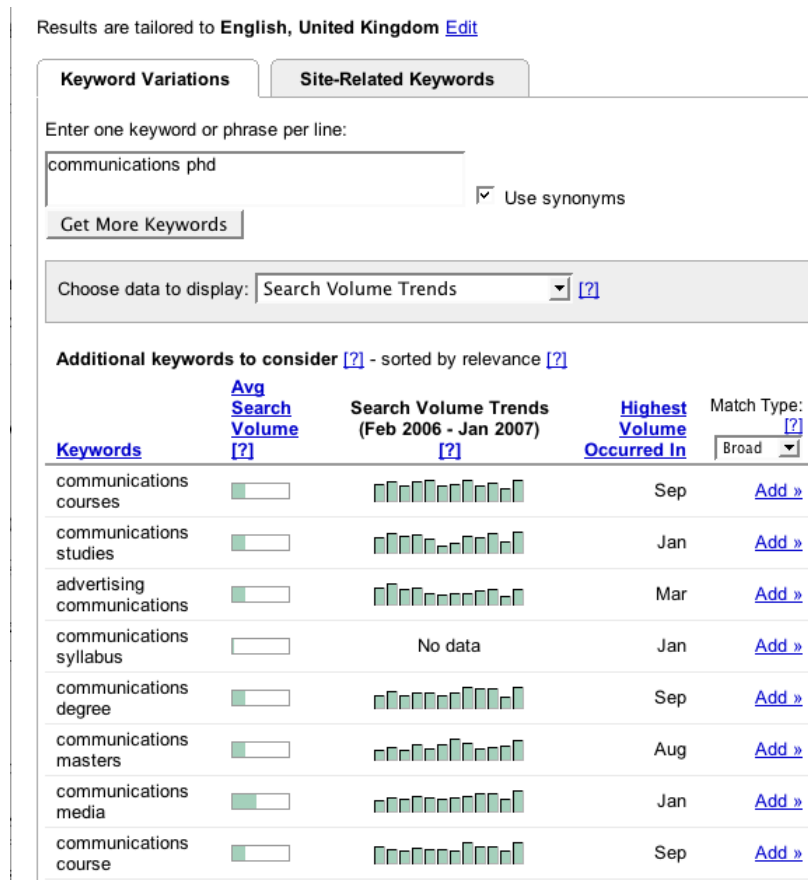
Retrieved 5 March 2007 from
https://adwords.google.co.uk/select/KeywordToolExternal?defaultView=2

In addition to the tools the search engines provide, third parties also provide competitive research tools which could be used either for advertising or for search engine optimisation (see Figure 25).

Figure 25: Search analytics tool adverisement.

From http://blog.searchenginewatch.com/blog/060628-202403, last accessed 20 March 2008.

The research provided by these programs into the language of searchers is then followed by the use of another program to assess the target website as it currently exists. This assessment reveals which words or phrases are repeated on the target site, and on which pages, and enables them to be compared to the language that the searchers are using. One such program, *Site Content Analyzer*, is shown in Figure 26.

Figure 26: *Site Content Analyzer* **example screenshot**

Once the target list of optimisation terms has been developed, the SEO gives advice to the client regarding how these are representing the website. This, in turn, leads to advice in terms of domain names, Web site structure (cf. Rogers & Zelman, 2002), and text on any given page – each can be subject to advice and change. Special technologies are available to assist with each stage of the process.

In addition to content on the site itself, many SEMs will develop networks of linked pages and websites which point to the target website. These may be other sites that the SEM solicits for links, or they may be sites created by the SEM for the sole purpose of linking to the target site through a series of "doorway pages" – although many search engines would consider this to be illegitimate (see Section 7.4).

Search engine companies such as Google or Yahoo are keen to draw a distinction between advertisers (whom they support and who are valued business partners) and

search engine optimisers, who are the subject of suspicion[57]. However, in practice, the people responsible for advertising and the people responsible for optimising are one and the same. My informants described the relationship between advertising and optimisation in the shape of two intersecting curves (see Figure 27). Early on in a relationship, an SEM may use keyword advertising because it is instantaneous. Meanwhile they will begin a programme of optimisation. As the traffic generated by optimisation kicks in, the advertising can be scaled down or even abandoned.

Figure 27: The relationship between traffic due to search engine advertising and search engine optimisation
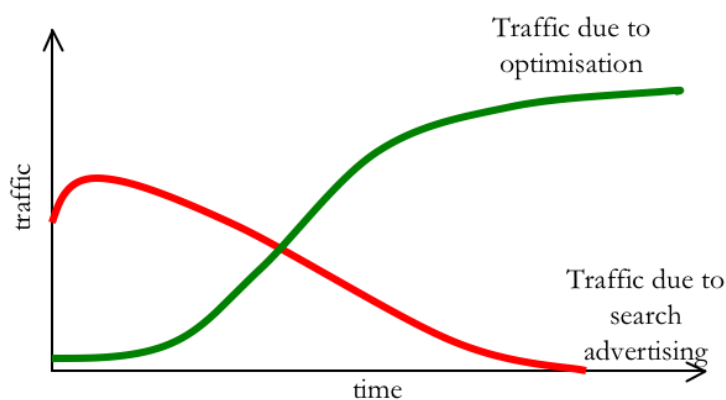
Source: Author, based on informant sketches

As with public relations, revenues generated from this process do not flow to the publishers (e.g. in this case the search engine firms), but to a range of small optimisation companies. It is in the search engine's financial interest to discourage optimisation and encourage advertising. The SEO and public relations business do have one large difference, however. In the world of print media, the editor or publisher has the final say about what goes into the newspaper or magazine. A search engine, being an automated process, has no such direct control. Therefore, what comes at the top of the list may not

---

[57] Google's help file (Google, 2007b) contains a long description of shady practices that it says SEOs may resort to, after a short disclaimer that "many SEOs…provide useful services for website owners" it suggests that site owners "seek out a few of the cautionary tales that have appeared in the press" before considering whether they will use an SEO. They then direct site owners to an article reporting a judgement against an SEO for failing to deliver on ranking promises, and also provide the FTC's address for complaints. Taken in total the page is anything but encouraging on the subject of using an SEO.

be exactly what the search engine provider would like. This observation leads the discussion into search engine spamming, the subject of the next section.

7.4 Working against each other: Search engine spam

The actions of search engine optimisers on the pages of their clients may cause the clients' websites to rise in the search engine listings, and most of these actions are seen by search engines and marketers as legitimate. However, there are some practices that are viewed as dubious or even fraudulent. For the purposes of this thesis, *search engine spam* is defined as follows: *any content on the Web that is intended specifically to deceive a search engine's indexing or ranking programs*[58]. Search engine spam has a long history and, in fact, has been one of the driving forces in search engine development. An engineer who was involved in creating one of the first search engines in 1993/1994 told me:

> *"Our…concern was, well, you want to make sure that the pages you serve are high-quality…Because before too long, you got people who tried to cheat the system in various ways. We called them word spammers. That actually took an extraordinary amount of time from us. For a small team that was damaging, you know, we couldn't work on the next generation search because we were dealing with those issues." (Interview C)*

In the early days, spammers used relatively simple techniques such as black text on a black background to fool search engines into ranking their pages more highly. But spam developed quickly. Another engineer who worked on a search engine in the mid-nineties told me that coping with spam, mostly detected by user complaints, was the day-to-day work of the engineering team:

> *"We definitely would see the spam get smarter. For example, we started seeing things that were obviously not readable but that were generated automatically: nonsense text that was designed to fool things that were trying to look for nonsense text. So they got very clever. So it's a tough one, it is an arms race." (Interview A)*

When Google launched in 1997, it was said to be unspammable, largely because instead of using primarily page content as the basis for its ranking, it used the link relationships

---

[58] This definition follows Perkins's (2003) quite closely. He defines search engine spam as "Any attempt to deceive a search engine's relevancy algorithm." He also defines what spam is not: "Anything that would still be done if search engines did not exist, or anything that a search engine has given written permission to do."

between different sites as a way to distinguish authoritative content (Brin & Page, 1998). This innovation, called PageRank, has come to dominate modern search engines. But Google and the rest of the search engines are as vulnerable to spam today as they were in 1994, for example, by techniques such as the 'link farm' in which a network of heavily cross-linked sites is set up in order to boost the rankings of all the sites included (Perkins, 2003). These link-farm sites then sell paid-performance advertising related to the original search query. The link farm page shown in Figure 28, a result from searching 'travel to Eritrea' on Google, has no content but rather a series of links and ads related to the initial search terms that are paid for on a cost-per-click basis.

Figure 28: A sample link farm page.
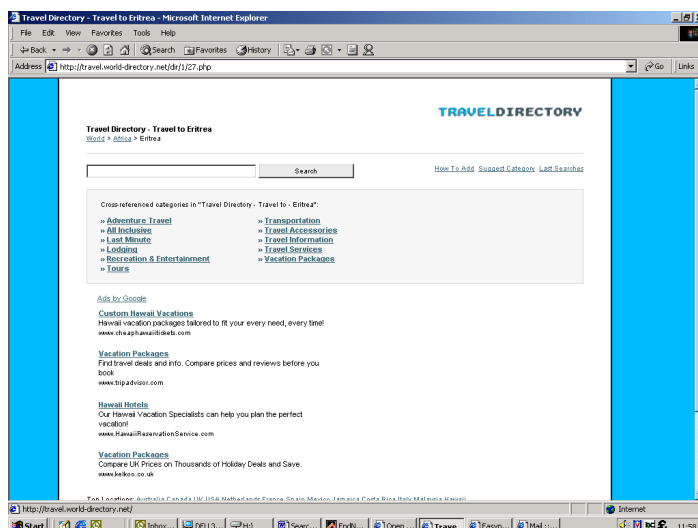
Source: http://travel.world-directory.net/dir/1/27.php, last accessed 1 July 2004.

Businesses built on search engine spam operate in a grey area – what they do is legal, but prohibited by search engine policy. In March 2008 the Canadian newspaper the *Financial Post* reported on a company, GeoSign, that operated just such a link farm as its primary business – with the additional caveat that, though it purchased keywords from Google to drive traffic to its link network, it diverted the traffic to Yahoo advertisements (Thompson, 2008). According to the report, GeoSign generated over $100 million a year in revenue, had 250 staff and had just signed a deal for $160 million in venture capital. In May 2007, Google forced the business to close by raising the price of keywords overnight.

The paradox that is demonstrated in Figure 28 is that while search engine relevancy teams battle against spam in the interests of bringing a high-quality result to the user, their own advertising formats – paid-placement ads – are at the heart of this new generation of spam. Paid-performance ads provide easy revenue generation for spammers who may operate hundreds of link-farmed sites. As content-free, link-farmed sites increase their presence in the top 10 or 20 search results, user satisfaction with search engines is likely to decrease.

Thus, one perspective is that search engine results are being subjected to a concerted assault in the name of commercial enterprise. This longstanding "arms race" consumes a great deal of the time and attention of the search engine providers, as they constantly 'tune' their indexing and ranking algorithms to try to exclude the new tricks of the spammers. It may even hamper or prevent the development of alternatives. A case in point is the service Nutch (www.nutch.org). Nutch is an open-source search engine, founded specifically in response to concerns about the quality of results produced by commercial search engines (which I refer to as bias in this thesis). A standard criticism of Nutch is that it will simply be swamped with spam, having done the spammer's work for them. Here is the Nutch reply:

> "*Won't open source just make it easier for sites to manipulate rankings?*
>
> "Search engines work hard to construct ranking algorithms that are immune to manipulation. Search engine optimisers still manage to reverse-engineer the ranking algorithms used by search engines, and improve the ranking of their pages. For example, many sites use link farms to manipulate search engines' link-based ranking algorithms, and search engines retaliate by improving their link-based algorithms to neutralize the effect of link farms."
>
> "With an open-source search engine, this will still happen, just out in the open. This is analagous to encryption and virus protection software. In the long term, making such algorithms open source makes them stronger, as more people can examine the source code to find flaws and suggest improvements. Thus we believe that an open source search engine has the potential to better resist manipulation of its rankings."
> (The Nutch Organization, 2004)

Nutch has since very quiet and its founder and lead developer left in 2007 to join Yahoo. Another open-source search engine called Wikia launched in 2008 and was immediately called "hackable" by industry commentators (Battelle, 2008). Thus, in 2009, the same

issues are still very much in evidence in terms of whether open-source search may be a viable alternative.

This section I have distinguished between legitimate search engine optimisation and illegitimate spamming practices. Because the laws are weak and both customers and optimisers may be unaware of the effects of particular actions, the distinction in practice is less clear. Legitimate business owners can become spammers almost by accident. One engineer working on search engine relevance at a large provider related the following story to me:

> *"I grew up in a [western US city], and one of the things my family would do, is we would go down to Southern Utah to stay at a couple of places. One of them was called the Hey-di-hi Inn[59]…Well, I ran across their site about a month or two ago, and just kind of looking at it because I thought it was kind of fun, it caught my eye, so I sat down to play with it. It turns out that they're spamming. And I actually called the guy, and what I found out is that they, along with you know, a zillion other tourist places in the Four Corners Area – Arizona, New Mexico, Utah, Colorado – all of them are basically hosted by this company…which was an outfit out of southern Utah. And you know, this company's been around for ten years, which is pretty much as long as the internet's been around. So they sound reputable, they've established good business relationships with all of these mom-and-pop motels, horse ride places, balloon ride places, everything, everything that needs reservations. What they provided was a generic reservations system. So, use our system, we'll do your reservations, so for all these places it's just great. And in addition what this company has branched out to is various forms of link spamming and keyword spamming. Now, it's not really hard-core, they're not the worst of the bunch, but what they're doing is pretty clearly spam. It's pretty obvious. And that's not a good thing to do. By the same token, even though our friend at the Hey-di-hi Inn is spamming, it's not what Mr Smith who's owned the place for 30 years – he doesn't know what he's doing. It's not a lot of spam. And as it turns out, heydihi.com is in fact the right URL for the Hey-di-hi Inn. So even though you're spamming, you're still the authoritative source. If you want to stay there, that's still where you want to go."*

The search engine is now faced with a quandary: what if a high quality source, an authoritative source, is also a fraudulent, deceptive source? What does a good quality result consist of in this case? The advertiser also may be uncertain as to what constitutes acceptable and unacceptable behaviour and has little redress in case of conflict; even large, reputable companies such as the automobile manufacturer BMW have been banned from Google for using spam techniques (Cutts, 2006).

---

[59] Not the real name of the inn.

Within the SEM community, this grey area between "legitimate" optimisation and "fraudulent" spam is one of active debate, with those in the former camp being referred to as "white hats" and those in the latter as "black hats."[60] The analogy is to the heroes and villains of the Wild West, with the Web implicitly identified as a relatively lawless land. While most SEMs I spoke with could point to spam activities, none self-identified as "black hats." Online, however, self-identified black hats are more easily located, at sites like www.blackhat-seo.com, www.seoblackhat.com, and www.bluehatseo.com. Spam, or "black hat" optimisation, is essentially about exploiting loopholes in the system which, while (probably) not illegal, are nonetheless likely to be disapproved of by search engines and clients alike. Black hat techniques include sophisticated technical hacks and there are even third-party tools for sale to help the would-be spammer become more efficient. Spammers may not work for clients or large organisations, but rather exist as small entrepreneurs for personal gain: "Let your greed meet your imagination," is the slogan of the advice site/blog www.blackhat-seo.com.

Nevertheless, many marketers who use techniques that the search engine companies would criticise as spam see themselves as simply professional marketers working in a new medium. They are able to mobilise technology and resources on behalf of their clients and themselves and, in many cases, these technologies and resources are provided directly by the search engine companies. The SEMs may also see the search engine companies as hypocrites. One of my informants commented on Yahoo's acquisition of comparison shopping engine Kelkoo, which used extensive networks of interlinked sites similar to link farms to promote itself:

> *If the search engines are that averse to it,…it certainly didn't stop the deal…This is the whole problem: search engines are companies…they have their own commercial impetus, so when they talk about spam, and you know when we talk about black hats and white hats, it's just a way of looking at our industry. It's no different from any other industry: PR's [public relations professionals] use all sort of techniques to get pieces out there…There are just different ways of doing it. (Interview K)*

---

[60] The white hat/black hat terminology was used by my informants, and is also used within the industry generally; for example, the Search Engine Strategies Conference held in December 2004 in Chicago, Ill. Included a panel called "Black Hat, White Hat, and Lots of Gray."

The interviewee suggests that the search engines companies are quite willing to use techniques they would disapprove of in others in order to improve their own business and further implies that the distinction between optimisation and spam is mostly one of perspective. In Chapter 6, the analysis suggested that some search engine engineers, while lacking a developed vocabulary, were concerned with issues such as the transparency of search results. Interestingly, SEMs often depict themselves and their clients as "the little guy" battling against the colossus of Google or Yahoo for a share of traffic. Thus, the legitimacy of various techniques of optimization is always capable of being called into question by both the search engine provider and the SEO and each may feel, at times, bitterly aggrieved by the other.

7.5 Dimensions of interaction

The previous sections have examined the ways in which search engine companies and SEMs co-operate, co-exist, and conflict when it comes to producing search engine results pages. In light of the conceptual framework in Chapter 2 and the discussion about technology-in-practice, several dimensions of routine interactions around the search engine results pages can be identified. First, there is an institutional and legal dimension comprising the "official" modes of shared work. Second, we find a technical dimension that includes the range of technologies provided by the producers and used by SEMs to help manage search advertising. Finally, there is a sociocultural dimension that includes a range of informal accommodations and actions.

Table 14, summarises the co-operative and conflictual elements that have been revealed by the analysis so far of the search engine marketer/search engine producer relation, along these dimensions. The examples in the table are taken from the foregoing sections. The dimensions identified can also be linked to the modes of interaction between agents and structures as discussed in Chapter 2. These modes of interaction reflect the way in which the choices that individuals make are likely to reinforce or potentially break down the social structure. Thus, the institutional/legal dimension associates the interaction of agents with structures of legitimation through sanctions; the technological dimension with structures of domination through the tools which enable or give power to various groups; and the socio-cultural dimension can be understood to link agents with structures of signification though communicative elements.

Table 14: Co-operative and conflictual elements in the relationship between search engine marketers and search engine producers

Dimension	Co-operative elements	Conflictual elements	(Mode of interaction)
Institutional/Legal	• Contracts for advertising • Preferential services for high-value advertisers	• Lawsuits re click fraud • Lawsuits re traffic	(sanction)
Technological	• Special interfaces (e.g. Google client control centre) • Keyword suggestion tools • Bid tracking tools	• 3rd party tools for algorithm and site analysis • Link farm creation tools and other spamming software	(power)
Socio-cultural	• Customer service representatives • Trade show and conference attendance • Branded gifts	• Policy of secrecy • Labelling (e.g., "black hats"; "spam wars")	(communication)

Source: Author.

As we have seen in the previous section, advertisers optimise, optimisers spam, and spammers advertise. Search engines companies themselves may also do these things on behalf of portions of their own product portfolios. Although there is a clear rhetorical stance on the part of the search engine companies against spam and optimisation and supportive of advertisers, in practice, the clear distinction breaks down.

What can we then say about the relationship between search engine providers and the search engine marketing and optimising communities in terms of the meanings they give to the search engine results pages? The search engine results pages we see are a product of both the algorithms written by search engine producers and an ongoing engagement by those who produce content online, be they website owners or their agents. These two communities, informed by user behaviour in the aggregate, both seek to develop the search engine results for maximum revenue. Although their discursive positions appear to be different, it is misleading to suggest that they are "at war." The spammer, the optimiser, and the advertiser are all dependent upon the search engine for their livelihood;

the search engine in turn is dependent upon the advertiser and the Web site owner for providing both funds and content.

As discussed in Chapter 2, Section 2.3.1, when treated as a boundary object, the search engine results pages can be understood as co-ordinating the practices of search engine producers and search engine marketers, making possible the intricate interdependence discussed in this chapter. Thus, the relations between producer and marketer may be strained and uneven in any given case; however, in the aggregate, the system appears to be much more balanced and the two communities co-operate to ensure that the maximum revenue is developed, based on the Web as a whole.

7.6 Conclusions

This chapter has inquired into the modes of interaction between SEMs and search engine producers. Although the relationship is sometimes characterised by conflictual terms such as "arms race," the idea of the "black hat" SEO and heated debates about click fraud, there are other ways to frame the interactions. There is much evidence that SEMs and search engine companies work together to produce the search engine as a vibrant commercial entity. This collaboration is most pronounced in the paid search advertising area, but it continues through the informal relations that search engine producers maintain with SEMs at conferences and in the way in which optimisation technology is freely available. The most notable cases of conflict, namely those surrounding click fraud and those surrounding "black hat" optimisation techniques, seem rather to be about the correct distribution of revenue to the marketplace, or, in other words, who deserves to be paid for traffic. Both search engine producers and search engine marketers see the internet as a legitimate marketing space or as "a big shopping mall," as the interviewee cited in Chapter 6 referred to it. At least part of their interaction is based on close observation and knowledge of the others' working practices, with a standard request from SEMs being for greater transparency.

The raw fodder of the search engine – the many billions of Web pages which make up the world-wide Web – is so vast and so dynamic that it does not permit rigorous control of individual search results pages on a massive scale. To some extent control is possible, as was discussed in Chapter 5 with reference to Google's self-censorship in China – but, to

some extent, the search results pages which the search engines return in answer to user queries must also necessarily be shaped by Web site owners and Web site optimisers. This is a crucial difference between navigational media and traditional media. From one perspective, it is as if the pages of the daily newspaper had risen up against the publisher. From another, it can be seen as the reclamation of media from the control of the conglomerates. In either case, it appears as though the search engine queries must of necessity be influenced or biased by advertising, optimisation, and spam.

The past four chapters have discussed the empirical evidence upon which this thesis is based; the next chapter analyses this evidence in the light the conceptual framework outlined in Chapter 2.

VIII

Conclusion

8.1 Introduction

This final chapter summarises the findings of this thesis, discusses the contributions that it makes to the field of media and communication studies and to the social construction of technology, reflects on the research process and suggests some ways forward for research. Throughout the thesis, the search engine has been viewed as a hybrid of technology and media. The conceptual framework developed in Chapter 2 drew on theoretical traditions appropriate to both the study of technology and of media. From the social construction of technology studies was drawn the concept of the *technology-in-practice;* from media and communication studies, specifically the political economy of media and communication, the ideas of *commodification* and *spatialisation* were used to understand major processes in the media industries. These two theoretical concepts were linked together drawing upon social theory, in particular, *structuration theory*, as an orienting device and as an analytical lens.

8.2 Summary of findings

In Chapter 1 of this thesis, the main research question was presented as: *Why does bias arise in search engine results?* In Chapter 2, this question was framed as a product of the conflict between different normative standards, rather than as a deviation from objectivity. From this perspective, the search engine results pages are transformed from a place of boring lists and links to a scene of high drama – the thrill of ranking first, the fury of being spammed, the greed for clicks, the disappointment of the missed opportunity. Onto this site of struggle, the ordinary user wanders first onto one site, then onto another,

frequently following the first link on the page and very often leaving a trail of invisible coins behind. Appendix E offers the reader additional insight into the significance of search engine bias for the user.

Following from that central question, Table 1 of Chapter 3 laid out the research design and the empirical questions. This section returns to those questions and summarises the findings of the thesis specifically with relation to the research questions as proposed in Chpater 3. However, a chief finding of this thesis is that search-engine bias is the result of a multiplicity of factors, none of them wholly under the control of a single group. The theoretical questions were: *How can we understand search engine results creation as a practice in which members of different communities participate and from which the search results emerge as a reified object? Where does bias stand in relation to that practice?* Chapters 6 and 7 addressed these questions.

In Chapter 6, these questions were approached through three key empirical sub-questions: *How do search engine producers conceive of search engine results? How do they make decisions about where to allocate resources and how to make changes to the search engines? What notions of quality exist, and with what consequences for the search results?* The discourse analysis of interviews with search engine producers in Chapter 6 showed that interviewees conceived of and judged their work along two related axes: first, how successful it was in the marketplace and, second, how "relevant" the search engine results were deemed to be. Thus, search engine producers appeared to construct the search index and search results so that the most important markets they serve will have the most relevant results and they will thereby achieve greater customer satisfaction and higher revenues. The evaluation of search technology using market criteria seemed to be naturalised; the "relevance" criteria, which might be seen as a non-market catch-all for quality criteria, seemed to be strongly interwoven with the market framework. Other potential ways of evaluating search quality seemed to be difficult for the interviewees to articulate. Striving for relevance and customer satisfaction as their highest goals, they undertook processes like index cleaning, blacklisting (the banning of particular sites irrespective of content), whitelisting (the promotion of particular sites irrespective of content), and the fine-tuning of the algorithm so that individual results matched desired outcomes. These processes were often carried out with high levels of secrecy. They create search engine results that are highly opaque to

users and a climate in which questions of the public interest in search engine results, as opposed to questions of technical efficiency or market success, are far removed from their everyday search operations.

In Chapter 7, the sub-questions addressed were *How do search marketers interact with search engine results? How do search engine results act as a boundary object between SEMs and producers? How else do these communities interact?* The chapter identified a range of independent, co-operative, and conflictual behaviours in the relation between the search engine producers and SEMs. The search engine results seemed to be used by SEMs as a kind of raw material upon which a scaffolding of various technologies was put in place to interact with the algorithmic technologies put forward by the search engine producers. A grey area between legitimate and illegitimate practices has spawned a whole industry of "black hat" marketers; but many businesses are involved in search engine spam, search engine optimisation, and search engine advertising, seeing each of them as tools to reach their online customers.

In my analysis, therefore, the search engine results can be seen as a tacitly negotiated product of both communities' (search engine producers' and SEMs') actions. The relationship between the two communities, though sometimes framed very negatively can, in many ways, be seen as a collaboration through which the search engine results are produced as a vibrant commodity. The money spent by companies to bring their websites closer to the top of the search engine results lists, and the position of search engines as important sources of leads for business, are crucial to search engine survival. While some practices may be objectionable, search engine marketing is an essential part of the search industry. The dynamics of the relation, with the search engine giants on one side and the typically small-sized marketer on the other, mean that search engines companies can afford to take some action against "spammers" while not threatening the industry system as a whole.

In relation to the findings of these two chapters, it would seem that search engine bias can best be understood as a by-product of a set of practices that emphasises the economic aspects of search engine results to the detriment of other aspects such as public interest aspects. Without a particular political agenda, search engine bias appears to emerge from

a desire to provide a profitable, efficient and competitive service (on the side of the search engine producers) and a desire to be as visible as possible to potential customers (on the side of the search engine marketers).

The second part of the methodology set out in Chapter 3, Table 1, focused on systems. The key question was as follows: *Which aspects and elements of the capitalist processes of spatialisation and commodification are linked to search engine bias and how?* Chapters 4 and 5 addressed this question, with Chapter 4 focusing primarily on commodification and Chapter 5, primarily on spatialisation.

In Chapter 4, the main empirical questions were as follows: *What is the ownership structure of the industry? Is the industry concentrated into certain companies? How has it developed over time? What is the role of advertising in the industry? How much does it contribute to revenue? How does it operate? How does the pure search product fit into the overall financial structure of the company? What contribution does it make to revenue?* In Chapter 4, I developed an historical perspective on the transformations of the search industry. Search engines originally were conceived of as technology products, but their producers quickly attempted to move them into the media space, using advertising based on traditional models, paying per page-view which was equated with earlier charging models based on circulation or viewer exposures. The engines had high audience figures and many search engines were acquired during the dot-com boom by traditional media companies seeking online synergies, such as Disney and NBC. However, it was not until the per-click, or *traffic model* of advertising was developed that financial success was achieved and, in the meantime, the search engines associated with large media conglomerates were closed. The switch in the main media commodity from audience (charging advertisers for unique viewers) to traffic (charging advertisers for clicks through to desired websites) is arguably a feature of the new 'logic' of navigational media, of which search engines are one example (other examples may be electronic programme guides, digital assistants, etc.). The advertising displayed by search engines is comprised of many smaller websites and constitutes a large network (which I have called "virtually integrated"), which very effectively captures this traffic for search engine advertisers. This chapter portrayed an oligopolistic industry structure through which a vast volume of internet traffic flows, all of which is exposed to tailored advertisements.

Without creating content, search engines have become the media giants of the internet, and their policies have a major effect upon the shape of the internet as we know it.

Chapter 5 addressed the process of spatialisation, answering the following questions, using four country cases as examples: *Is the industry concentrated into certain companies and/or geographies? What is the relationship between global search provision and local search provision? Can we distinguish a centre or periphery in current search engine operation?* The analysis in the chapter showed a strong presence by the big three American search engine companies (Google, Yahoo, and MSN), especially Google and Yahoo. Google was the major search provider in Germany and Yahoo was the major provider in Japan. In China, local search engine Baidu had the most traffic while, in South Africa, although it was difficult to determine, Google seemed to be doing very well, with local search engines such as Ananzi used for investigations within the South African context. The relationship between global and local search provision was characterised initially as two-tier system, where US-based providers competed in markets with large populations and strong advertising markets (potentially strong in the case of China), while national providers competed at the local or regional levels. A further analysis of the geographic organisation of the US-based providers suggested that, operationally, search provision can be split into those countries that *produce* search technology and those that receive it, with the receivers further divided between those that had dedicated services and those that did not. The chapter therefore offered some evidence for some centre-periphery dynamics such that users in those countries who are passive receivers of search would appear to be disadvantaged in seeking local content. Again, the picture was one of a growing oligopolistic industry, although it was inflected by local governmental regulation and direct intervention.

The final question set out in Chapter 3, Table 1, was *How can we understand the structuration of search engine bias? In other words, how is bias routinely embedded into both everyday actions and large-scale processes?* This following sections address this question.

8.3 The enacted search engine

Returning to the conceptual framework presented in Chapter 2, insights from Giddens's structuration theory were further developed, drawing on Wanda Orlikowski's theory of technological structuration. In particular, her idea of a *technology-in-practice* was used as a

basis upon which to frame the research question. The technology-in-practice concept links the local microcosms of technology usage with the larger macrocosm of technological development within the capitalist system. In this section I use the same concept to inform an interpretation of the data in Chapters 6 and 7, so as to build on and offer a critique of the robustness of the concept.

The different worlds of the search engine producer and the search engine marketer were examined in Chapters 6 and 7. In Chapter 6 it was argued that the discourses of search production are such that a discussion of the concept of bias is difficult, since it seems irrelevant to the major interpretative schemas that search engine producers seem to use in their everyday discourses: those of the market, of science and technology, and of war or conflict. Similarly, in Chapter 7 the search marketer, acting in a professional capacity to build traffic on behalf of a client or on his or her own behalf, was shown to have little concern for the overall bias of the engine except insofar as it may affect his or her own business. That advertising is not neutral is surely a tautology; and, once again, the irrelevance to marketers of the concept of bias in search engine results is very clear.

Applying Orlikowski's technology-in-practice framework encourages us to take each group in turn to examine the structuration process of search production, (see Chapter 2, Section 2.3.1). Figure 29 suggests how Orlikowski's model might apply in the analysis of search engine producers. At the very bottom of the figure is the box containing a description of the ongoing, situated use of technology in the search engine production context (see Figure 2 in Chapter 2). This describes the everyday, routine actions of those engineers involved in search production and is analogous to the level of "interaction" in the model that Giddens presents (see Chapter 2, Figure 1). The analysis of the interviews with search engineers suggested a daily routine of relevance judgments, weighting, tuning, developing the index and combating spammers, as well as adapting search to new formats such as video in order to improve the twin measures of relevance and customer satisfaction (see Chapter 6).

Figure 29: Online information retrieval technology-in-practice enacted by search engine engineers

Structure

Income generator

Competitive advantage

Information science achievement

Advertising platform

Online Information Retrieval Tool
Technology-in-Practice

Agency

Facilities, *e.g.,*	**Norms,** *e.g.,*	**Interpretative Schemes,** *e.g.,*
• spider	• develop relevant search results	
• index		• search is a business (market)
• ranking algorithm	• create customer satisfaction	
• syndication facilities		• search is a problem to be solved (science-technology)
• advertising engine	• protect search against marauders	
• massively parallel banks of PCs in various locations	• beat the competition	• search is a struggle (war)

Search engine producers consistently modify and tune the search engine to create greater relevance and customer satisfaction

Source: Author

Moving up the figure, the next box refers to the modalities of interaction – the *interpretative schemes*, *norms*, and *facilities* as developed by Giddens. Giddens says of this level, "Actors draw upon the modalities of structuration in the reproduction of systems of interaction, by the same token reconstituting their structural properties." (Giddens, 1984, p. 28). Orlikowski discusses these three modalities in terms of first, assumptions and knowledge about technology, second, protocols and etiquette related to technology and third, hardware and software, respectively. The analysis in Chapter 6 related the interpretative schemes as those of technology as a business (what I term the *market* schema), technology as a problem to be solved (the *science and technology* schema), and technology as a struggle (the *war* schema). The norms observed are related to these – the creation of relevant searches that boost customer satisfaction and the desire to protect search against spammers and beat the competition. Giddens emphasises that "the communication of meaning in interaction…is separable only analytically from the operation of normative sanctions." (Giddens, 1984, p. 28). The facilities which help to bring these norms and interpretative schemes to fruition include the algorithms that run the search as well as

syndication and advertising algorithms and extremely large banks of computers and electrical power situated worldwide.

The top box of Figure 29 describes the structures that are instantiated in the routine situated use of search engine technology by search engineers. In this context, a *structure* is a rule-resource set that exists because it is mobilised by agents in their ordinary practice, and it is this kind of structure to which Orlikowki refers by the term "technology-in-practice" (Orlikowski, 2000). The situated action that motivates and is motivated by the modalities discussed above can be understood as enacting the search engine primarily as a tool for online information retrieval and, recursively, it is in pursuit of better online information retrieval (for example, adding the ability to search for video, or search within local geographic boundaries) that much of the day-to-day work of the search engine engineer is structured. Other technologies-in-practice, however, are also instantiated and a consideration of the evidence in Chapters 4 and 5 suggests that these include, at least, search as an advertising platform, search as an information science achievement, and search as a competitive advantage for the corporation. Finally, the search engine as a technology is an income generator – the source of a search engineer's livelihood.

In Figure 30, a similar diagram is constructed for search engine marketers. The lowest box shows that the situated use of the search engine for search engine marketers is as a publicity venue or advertising publisher for themselves or for their clients (in the case of the SEM agency). Their modes of interaction with the technology are in some respects different and, in other respects, similar to those of the search engine producer. The primary interpretative schemes are once again the market-based and science-technology schemes, but with a different inflection. The search engine marketer was found to account for his or her interactions with the search engine technology in terms of professional marketing practice; the focus of the practice is either the number of visitors or sales (for those working on their own behalf) or on client satisfaction for those working as agents. The search marketer may also see himself or herself as an expert in a kind of information science in which the search engine as developed by search engine producers is the research object. The associated norms are related to maintaining the business relationship – having client relationships and relationships with the search engines, having solid legal foundations (contracts, etc.) and maintaining ethical business

practices. Ethics are hotly debated in the search marketing industry, so in certain cases tending more towards the "black hat," one may say that the norm is to keep within *legal* boundaries, and the emphasis would shift to the effective exploitation of search traffic. The hardware and software facilities that are engaged to enact these schemes include the search engines themselves and tools provided for marketers by the search engines, various third party software tools with additional facilities, the software necessary for coding Web sites and registering and maintaining domain names, and the software needed to measure traffic going to the site. Relatively modest amount of hardware are needed – certainly nothing like the scale needed for search engine production. Taken together, these modes of interaction can be understood to structure and be structured by the search engine-in-practice as an internet marketing platform. At the same time, the search engine is enacted as an object of research interest as the knowledge about search engines is seen as good in itself. Finally, as for the search engine producers, search engines are the source of the livelihood of the SEM, although indirectly.

Figure 30: Internet marketing tool technology-in-practice enacted by search engine marketers
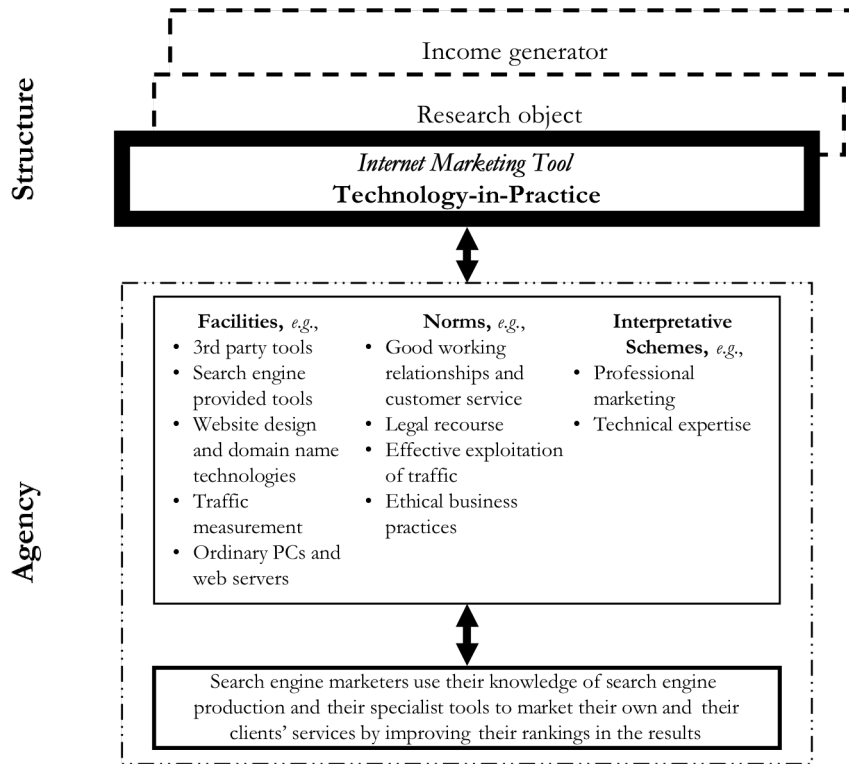
Having discussed how the technology-in-practice model can be applied to search engine producers and marketers, the analysis above presents us with a number of problems. Firstly, in this model the different communities, that is to say, the search engine producers and the search engine optimisers[61], are presented as being separate, whereas we know from the analysis conducted in Chapter 7 that they interact along multiple dimensions, sometimes conflicting, sometimes co-operating, and sometimes operating relatively independently. Secondly, although the model captures a process, it gives little hint as to the dynamic elements of the process – in a sense, it implies that the structuration process described is self-sustaining. Thirdly, the relation of the structuration processes captured in the model to the wider social systems such as those examined in Chapters 4 and 5 of this thesis is not revealed using the model. This is related to the particular interpretation of structuration theory upon which Orlikowski builds her technology-in-practice mode,

---

[61] These could be called *relevant social groups* (Bijker, Hughes, & Pinch, 1989), see also Chapter 2.

which is, in turn, related to her focus on the organisation as the main analytical unit. Giddens's own discussion of this theory, by contrast, allows for a range of levels including not only individual and group processes, but also systemic and societal processes.

The distinction between structure and system in Giddens's writing is fundamental. Structures, he says, are: "Rule-resource sets, implicated in the institutional articulation of social systems" (Giddens, 1984, p. 377); whereas a system is: "The patterning of social relations across time-space, understood as reproduced practices" (Giddens, 1984, p. 377). In my conceptual framework, developed from an interpretation of Giddens' theory, the questions of how structures inter-relate and how they continue or are replaced or eliminated are regarded as *systemic* issues, not structural ones. All three of the problems discussed in the previous paragraph (of inter-relation, of dynamism, and of history) are important to an understanding of the search engine's systemic context. Although Orlikowski does not deal issues relating to structure and system integration, it is important to do so to understand how search-in-practice may be related to the larger search system and to gain insight into additional features of the production of bias.

8.4 The development of technological structures

This section begins to develop the connection between the local structuration processes in search, described above, and the wider system of the search industry, as discussed in Chapters 4 and 5.

One essential insight is that this connection is *cultural* and that structure and system are in many ways organised and linked through cultural practices. In his formulation of structuration theory, discussing how system and structure are inter-related, Giddens does not mention culture but rather he emphasises practice and routine. He suggests that "structure has always to be conceived as a property of social systems, 'carried' in reproduced practices embedded in time and space" (1984, p. 170) and that "…it is not the case that actors create social systems: they reproduce or transform them, remaking what is already made in the continuity of praxis." (1984, p. 171). However, anthropologists have elaborated on Giddens's arguments by emphasising the importance of culture as a key to organising practice. One classic definition of culture is that of the anthropologist Geertz: "an historically transmitted pattern of meanings embodied in symbols, a system of

inherited conceptions expressed in symbolic forms by means of which men [*sic*] communicate, perpetuate, and develop their knowledge about and attitudes towards life" (1973, p. 89). While Geertz's definition of culture is primarily symbolic, most anthropologists would also acknowledge a practice aspect and a material aspect to culture[62]. Each of these aspects can be understood to contribute to the structuration process across the social system. This section discusses how the conceptual apparatus for this study developed at the outset can be productively linked to developments in cultural theory and in theories of technological innovation, which, in turn further leads to a development of the conceptual framework presented in Chapter 2.

The starting point for the analysis in Chapter 6 of search engine producer discourse was the "interpretive scheme", which Giddens characterises as a way of accounting for and explaining the world. According to Giddens, "'interpretative schemes' are the modes of typification incorporated within actors' stocks of knowledge," (1984, p. 29). In my reading of structuration theory, interpretative schemes help to govern the allocation of resources and thereby reinforce or potentially change larger structures. From the social shaping of technology studies perspective, a somewhat related concept has been developed known variously as the *technological frame* (Bijker, 1995) or *technological frame of reference* (Orlikowski & Gash, 1994). According to Bijker, a "technological frame" refers to the "shared cognitive frame that defines a relevant social group and constitutes members' common interpretation of an artefact." (1995, p. 123). Similarly, Orlikowski and Gash argue that a "technological frame of reference" governs the way in which people perceive technology as appropriate or inappropriate; thus, their definition of the technological frame of reference "includes not only the nature and role of the technology itself, but the specific conditions, application, and consequences of that technology in particular contexts." (1994, p. 178)

The anthropologist Sewell (1992) develops structuration theory by suggesting how culture functions to help maintain social structures (or "rule-resource sets" as Giddens terms them). According to Sewell, while material objects (such as money or buildings or artefacts) have an independent existence, their simple existence is not enough to create

---

[62] Indeed, Geertz's pioneering use of "thick description" (Geertz, 1973) puts culture as the object of empirical study squarely in the realm of observable practice.

value or power. For that, cultural or symbolic "schemas" are needed. To take one example, paper money has no value without its meaning as a medium of exchange. Resources like money and technology, therefore, "embody cultural schemas," but not unambiguously. The value of a resource is dependent upon the ways in which cultural schemas can mobilize that and other resources. Reciprocally, the effectiveness and persistence of a cultural schema are bound up with how well it can organise and mobilise new and existing resources.

A key feature of these cultural schemas is that they are generalizable or, as Sewell puts it (drawing from Bourdieu), *transposable*; that is, "that they can be applied to a wide and not fully predictable range of cases outside the context where they are initially learned" (Sewell, 1992, p. 17). In the discourse of the search engine producer, the market, science, and war can be seen as precisely such transposable allegories, whose demands constrain the pursuit of alternate standards of search quality. For, as Giddens and Sewell both note, schemas are crucial elements of structure that mobilize resources on behalf of their users. Returning to Giddens and Sewell's portrayal of the relationship between rules and resources, therefore, we might suggest that the "technological frame" or "technological frame of reference" is the way in which producers cognitively organize technology not only to interpret it and give it meaning, as Bijker and Orlikowski and Gash suggest, but also in order to mobilize other resources around their interpretation.

Viewing a technological frame in this way also allows us to account for the *strategic* use of different schemas. The discourse analysts Potter and Wetherell (1987) argue that each speech act is an achievement on the part of the speaker and that different "interpretative repertoires" have different functions for speakers and will appear in different contexts. It would seem that the strategic use of interpretative repertoires fits much more closely with the empirical analysis in this study than the unified technological frame approach of either Bijker or Orlikowski and Gash.

As Chapter 6 suggested, search producers find the market schema is especially productive when it comes to mobilizing resources within the pre-existing structure of a business. Nevertheless, producers do not portray themselves as experts in the market. They may be engineers, researchers, or designers and they may have become managers or even business

owners over the years, as happened to several of my interviewees. However, they did not appear to identify themselves as business people and when working on search engines their discourse suggest that they felt their ability to affect major features of search was circumscribed as soon as their companies began to make significant amounts of money from the search. They acted discursively to reclaim their abilities to influence search technology by claiming expertise on an 'objective' measure of search quality, 'relevance,' which they linked to customer satisfaction and therefore revenue.

Thus, although search producers and search optimisers, like all other cultural actors, have a multiplicity of transposable schemas at their disposal (examples might include art, or public service, or family, or many others), the implication is that these schemas are not a simple matter of choice – they have material consequences for their users.

The theoretical insights into structuration processes suggest that technological schemas and associated norms will have an effect on how resources are directed within society, helping to maintain old structures or create new ones. In the case of search production, it is clear that a considerable amount of resource is dedicated to maintaining search quality in the form of relevance and of customer satisfaction. In the case of search optimisation, the search engine is valued primarily as a gateway to a particular website and resources are directed to improving access to that particular website. And as Sewell maintains, the accumulation of resources, in turn, *validates* schemas and fosters their continued use. As he puts it: "Schemas not empowered or regenerated by resources would eventually be abandoned and forgotten, just as resources without cultural schemas to direct their use would eventually dissipate and decay." (Sewell, 1992, p. 13) This suggests that pre-existing systems, such as those of corporate hierarchy or, indeed, capitalism, will profoundly influence of technological schemas and norms become widespread and influential.

8.5 The search system

The intertwining of symbolic and material culture through practices, suggested by in the cultural approach to technological structuration, has implications for understanding how the everyday decisions and actions of search producers and search optimisers may be related to the historical development of search engines. As Chapter 4 showed, the history of search must be considered in the context of the dot-com boom, the explosion of

investment into internet technology which peaked in 2000 and was shortly followed by the dot-com bust. Perez (2002), in her historical analysis of the similar booms and busts that have occurred regularly since the 1770s with the start of the Industrial Revolution in England, discusses the linkage between technological, cultural, and financial elements that characterise these economic cycles or long waves[63]. Each cycle, Perez suggests, follows the sequence *technological revolution – financial bubble – collapse – golden age – political unrest* and is driven by developments in technology leading to major increases in productive capacity, and also by associated speculation by the financial markets into firms using those technologies. The *technological revolution* is defined as "a powerful and highly visible cluster of new and dynamic technologies, products, and industries, capable of bringing about an upheaval in the whole fabric of the economy and of propelling a long-term upsurge of development" (Perez, 2002, p. 8). Importantly, associated with this revolution is an associated change in practice and meaning, which Perez calls the *techno-economic paradigm.* She defines as techno-economic paradigm as "a best-practice model made up on a set of all-pervasive generic technological and organisational principles....When generally adopted, these principles become the common-sense basis for organising any activity and structuring any institution" (p. 15). Although Perez does not discuss the idea of practice, the "common sense" that Perez refers to can be related to the idea of technology-in-practice, if it is interpreted as a new schema of the kind discussed in this thesis; in other words, as a cultural vehicle that enables proponents to marshal resources that have been transformed by the technical revolution.

Perez's argument is one in which the transformation of capitalism is spurred by technology (creating resources) which combine with a changed "common sense" (new rules) to produce new structures (rule-resource sets), including both institutions and material goods. Following this argument, the principles of creating a prototypical business in the "network society" can be seen to be based in part on the "common sense" of the computer engineers who created the technological revolution of the microchip. If that is so, the roots of the search engine culture are to be found in Silicon Valley.

Castells (2001) suggests that "internet culture," by which he essentially means the culture of internet developers, based in Silicon Valley, has four aspects. The first aspect is a

---

[63] "Long waves" were first proposed by the Soviet economist Kondratiev in 1925.

culture of techno-meritocracy, derived from academic norms, in which discovery, especially the kind of discovery that leads to technical improvement, is the highest value. The second aspect is the "hacker" culture that values freedom, especially freedom to create, regardless of organisational rules. The third aspect is the culture of the virtual communitarian which values lack of hierarchy, free communication and self-directed networking as the ideal community structure. The final aspect is the culture of the entrepreneurs. One might call these levels those of the programmer, the artist, the citizen and the businessperson. According to Castells, the entrepreneurial culture is essentially intertwined with the other elements of the internet culture: "The strategy is to change the world through technology and to be rewarded with money and power, via the workings of the financial markets" (2001, p. 57). He describes a social system in which money fuels technology and comes to define personal success and intellectual achievement: "the fact that the reward is external (money) rather than internal (Puritan ethic of self-improvement by honest work) has considerable consequences for the culture…so that ideas, work, and the personal accumulation of wealth tend to be associated in the same movement" (2001, p. 58). The system of entrepreneurship and innovation includes inventors, technologists, and venture capitalists, who "come together in a process of production and innovation that ultimately creates companies, makes money, and, as a byproduct, delivers technology, goods, and services. In this process, the relationship between capital and innovation is internalised" (2001, p. 58).

Castell's association of technological development and economic success with virtue and his depiction of the intense relationship between capital and innovation in the culture of internet production is resonant with the results of the analysis in Chapter 6, where technical virtue, financial success, and the persona of the guardian of the Web were shown to be linked in search engine producer discourse. The internet culture Castells describes leaves space for freedom and individual networking but places a premium on individual intellectual achievements. In terms of the factors contributing to search engine bias, this same culture can be understood to hold together the idea of intellectual, technical achievement to achieve high monetary returns and rewards. This recalls the equation of relevance=customer satisfaction=revenue that was discussed in Chapter 6. The public aspect of internet culture, exemplified by Castell's virtual communitarian, is one in which the individual is free to associate and network on an *ad hoc* basis, not one in

which permanent and binding associations are formed, and not one in which public discussion and regulation are expected to place limits on behaviour.

From this perspective, therefore, questions of bias in search are associated with larger trends in capitalism and the large-scale or macro-structure of the search engine business. These concerns seemed to be 'naturalised' through specific technological schemas that are created through participation in the culture of search engine production and reified in search results.

8.6 Rethinking the conceptual framework

This section reflects on the construction of the research, especially the conceptual framework, and draws out some of the theoretical implications contained in the study.

In Chapter 2, I presented the conceptual framework visualised as two circles, the inner circle being practice and the outer circle, context, with the two circles being related by structuration processes. The analysis above suggests a revision of this framework. First, the initial framework considered one level as "context" and the other level as "practice." In a modified framework, both are levels of practice, but there appear to be hierarchies of practices, with different people participating in different practices. When considering a single group of people (e.g., search engine producers or search engine marketers), the levels of practice blur into one another, so that the patterns observed when viewing the historical construction of the search engine industry leave their imprint on the practices of daily life, and the practices of small groups in turn echo up to the larger scale of the industry. The depiction of the conceptual framework therefore can be changed to show two sets of parallel processes, each investigated independently, while the relation between them is now more in focus as an enquiry in its own right. I now refer to the levels as "practices" and "systems" while bearing in mind Giddens's conceptualisation of systems as structures extended through space-time (Giddens, 1984, p. 377). In addition, the objects of inquiry are different at each level. The revised framework shown in Figure 31 also more closely matches the research design contained in Chapter 3, Table 1, where Level A (agents) is linked to practices and Level B (structures) is linked to systems.

Figure 31: Revised conceptual framework

Analytical Level	Objects of inquiry	Processes of interest
Practices	technology-in-practice in relevant social groups	Participation Reification
	relation between practice and structure	Structuration
Systems	industry history and industry geography	Spatialisation Commodification
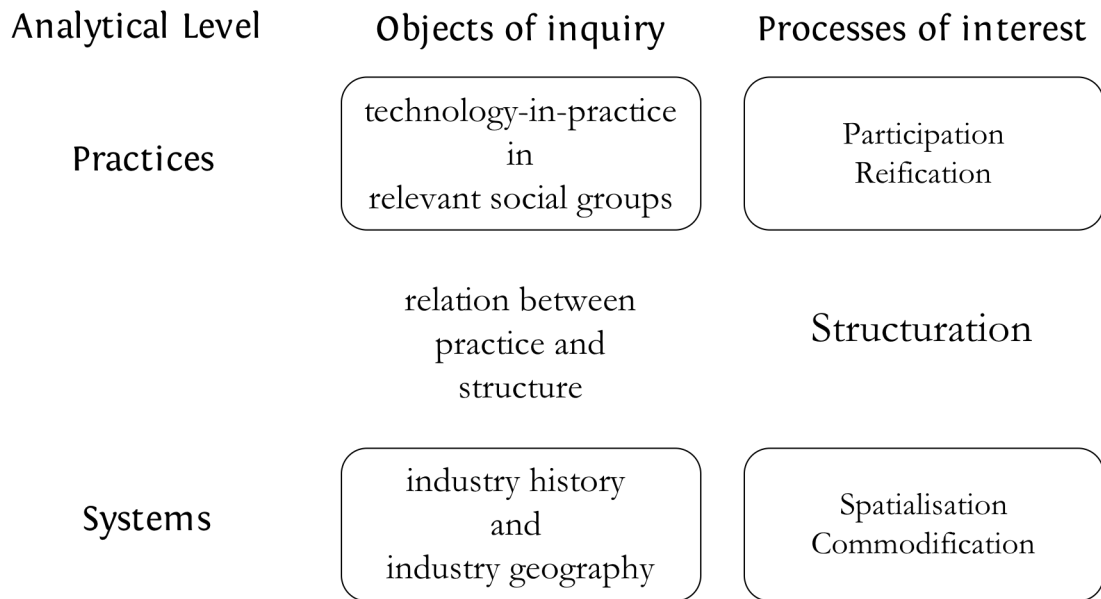
In Sections 8.4 to 8.5, I argued that the relations between practices and systems, the middle level of the diagram, are heavily influenced by culture; specifically, the local culture of the groups in question. This analysis suggests that although Giddens characterises resources and rules (interpretative schemas and norms) separately, it is productive to view them as a mutually constitutive duality. The schemas chosen and used by relevant social groups, then, are likely to be strategic, in that certain schemas are likely to liberate more resources. This choice, however, will be highly influenced by pre-existing rule/resource sets. In the case of search engines, I have argued that it is predominantly the culture of Silicon Valley that provides the "transposable schemes" (to use Sewell's phrase) that the search engine producers adopt in making meaning from their daily practices.

Although structuration theory often serves more as an "orienting device" than as a methodological approach, it was through a critical engagement with structuration theory itself as it informed this study that the modifications to the conceptual framework emerged. First, the idea of the "technological frame," as developed by Pinch and Bijker and Orlikowski and Gash, was re-conceptualised as a set of "technological schemas" – a non-unitary but complementary set of meanings that help actors to make sense of technology. This, in turn, was related to Giddens's concept of the "interpretative scheme" as a key element of structuration. Helpful, also, was Sewell's concept of the *duality* of rules

and resources. By considering technological schemas to be intrinsically linked to the availability of resources, the apparently strategic use of different schemas by my interviewees was made more explicable. These schemas, in turn, were defined by Sewell as "cultural schemas" – in the case of the technological schema as I defined it, it made sense to think of the local culture as a pool of potential technological schemas, of which some eventually were used by the interviewees.

Finally, Perez's discussion of the dynamics of capitalism on a large scale which implicates technologies and her concept of the techno-economic paradigm, led me to suggest that certain *complexes* of schemas might be important in a powerful liberation of new resources, becoming the "common sense" of a new techno-economic paradigm. And, turning to Castells's characterisation of the 'internet culture,' the equation of personal virtue, technical proficiency, and money-making also seemed to buoy up the idea that, of the multiplicity of potentially available schemas or complexes of schemas, some may be more widespread or even, in some sense, dominant, in that they reflect the default model to be adopted, for example, when operating a business or producing a specific technology such as the search engine.

Indeed, some of these local cultural characteristics seem to inflect the business model of search: in the early stages, the ready access to stock market money shows (once again) the close associations between technology and finance in Silicon Valley. The preference for loose affiliations, characteristic of Castells's "virtual communitarians," can be seen in the way the search engine firms are networked with others in a Web of commercial ties that I have called 'virtual integration.' This very looseness suggests a re-configuration of some of the insights derived from political economy which often seem to assume stable firms with relatively clear ownership. In the new media market of search engines, I have suggested that constructing an audience value chain is analytically more helpful in many respects than focusing on the value chain for content, which, after all, may contain a single blogger or other micro-publisher.

8.7 Reflections on the research

During the writing of this thesis, the amount of literature on search engines has been steadily on the rise, signalling a welcome concern with the role of search engines from academics, public bodies and private individuals alike. There also seems to be an increase in popular concern prompted by Google's ever-increasing share of the search market and there have been persistent rumours of changes in ownership structure, such as the acquisition of Yahoo by Microsoft.

Despite this increase in attention, this study is distinctive in that it offers a scholarly study that includes primary qualitative data from search engine producers and an extensive treatment of search engine optimisers and their role in shaping search results. It also presents an overview of search engine history and a comparison of the search environment in several different countries. Empirically, then, this thesis charts new ground and forms a starting point for other researchers interested in this important class of industry actors.

The research has yielded theoretical insights. In Section 8.6, my rethinking of the initial conceptual framework has resulted in an extension of Orlikowski's concept of technological structuration by integrating ideas about systems and cultural issues. While the local processes of structuration are important for research, this study emphasises that local structures fit into larger systems. This is perhaps what Giddens means when he refers to 'structuring principles.' In any case, an investigation of whether the same structuring principles are at play in different scenarios of technological development would be an interesting direction for future research.

The initial conceptual framework and research design for this study proved to be robust insofar as it guided the juxtaposition of elements of media and communication theory with those drawn from studies of the social construction of technology. The latter focuses mainly on the development of the technological form, while the former often takes the realm of symbolic content to be its object of study. And though McLuhan (McLuhan, 1964/2001) declared many years ago the importance of form to content by saying "the medium is the message" this study suggests that the message becomes the medium: it is not only the culture of the producers, but the interaction of others (such as

search engine marketers) with the content they produce that creates the technological form. Message and medium, it seems now, are interdependent. A converged research design is necessary to deal with the complex dynamics of the emerging media form of the search engine.

Other insights also have come to light. Conceptualising the competition in media as one for audiences rather than for product sales is not a new approach. Nevertheless, using the technique of value chain analysis to focus on the audience rather than content has proved particularly helpful in the case of search engines. As media content production grows more fragmented and is produced more and more by individuals rather than by institutions and, thus, becomes cheaper and more plentiful, its significance in the strategies of media companies may diminish. In tandem, the ability to attract or direct a large audience may grow in value. Thus, this approach can be applied to understand the strategies and development of other media actors.

Search engines, I have argued, are interesting in their own right: these companies are major new entrants into the media industry, they provide the major destinations for internet users around the globe and they are a major source of customers and readers for retail and publishing businesses. However, I have also referred to them as navigational media, and by this term have sought to indicate a type of technically-based media actor that organises and directs audiences or users to various types of content. Search engines can be seen as a class or case of technological development, but they are not simply that, just as a media business is not simply any other kind of business. Nor, I would argue, can search "content" be understood without taking into account its technological roots. At the heart of this study is the idea that form and content cannot be easily separated. Bias is an issue then has particular resonance when the audience or user depends on the service to alert them to valuable content, however that is defined.

The dynamics observed in this study – the power of oligopolistic companies, the potential centre and periphery of global dynamics, the manipulation of results by members of the network and the continual development of the algorithm for improved quality within the cultural schemes of the producers – may well be in evidence with other navigational media. Examples might include services such as the product search interface provided by

eBay and Amazon and the music and video search services provided by Apple's iTunes and YouTube (a Google subsidiary). Various other types of searching and ranking mechanisms may share similar dynamics, especially when the results have a clear monetary value and when user feedback and response is possible. It may also have resonances with other cases which are less obvious: online games, for example, have shown that 'virtual' economies can create real jobs and real economic advantages for those who know the technology and can manipulate it, and the same may be true of social networking technologies, where certain users (be they individuals or companies) are able to direct large amounts of traffic, in effect, becoming navigational entities.

Conceptualising a virtually integrated media organisation, in which control of content is non-existent and presumed to be a non-issue, may provides a new basis for analysis of other media firms and for developing public policy which, in the past, has been largely focused on content regulation. The regulation of access to content and remedies for those who feel that they are unfairly inaccessible are likely areas for policy development.

Methodologically, this work integrates many techniques. These range from the individual interview to use of data to depict large-scale patterns. The use of this variety of methods has been challenging, but on reflection I would not change this aspect of the study. The insights gained from the juxtaposition of the findings from the different methods enabled me to uncover new empirical and theoretical insights through a work based on interdisciplinarity, integration and synthesis.

However, there are some methodological issues that are worthy of consideration by future researchers. If one were to begin a similar research project, my advice would be first, to locate oneself in California. Search engines are still a local business and, as this research shows, local culture is important. More than that, in-depth interviews and participant observation depend upon long preparation in the field. If I had spent two or three years in California as a part of this research, I might have been able to develop a deeper analysis and to obtain more data than I was able to do working from London. A key informant would be high on my list of priorities for any prospective researcher.

Second, I underestimated the challenge of investigating search across different geographies. The findings contained in this study are relatively limited and the data

gathering process consumed a great deal of time. The internet has a different history in every country and the large multinationals interact differently in each geography. In this respect, setting up a network of research colleagues might be a helpful future approach, although perhaps not one available to the PhD student researcher. In retrospect, perhaps a travel budget and time spent at a foreign university would have helped to extend the data available for the study.

8.8 Directions for further research

My aim in writing this thesis was not to "solve" the "problem" of search engine bias, but rather to raise and examine it as an issue, so that the results can serve as a platform for both future research and debate. More work on the topic is clearly indicated. For example, if search engine bias is to be addressed within policy circles, then documentation of its implications is paramount – in effect, asking the question *"Who does search engine bias harm, in what manner, and to what degree?"* This is the area of search engine effects addressed in Chapter 1, section 1.2.4. Continuing research in this area, and reviews of the existing literature, are a natural outgrowth of this thesis.

As discussed in Chapter 4, advertising, it seems, may function slightly differently online to how it has been conceptualized for other electronic and print media. The specific industrial form of online advertising and its relation to online content of various types is a subject that has yet to be investigated in depth. This too is an area of research that should be developed.

Also in Chapter 4, the thesis began to investigate the relations between social media networks and search engines, and to use the concepts outlined in my presentation of navigational media to help analyse social network business strategies. The relations between search, social networks, and other media will continue, in my view, to be a fruitful subject for research.

During the course of this research it also became apparent that there is a dearth of research into the shape of the online space in various countries – which sites are important, how does the money flow, and what is the history of the internet in various geographies? This is a basic research area for those who study the Web.

8.9 Ways forward

It was always my intention to be cautiously constructive with respect to the "problem" of bias in search engines. If bias in search is a problem (as I have suggested, although others may remain unconvinced), then what are the possibilities for addressing it? From the research in the previous chapters, it is clear that no one group is entirely responsible and that potential solutions, if warranted, will be complex.

First, search thrives upon the commodification of the whole Web, bringing ever more advertising to ever more websites, creating a larger and larger virtually integrated portal. In the long term, this will mean that more online and possibly offline media will be added to the search index. Sites with more traffic will always remain top priority, however, and we may expect countries or areas without large advertising markets or the potential to develop such markets, to remain disadvantaged in terms of coverage in the search index in the absence of some kind of intervention.

Second, the syndicated advertising model creates many opportunities for search manipulation and spam, as well as for click fraud and various unsavoury business practices. Further, the value of traffic for online retailers and publishers is such that search engine marketing is here to stay. Any solution that might be proposed in an attempt to address bias would need to consider the secondary actors of the search value chain – the search genie cannot be put back in the bottle.

Finally, the prevailing wisdom and common sense within the search engine production community presents challenges for those who wish to discuss public interest or information ethics in connection with the search engine industry as questions in this area infringe on the values of "relevance" or of "customer satisfaction". It will be the role of actors outside the search engine community, no doubt with the support from certain actors within it, to raise awareness of issues around bias and its implications for the public interest and to provide potential alternatives.

From the foregoing discussion it will be clear that a technological solution to the problem – a better ranking algorithm, for example – will at best be partial. There seems to me to be a need for better transparency – just what, who, and how are sites included or

excluded, weighted up or reduced in weight in creating ranking decisions? The economic value of links also cannot be ignored, but search traffic needs to be managed to benefit the consumer and the citizen alike.

In Chapter 1, the question of the public interest was raised in conjunction with the justification for public-service television and online. Little public discussion has taken place with regard to the potential for internet media to also be subject to any form of public service goals. Part of the explanation for this may lie precisely in the strength of the culturally-bound schemas which give the search engine and new online media businesses their competitive advantage. The "internet culture" is predominantly individualistic, market oriented and opposed to most forms of regulation. Putting the case for a public interest in online media is fraught with difficulty. During this research I was met with bemusement and puzzlement, as well as encouragement, from many sources. Having said that, search engines operate in many parts of the world, notably in Europe, where the tradition of public service media is strong as compared the to the US. These parts of the world may be the best places to encourage discussion about how innovative approaches can be used to create a search engine industry that can help to deliver less biased search results and, arguably in consequence, a fairer information society for all.

Appendices

Appendix A: Search engine operations: A brief overview

It is be helpful at this stage to review the operation of a search engine from a technical and organizational perspective. This is a simplified view, but may be helpful in making sense of the chapters that follow. This Appendix is based on documents, observations and interviews with search engine producers, distributors, optimizers and commentators (see Chapter 3, Methodology).

There are three elements that characterise any internet search engine, whether the search is powered automatically, by human editors, or by a mixture of the two.

1. The *index*. The index is the database that stores all the information about the documents, pages, or files that the search engine finds.

2. The *search algorithm*. This is the set of programming rules that searches the index, finds matches to the user's query, and ranks them.

3. The *front end*. This controls how the search results delivered by the algorithm are displayed – for example, in large or small text, red or blue, boxed or unboxed, on the left or right.

None of these elements is unproblematic: each is complex, may overlap to some extent with the other stages, and presents many possibilities for intentional and unintentional intervention by interested parties. Below I describe each element in detail.

The Index. The search index must be populated, or filled with data, either by human editors or by an automated or semi-automated process, called a crawler. Editors were more common in the early history of search; for example, when AskJeeves launched it had human editors, as did LookSmart which functioned as MSN Search's index for some time. Normally, the key quality measures for the index are *size* (how many documents it contains) and *freshness* (how recently the links to the documents were checked).

The index must also be as "clean" (Interview B) as possible, or free of duplicated content or content deemed unsuitable or "malicious" (Google, 2004b) – for example, spam. A team of engineers, sometimes called content management engineers, is in charge of programming the crawlers so that they clean the data as they are collected, and then once again before it is added to the main index. Cleaning techniques include blacklisting known spammers and known spam sites, as well as implementing various heuristics that look for unusual patterns. Simple examples of cleaning heuristics would include disregarding black text on a black background, or font sizes too small to be read; complex examples would include checking document text for word patterns which do not conform with standard word patterns, or checking for outbound links to known spam sites (for further discussion of anti-spam heuristics, see Fetterly, Manasse, & Najork, 2004; Henzinger, Motwani, & Silverstein, 2002).

The index often has multiple partitions for international searching so that the search algorithm might run against the UK partition, or the German partition. Thus the size, freshness, and hygiene of the partitions may be different, so that searches might be poor for certain countries and good for others (Vaughan & Thelwall, 2004).

One potential cause of bias in a search engine is the inclusion or exclusion of particular websites in the index (Machill, Neuberger, & Schindler, 2003). There are many reasons for exclusion of particular pages, websites, or domains. Google, for example, offers the following potential reasons for exclusion to concerned Web site managers (Google, 2004b): 1) the pages are dynamically generated; 2) the pages are doorway pages (that is, pages which simply forward the user on to another page); 3) the pages use frames; 3) the site might not have enough links to it from other sites; 5) the pages might change between different versions of the index (updated about every four weeks); 6) the crawler might not have been able to reach the page because of network problems; 7) temporary technical faults in Google's crawler; and 8) other reasons, including a manual removal because the page did not conform to "quality standards". In explaining this last, Google says:

> "We will not comment on the individual reasons a page was removed and we do not offer an exhaustive list of practices that can cause removal. However, certain actions such as cloaking, writing text that can be seen by search engines but not by users, or setting up pages/links

with the sole purpose of fooling search engines may result in permanent removal from our index." (Google, 2004b)

There is also a secondary index for advertisements, where inclusion is based on payment but additionally on certain other kinds of rules, such a limited amount of capital letters or a certification that one owns the trademark one is attempting to place in one's ad.

The Search Algorithm. The search algorithm, or set of computer logic controlling the search engine queries, is often known as the "core search technology." Put simply, it matches answers to the queries that users type in. It may be divided into two parts: the parser and the ranker. First, the *parser* reads and "understands" the query based on logic to do with language. For example, the German word *Geschwindigkeitsbeschränkung* (speed limit), must be split into its component elements before being matched to the content of the index. Similarly, the French adjectives *beau* and *belle* (beautiful) must be recognized as variants of the same word by the parser. Second, the *ranker* decides which match should come first, second, etc., in the list of results. Each ranker is slightly different, and all of them are secret; however, the ranker operates both with document contents and with the position of the document relative to the whole Web. Document contents include such items as page titles, page text and keywords embedded in the source code of the Web pages. The second element, the position of the website relative to the Web as a whole, is the key to the modern search engine and was the original technical innovation which separated Google from its competition. The position of the website is determined by the number of links to the page from other sites ("inlinks") and the number of links to from the page to other sites ("outlinks"). Sites with a high inlink density are called "authorities," whereas sites with a high outlink density are called "hubs," terminology developed by the mathematician Kleinberg (1998). There are two major known algorithms to assess this position: Kleinberg's HITS algorithm, and Google's PageRank (Brin & Page, 1998). These algorithms have developed since publication in 1998; the trend seems to be to assess groups of sites, or communities, relative to each other (see for example Fujimura, Inoue, & Sugisaki, 2005).

The search algorithm is typically assessed by the *relevance* quality metric (see Chapter 6)

The secondary index of advertisements uses a different and much simpler search algorithm: the advertisers themselves decide what user queries they wish to be related to,

and the minimum position of their results (number 1 may be considerably more expensive than number 3 on the list). However, some search engines, like Google, also rank advertisements based on the number of users who click the ads.

The Front End. The front end comprises the design of the results. "We see the front end as being us" says one major search distributor. That is to say, the front end can easily be managed by a different team, or even a different company, than the index and the search algorithm.[64]

The front end is what we see. The front end of Google's search – the search engine results page or SERP – is a list of links with ads down the side (received from the advertising search algorithm); the AskJeeves SERP is a series of questions with options in a drop-down list.

Front end choices respond to aesthetic concerns, users' functional needs, and business and legal constraints. Measures are not normally calculated with precision although accessibility (for example, for partially-sighted users) and usability are normally issues for the front-end design teams. The balance of advertising and "editorial" content such as search engine results is a front-end issue, although the decision is normally not taken by the designers, but by the product manager.

Organizational structure. Search overall is normally managed by a "product manager" – someone whose function is primarily to ensure that the product is "competitive" by balancing business needs (costs, revenues, and overall business priorities) with technical possibilities and constraints and an assessment of user needs. The search product manager is rather like the editor of a newspaper, responsible for the overall quality of the search. Thus, the product managers normally will have both engineers and user interface specialists reporting to them. Advertising sales will not normally report to the product manager, but will have some claim on the way that search is developed – for example, spaces will be left for advertising messages in the overall design. In some organisations, there is a separate product manager who manages the keyword search advertising product. In other organizations, this may be outsourced and purchased in. In either case, the

---

[64] Google, for example, publishes an "API" (application programming interface), which allows programmers to bypass Google's front end and put results from the search algorithm directly into their own programs.

search product manager must negotiate with others how advertising appears in the search product (see Chapter 6).

Appendix B: Literature Map

Search Engines

Programming
- Specialist engines
  - Scientific
    - Genes
    - Molecules
    - Algorithms
    - Sensors
    - Earth science
    - Astronomy
  - Single-site
    - Intranet
  - Images, multimedia
  - Within programs
  - Professional
    - Legal
    - Educational
  - News
  - Library
  - XML schema
- Improving
  - Queries
    - Categorisation
      - Metadata
      - AI/Learning
      - By users
      - Automatic
      - Hierarchic
    - Expansion
      - Query mining
      - Keyword comparison
      - Fuzzy matching
      - Term suggestion
    - Spell-checking
    - Personalised
    - Contextual
      - Based on user groups
    - Browse
      - Image
    - Metaengine
    - Avatar
    - Translation
    - Local
      - Alternate script
  - Results
    - Alternative display
      - 3D
      - Voice via avatar
    - Categorisation
    - Pie charts
    - Ranking
  - Index
    - Mirror awareness
    - Filtering
    - 'semantic' v 'document'
    - Metadata inclusion
    - Clustering
    - 'knowledge'-based
    - Relevant document data
  - Operations
    - Compression
    - Storage and retrieval
    - Server load
    - Cache
  - Architecture
    - XML
    - Agents
    - Distributed
    - Data model
    - Algorithms
      - Genetic
      - Evolutionary
      - Sequential
      - Link-structure (Kleinberg)
    - Object collector
  - Process
    - Interdisciplinarity
  - Mathematics
    - Graph theory
    - Set theory
    - Fuzzy sets

Role — 'make accessible' — Power

Industry
- History
- Partnerships — Government

Use
- Sampling the Web
  - Medical
  - Research — Recruitment
- Promotion — Page structure
  - Metadata
  - Images
- Deliver relevant links — Metadata policy
- Intelligence — Commercial availability of hazards

Evaluation
- With people
  - Undergraduates
  - Children
  - Library patrons
  - 'users'
  - Master students
    - Cognitive styles
    - Level of experience
    - Internet perceptions
    - Study approaches
    - Age
    - Gender
    - Search strategies
  - 'Expert' searchers
- Versus criteria
  - Info correctness
  - Replicability
  - Site duplication
  - Dead links
  - Availability
  - Response time
  - Number of sites
  - Time to site
  - Satisfaction
  - Precision
  - No of successes
    - Recall
    - Precision
  - Index quality
    - Random
    - Representative
- Against each other
  - Specialist v general — Medical
  - Directory v engine
  - Pay v no-pay
  - Index quality
  - Mobile v traditional
- Via query logs
- With different tasks
  - Complexity
  - Image searching
- With different page types — Metadata v non metadata
- Versus offline promo

C:\Documents and Settings\E Van Couvering\My Documents\Research\Search Engines.mmp - 28/08/2003 - Elizabeth Van Couvering - e.j.van-couvering@lse.ac

Appendix C: Initial Approach Letter

[LSE Letterhead used]
        [Date]

[Name and address removed]

Via fax : [number removed]

Dear [Name removed],

I am a student in the third year of my PhD at the London School of Economics. The topic of my research is search engines, and for one of my chapters I am reviewing the history of the industry to date. As you were intimately involved in that history, I would be very interested in interviewing you for my research, and wonder if you would be able to spare me an hour in the coming weeks for a telephone interview. This letter is a follow-up to the message I left on your voicemail earlier this morning; I am sending it via fax as the switchboard is instructed not to release your email address. I am sure that you are very busy, and I'm very happy to accommodate your schedule.

That's the nuts and bolts of it, on the following page I describe more about my research. Please do let me know if you have any questions I don't address below.

Best regards,

Elizabeth Van Couvering
PhD Student
Department of Media & Communications
Email: e.j.van-couvering@lse.ac.uk
mobile: +44 (7768) 890 382 (GMT)

ABOUT MY RESEARCH

I am conducting research into the search engine industry. Search engines are a major force in new media but there has been very little academic work except from a computer engineering standpoint. I am situated in the media and communications department, and my approach is media-related rather than technology based. Recent studies from computer science have indicated that search engine results are "biased" - for example, all other things being equal, they are more likely to favour large, commercial, American websites. I am interested in the variety of reasons - technical, economic, historical - as to why that might be so. This can be compared to the many production studies of newspapers which have sought to understand how journalists operate as they do and how the wider context of newspaper publishing might affect particular journalists.

ABOUT THE INTERVIEW

The interview can be, at your choice, entirely anonymous, with your name and identifying details disguised if I quote you directly or indirectly. I would like to record the interview to aid my analysis later, if that is OK with you. Topics I would like to cover are:

1 - your general background, how you got into the industry, etc
2 - how things worked at [company name removed] (this is for purposes of comparison with other engines, and to see how things have changed). In particular, who was involved in making decisions about search engine results, combatting spam, etc; also how advertising worked and related to the engineering and product areas (if you called them that...)
3 - how you think things have changed - how the industry currently is different than you recall

ABOUT ME

I am a third-year PhD student. Prior to starting at the LSE I worked for 12 years in marketing, 8 in the commercial internet sector in the UK, including one year at Excite as Marketing & Communities manager for Europe (1998). There is more about me at my home page: http://personal.lse.ac.uk/vancouve/

WHAT THE INTERVIEW WILL BE USED FOR

The interview will form a part of the evidence for my doctoral thesis, which I intend to submit in 2006. It may also be used for other academic publications either before or after that date..

Appendix D: Interview Schedule

- Reiterate aims of project, right of refusal, etc.
- Can you tell me how you came to be at [company name], and what's your job there?
- What does that entail?
- Can you tell me who else is involved?
- Can you tell me about a particular change in the search engine? Who instigated that change and why? How was it carried out? How was it evaluated?
- How does search interrelate with the rest of the [company] website?
- How does search interrelate to advertising? To product? To website design?
- What is different about the process in US search versus search in other countries?
- How has search changed in the years since you've been in it? Do you think it is better or worse?
- Do you have anything else you'd like to tell me?
- Can you recommend someone else I can talk to?
- Thanks

Appendix E: The Significance of Search Engine Bias

What role do search engines play for ordinary people using the Web? What does it mean to suggest, as some have done[65], to say that search engines are biased? Might it mean that access to the internet is somehow fundamentally compromised? Or, on the other hand, might it mean very little to the average user?

To answer these questions, we need to know a little more about how people navigate the Web overall. While there are a great many studies which deal with navigation behaviour within individual websites, there are comparatively few that address online behaviour holistically across across a range of websites and tools. This Appendix synthesises a range of empirical studies of searching and browsing behaviour.

In general, the empirical studies that we have draw from three sources. These are, in order from least to most specific: user tracking studies, task-based studies, and log-mining studies. User tracking studies record user interaction with the Web for shorter or longer periods of time and then categorize the behaviour they observe. Task-based analyses set a group of users (often college students) specific search tasks such as finding local movie showtimes or contrasting theories on climate change, often in order to determine either the skill base of the user or the usability of the search interface. Log-mining studies use statistical analysis of huge query log databases provided by the search engines to gain insight into user searching strategies.

The three approaches to understanding search engines are not a unified field of work, although they do share some common assumptions.[66] Some conclusions about how

---

65 In particular, studies have shown that search engine results pages represent only a portion of the Web and over-represent popular sites, commercial sites and American sites (Bergman, 2001; Kleinberg, 1998; Kleinberg & Lawrence, 2001; Lawrence & Giles, 1999; Vaughan, 2004; Vaughan & Thelwall, 2004). This research is analysed more fully in a separate section of the thesis.

66 Most of these studies are working within an information-retrieval paradigm; thus they are applying or developing theories of learning where information "gaps" are crossed by a variety of "sense-making" activities, which are often operationalised as skills (e.g., Savolainen, 2006). Thus one can succeed or fail at a given search task within the paradigm of these studies. This is in marked contrast to media studies literatures where media texts are "interpreted" to give them "meaning" within daily life, and success and failure are not appropriate categories.

people interact with Web content and an assessment of the role of search engines within that overall interaction are presented below.

Overall patterns in online navigation

The most comprehensive data for understanding Web navigation comes from a sophisticated French study (Beauvisage, 2004). This study tracked the browsing behaviour of three separate user panels over a substantial period of time, by installing a tracking device on their computer which examined not only Web activity but other internet applications as well (such as email and chat). The first and largest panel consisted of 3,372 representative French internet users whose Web activity was observed between January and October 2002 (10 months). The second panel consisted of a longitudinal study of 597 representative French internet users between 2000 and 2002 (34 months). The final panel was composed of 72 users of electronic libraries whose Web activity was observed between April and December 2002 (6 months), and who were also surveyed about their internet usage in general. 16 members of this final panel were also interviewed in depth about their Web and library usage as well as their online activities.

The recorded behaviour from all the panels was divided into sessions (periods of activity separated by more than 30 minutes of inactivity), and all pages in a session were further categorised by content which was derived automatically from a combination of two online directories (Yahoo! and Nomade) with the addition of an additional category for sexually explicit content. The sites were then analysed by correlating the number of times they were viewed and the timespan between viewing.

On the basis of this study, each user's experience of the Web seems to be quite different, as judged by the sites they look at and the tools they use. Beauvisage reports that in his 34 month-panel, participants consulted over 192,000 sites. Two-thirds (66.3%) of these sites were seen by only a single user; a quarter by only a handful of users (25.8%), and less than a tenth (7.7%) by between 6 and 99 people. Only 0.2% of these sites were seen by more than 100 research participants - as seen in Figure 32:

Figure 32: Number of sites sharing visitors over 34 months
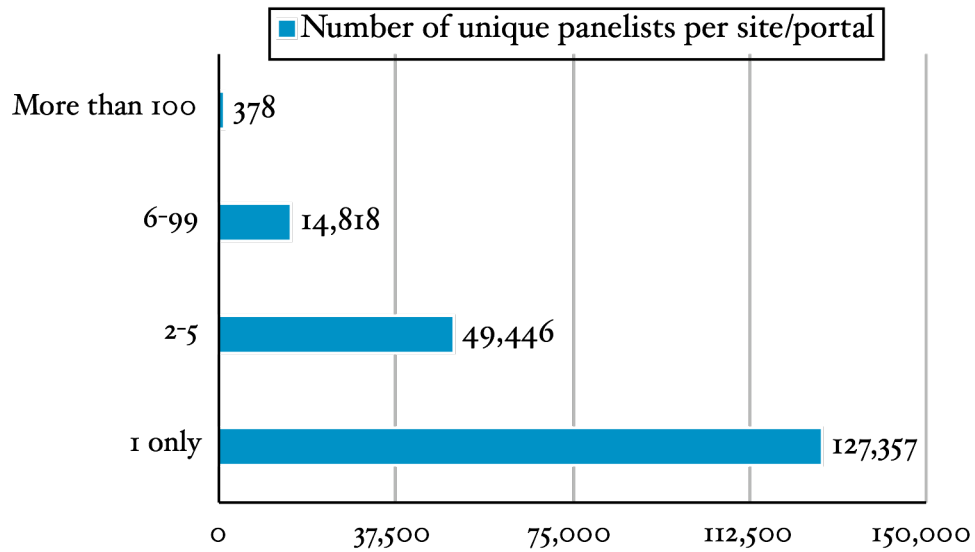
Source: adapted from Beauvisage, 2004: 243, panelist n=597

The 378 sites viewed by more than one panelist included general portal and search engine sites (MSN, Wanadoo, Yahoo and Club Internet).

The great diversity of Web usage that Beauvisage reports seems to be replicated even when users are quite homogeneous, as we see in another study of 17 members of a computer science department whose Web behaviour was observed over four months (October 1999 - January 2000) (Cockburn & McKenzie, 2001). In this group, the total Web "vocabulary" was 17,242 distinct URLs. However, only 9.2% of these URLs were visited by more than one person, and only .52% (89 URLs) were visited by eight or more members of the department. No single page was visited by all the members in the group during the observation period. A further long-term study which analysed the Web transaction logs of 206 students of higher education over 10 months during the 1997/98 academic year (Cothey, 2002) also found that as the students increased their experience of browsing the Web, their tastes became more eclectic and they accessed fewer of the same sites as their peers.

Cothey's study leads on to a second finding of the Beauvisage study: we can also say that people have a small range of subjects in which they are interested. The least active

quartile of users in the 34-month panel, for example, viewed on average sites from only only five different categories of content[67], while the most active quartile viewed content from slightly more than eighteen categories (Beauvisage, 2004, p. 249). For the library usage panel, the average session was composed of pages from 15 directory categories, but only 5 categories were reflected in more than 5% of sessions.

THE HOME RANGE

Beauvisage uses the metaphor of "territory" to describe Web usage patterns, noting that Web users stake out their own territory online, a very small set of familiar, comfortable sites at which they spend most of their online time, and which may be said to be their online *home range,* divided into a central home range and an extended home range. Outside the home range there is a second tier of *activity-based sites* which are visited several times over a very short period, and then typically abandoned and never visited again. Finally, there is a third tier of websites which are associated with *exploration* and are typically visited only once (Beauvisage, 2004, p. 247). Exploratory Web sites comprise over 75% of the total popularion of Web sites visited. Figure 33 represents these three tiers in graphical form.

[67] Based on Yahoo! and Nomade categorisations, plus a category for sexually explicit material.

Figure 33: A model of Web territories
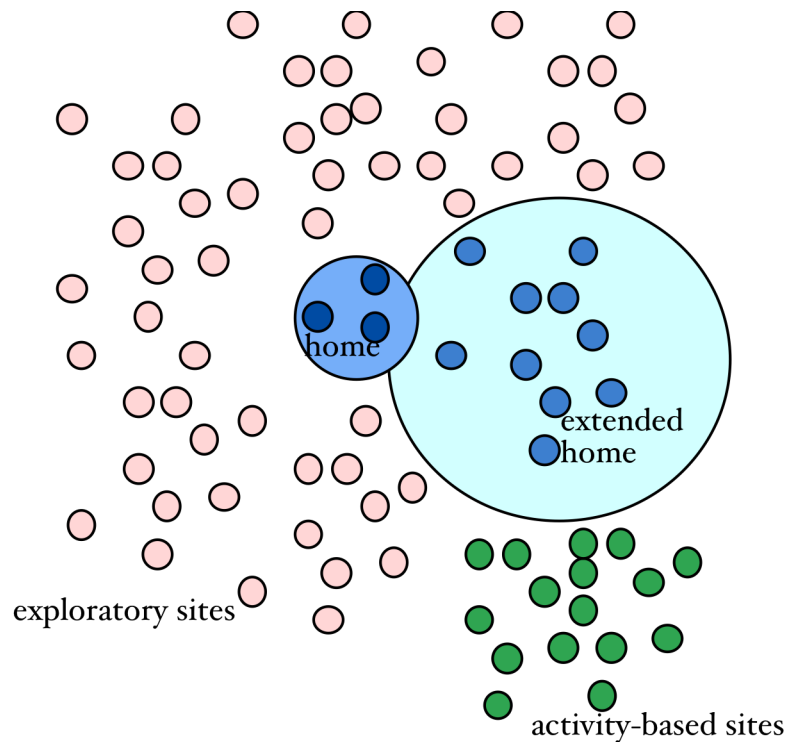
The home range and extended home range comprise the bulk of the users' time online. Activity-based sites are viewed frequently for a short time and then discarded. Exploratory sites are visited only once.

Source: Author. Figure based on data from Beauvisage, 2004.

The home range, the core of the user's territory according to Beauvisage, is a set of a few pages which are visited consistently over a long period of time. These sites form the bulk of the user's interaction with Web content. Web sites visited consistently over a period of more than a year were contained in 91% of the sessions in the three-year panel (Beauvisage, 2004, p. 252). Some sites in this category are also visited very frequently during this time – these sites form the *central home range*. In this category are generalist portals and search engines, Web applications such as Web-based email, and also specialist sites on topics of interest (employer home pages, forums, movie listings, etc.) These sites number less than five in half the cases and are rarely augmented (Beauvisage, 2004, p. 247).

Outside the daily territory of the central home range, the *extended home range* is set of pages which are visited repeatedly but only occasionally - for example, pages that are useful when conducting a particular task, such as shopping or checking train times (Beauvisage,

2004, p. 257). These pages may be bookmarked, or they may be found through tracing an already-known pattern of links. Another study that investigated "re-finding" behaviour in a small sample of US undergraduate students suggested that during occasional tasks "users might recall or recognise *waypoints* – important or memorable Web sites along the path they took when they found the information the first time" (Capra & Pérez-Quiñones, 2005, p. 38, emphasis original). A second, smaller study conducted amongst 10 members of the Carnegie-Mellon University community reinforces the sense of territory: "Hunts targeted at specific pages tended to start at a known location. This provides strong evidence that users know something about where the pages they frequent lead... Furthermore, people did not seem to use search engines to browse for things of interest, but instead use them for more targeted searches" (Byrne, John, & Joyce, 1999, pp. 27-28).

Sites in the extended home range do increase as users try sites in their excursions of discovery, but the number is still quite small: a mean of 72 sites, with a minimum of 16 and a maximum of over a hundred (Beauvisage, 2004, p. 247). In terms of overall usage, in the 34-month panel these home range sites represented only 7.5% of the total sites viewed – a figure that radically understates their importance, for in terms of duration, home range sites sites, both central and extended, take up most of a user's online time in 70.1% of sessions (Beauvisage, 2004, p. 253).

While Beauvisage has the most comprehensive data, other studies back up his conclusions about the home range. Cokburn & McKenzie (2001) also found the same overwhelming revisitation of favourite pages in their 2-week observation, as the top three pages for each user accounted for 24% of the total visits (ranging from 8.9% for one user up to 48% for another).

Routine, occasional, and discovery sessions

Beauvisage further statistically clustered not just websites but user sessions, using three variables: the frequency of visit, the timespan over which the user visits the site, and the rarity of the content type. He identified three major clusters which he calls "routine" sessions (51.6% of all sessions), "occasional" sessions (36.5%), and "discovery" sessions (11.9%). In routine sessions, the user visits generalist portals, messaging services, news sites (either on portals or in online newspaper), or financial sites. They may occasionally

follow a link out from a story or use a search engine to follow up on something (in 12% of "routine" sessions") but tend to remain within their central home range. In essence, during these sessions the user keeps updated on content which changes frequently. Beauvisage says:

> "...this type of content reception is not far different from that of traditional media: once the user has identified the channels which interest them, they will 'consume' the content on this basis, as they do with television or radio programmes, or newspapers and magazines. The difference here is on the one hand is in the availability of Web content and the possibility of getting it at the user's convenience, and on the other hand the way in which this routine content opens the way to the unfamiliar via hypertext." (Beauvisage, 2004, p. 260, my translation)

In "occasional" sessions, the user goes to sites in the extended home range, sometimes (in a third of "occasional" sessions) using a search engine, normally to help find a site whose address has been forgotten. These sessions often include a relation to offline activities, such as cultural activities or shopping. Sometimes a session like this will include multiple windows, in which different sites - for example, air travel sites - might be compared.

> "In this type of path, the understanding of Web content is close to a 'ticket office,' a 'yellow pages,' or a 'shopping mall': the included activities are relatively rare for each user, who nonetheless adopts a relatively stable behaviour in this context, and often visits the same site or group of sites." (Beauvisage, 2004, p. 262, my translation).

Finally, in the "discovery" sessions, the user is seeking unknown information and expanding their territory. Search engines are key to this kind of session and are present in 60% of all discovery sessions. Beauvisage distinguishes discovery sessions between the "predator" and the "browser" (*flâneur*). The predator seeks specific information and goes rapidly through a range of sources until they get a result. For a browser, the query is often more open, pages are visited more slowly and digressions are more numerous. This type of session is particularly related to viewing sexually explicit pictures – the task being not to find a specific photo but an appealing and new site. Hence, during the discovery sessions:

> "[T]he Web is considered not so much as a space of documents but rather as a collection of possible sources at the centre of which one may find the most complete and the most reliable for a current investigation." (Beauvisage, 2004, p. 267, my translation)

Navigation and search

Search engines seem to occupy a complex place in Web navigation. They are rarely used in an everyday session, although they may be part of a portal or ISP site which is scanned for news or entertainment content and may supply an outward link for following up a topic of interest. They serve as an online *aide-memoire* or "waypoint" when it comes to occasional needs, supplementing or replacing bookmark files or other lists of websites. But they are absolutely central when it comes to acquainting oneself with a topic, seeking a new point of view or new source of entertainment, or researching a specific fact or piece of information.

The next section reviews the interaction between search engine and user.

Search engine usage

According to Beauvisage (2004, p. 302), search engines were included in one-fifth of user sessions and search engine results pages comprised 1.3% of total overall pages viewed. 85% of his panel used a search engine during the period of study, a figure that is comparable to the 85% that the Pew Internet reports for US internet users (Fallows, 2005). Amongst those not using search engines according to Beauvisage, children under 15 and women were prevalent, and there was a very strong correlation between overall intensity of internet usage and intensity and diversity of search engine usage. That is to say, those who used the internet rarely relied on a single search engine, whereas intense users used multiple search engines.[68]

It seems that internet usage and skill with search engines are related. A Swedish study found that compared to expert internet users, novice users are highly unlikely to enter the URL of a search engine when asked to complete a search task, clicking instead on the 'search' link provided by the browser. Instead of examining pages from the resulting lists of links, as experts did, they instead retried the search (Hölscher & Strube, 2000). Those novices who lacked experience with the topic they were investigating also made "only small and ineffective changes to their queries, forcing them to reiterate repeatedly." This study also highlights the importance of what might be termed a flexible search repertoire,

---

[68] At this time, Google does not seem to have been available in France. Search engines used included AltaVista, Voilà, Wanadoo and Yahoo, among others.

saying "It is not fully clear, if novices browse less useful material than the experts, but once they face a dead end their only way out is to go backwards, while experts have more flexible ways of reacting." Novice users also appear to be entirely reactive to what they see on the screen, whereas expert users plan ahead (Navarro-Prieto, Scaife, & Rogers, 1999).

It may be that some users, at least, cling to their home range because of difficulties in navigational skills. This hypothesis is given some support by another US study. Hargittai (2002) studied a random sample of internet users in Bergen County, New Jersey. She found that "the general user population lack the basics of surfing the Web." Common among her sample was extensive use of the Back button, using only links and browser functions to navigate (i.e., not using search terms or URLs) and having difficulty entering valid terms into search engines because of common mistakes, including poor spelling and the entry of search terms without spaces between them. Other users were unable to differentiate between the location bar and the search engine search field.

Capra and Pérez-Quiñones (2005) also suggest that information finding strategies are strongly related to the familiarity of the task. They propose a four quadrant model based on dimensions of familiarity of the topic and frequency of the search, which leads to different user behaviours (see Figure 34).

Figure 34: Matrix of finding and refinding behaviours

familiarity

II: Refinding with access patterns +	+ III: Refinding with shortcuts -
  I: Finding	IV: Inherently - difficult tasks

(row label, vertical: **frequency**)

Source: Adapted from Capra & Pérez-Quiñones, 2005, p. 41.

Quadrant I contains finding behaviour for information unfamiliar to the user and for which they have never searched - search engines would be used along with other information "foraging" strategies in this quadrant. Quadrant II contains familiar but infrequent searches which the user finds though known links and other "access patterns". Quadrant III contains familiar and frequent tasks, for which the user uses shortcuts - bookmarks, links, or URL typing. Quadrant IV contains frequent and unfamiliar tasks, probably tasks the user cannot yet perform proficiently or which are inherently difficult. In this model, search engines are probably used most highly in Quadrants I and II, and possibly IV.

These studies, taken together, also suggest that search proficiency builds on itself: as users gain skills, they add to their repertoire of familiar tasks and Web sites which, in turn, helps them to gain new skills.

It is interesting to note data from log-mining studies: for those users who do use search engines, most use simple searches, with an average of two terms per query, two queries per session, typically not using complex query syntax and viewing no more than ten documents from the results list (Jansen & Pooch, 2000). On the basis of a large sample of

internet queries (over 1 million) another study reports that searchers rarely go beyond the first page of results (H. C. Ozmutlu, Spink, & Ozmutlu, 2002; Spink, Wolfram, Jansen, & Saracevic, 2001). If users are formulating extremely targeted searches for Quadrant II or III activities, then some of this behaviour could be explained not through lack of skill but on the contrary precisely the opposite: a skilled and speedy location of known resources.

Users seem broadly satisfied with search services. A memo from the Pew Internet project in the US says that 87% of Americans who use search engines find the information they are seeking most of the time. 32% said they could not live without them. And 68% said they thought internet search engines were a fair and unbiased source of information. 92% said they were confident in their use of search engines (Rainie & Shermak, 2005). A German study found 66% of a random sample of internet users considered themselves "advanced" or "expert" at search engine use (Machill, Neuberger, Schweiger, & Wirth, 2004). There is little similar data for other countries, though it might be reasonable to assume that US searchers would be more satisfied as most search engines are US-based and contain more features and more local content in their US versions.

A series of studies has linked changes in search behaviour to poor design of search engine pages and search engine algorithms (Hsieh-Yee, 2001). A large study of European search queries (over 1 million queries) reported that only 50% of pages viewed as a result of a search were topically relevant (Jansen & Spink). And Hargittai's study, reported earlier, suggests that perhaps people say they are more expert than they really are.

The Implications of Search Engine Bias

It is possible to overstate the importance of the search engine for the average user. For example, although search engines are the largest Web sites online in terms of traffic, the search function does not comprise the bulk of an ordinary users' Web traffic which, instead, is focused on a narrow home range of sites with constantly changing information.[69] Nevertheless, search engines do have several critical functions in the ordinary online behaviour of internet users:

---

[69] However, it is important to realise that the "news" functions of large portal-search sites like Yahoo! are likely to be included in the central home range of many users.

1. They are essential in discovering new information, not necessarily timely information, but information which is new to the user. They are particularly important when time is critical and any "expert" usage of the Web implies a concomitant expertise in using search functions – for example, search skills are key components of information literacy (Livingstone, Van Couvering, & Thumim, 2004).

2. They are essential in discovering new Web sites which may be added to the extended home range, particularly small sites that are unlikely to have many linkages and therefore would be relatively inaccessible through browsing/linking behaviour.

3. They are important for re-finding old information which users may have misplaced.

4. Bias in search engines might therefore affect inexperienced and unskilled users most strongly. It might also be expected to affect expert searches where time is of the essence, for example for journalists using the Web for research while writing to deadline.

It might also adversely affect people's ability to find small websites. That is, while people might know how to find giants of online retail like Amazon.com, small websites that are off the beaten track rely on search engines to ensure that their content is accessible (Fortunato, Flammini, Menczer, & Vespignani, 2006). In this context it is particularly ironic that one of the elements of search engine bias is that search engines rank websites by giving higher weight to more popular and more highly-linked sites.

It seems likely that a bias in search engines might easily go undetected and be relatively invisible to the average user, given that its most noticeable effects would appear when relatively unskilled users are looking for unfamiliar sites. Nonetheless, the effect might be profound in the long term as people's attention is drawn consistently to mainstream, commercial websites as they search for new unfamiliar information under the pressure of time.

References

Abondance. (2002, 26 February). Transcript du chat avec Matt Cutts et Stephanie Kerebel, Google. Retrieved 1 March, 2004, from http://chat.abondance.com/google.html

AFX News. (2004, 21 June). Yahoo sees limited revenue from new China search engine Yisou in short term. *AFX.com* Retrieved 24 May, 2006, from Lexis Nexis Executive database.

Alexa. (2006). Top sites South Africa. *Traffic Rankings.* Retrieved 4 September, 2006, from http://www.alexa.com/site/ds/top_sites?cc=ZA&ts_mode=country&lang=none

Aljifri, H., & Sanchez Navarro, D. (2004). Search engines and privacy. *Computers & Security, 23*(5), 379-388.

Amnesty International. (2006). *Undermining freedom of expression in China: The role of Yahoo!, Microsoft and Google.* London.

Ang, I. (1985). *Watching Dallas : soap opera and the melodramatic imagination.* London: Methuen.

Ang, I. (1996). *Living room wars: rethinking media audiences for a postmodern world.* London and New York: Routledge.

Arrington, M. (2008, 29 February). Technorati to launch blogger advertising network. *TechCrunch* Retrieved 26 April, 2010, from http://techcrunch.com/2008/02/29/technorati-to-launch-blogger-advertising-network/

Asia Pulse. (2005, 23 June). Microsoft Japan to revamp search functions. Retrieved 5 June, 2006, from Lexis Nexis Executive database.

Ask Jeeves. (2004, 24 August). Ask Jeeves launches beta version of ask.jp. *PR Newswire* Retrieved 15 June, 2006, from from Lexis-Nexis Executive database

Associated Press. (2005, 19 December). Japanese government considers launching Internet search engine. *MercuryNews.com* Retrieved 9 June, 2006, from http://www.siliconvalley.com/mld/siliconvalley/business/technology/13443136.htm

Bar-Ilan, J. (2000). Evaluating the stability of the search tools Hotbot and Snap: A case study. *Online and CDROM Review, 24*, 439-449.

Bar-Ilan, J., & Gutman, T. (2005). How do search engines respond to some non-English queries? *Journal of Information Science, 31*, 13-28.

Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture.* London: Nicholas Brealey.

Battelle, J. (2008, 3 June). Wikia Announces Hackable (ie Open) Search. *John Battelle's Searchblog* Retrieved 7 May, 2009, from http://battellemedia.com/archives/004479.php

Beauvisage, T. (2004). *Sémantique des parcours des utilisateurs sur le Web [Semantics of user paths on the Web].* Unpublished PhD, Université de Paris X - Nanterre, Paris.

Bergman, M. K. (2001). White Paper: The deep Web: Surfacing hidden value. *Journal of Electronic Publishing, 7*(1).

Bermejo, F. (2009). Audience manufacture in historical perspective: from broadcasting to Google. *New Media & Society, 11*(1&2), 133-154.

Bijker, W. E. (1989). The social construction of Bakelite: toward a theory of invention. In W. E. Bijker, T. P. Hughes & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 159-187). The MIT Press: Cambridge, MA.

Bijker, W. E. (1995). *Of bicycles, bakelites, and bulbs : toward a theory of sociotechnical change.* Cambridge, MA: MIT Press.

Bijker, W. E., Hughes, T. P., & Pinch, T. (1989). *The social construction of technological systems: New directions in the sociology and history of technology.* Cambridge, MA and London: The MIT Press.

Blevins, J. L. (2004). Battle of the online brands: Disney loses the portal war. *Television & New Media, 5*(3), 247-271.

Bowker, G. C., & Leigh Star, S. (2002). *Sorting things out: Classification and its consequences.* Cambridge, MA and London: The MIT Press.

boyd, d. m., & Ellison, N. B. (2008). Social network sites: Definition, history, and scholarship. *Journal of Computer Mediated Communication, 13*, 210-230.

Boyd-Barrett, O. (1998). Media imperialism reformulated. In D. K. Thussu (Ed.), *Electronic empires: Global media and local resistance* (pp. 157-176). London: Arnold.

Breed, W. (1999). Social control in the newsroom: A functional analysis. In H. Tumber (Ed.), *News: A reader* (pp. 79-84). Oxford: Oxford University Press.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and Isdn Systems, 30*(1-7), 107-117.

Brown, L. E., Dubois, A., & Shepard, D. B. (2008). Inefficiency and bias of search engines in retrieving references containing scientific names of fossil amphibians. *Bulletin of Science, Technology & Society, 28*(4), 279-288.

Burbridge, M. (2006). Naspers ready for internet boom with 24.com. *Mail & Guardian Online.* Retrieved 4 September, 2006, from http://www.mg.co.za/articlePage.aspx?articleid=278738&area=/insight/insight_tech/

Business Day. (2006, 15 August). Google searches for new SA office staff. *Africa News* Retrieved 4 September, 1006, from Lexis Nexis executive database.

Buzan, T. (1982). *Use your head* (Rev. and extended ed.). London: Ariel.

Byrne, M. D., John, B. E., & Joyce, E. (1999). A day in the life of ten WWW users. Retrieved 4 May, 2005, from http://chil.rice.edu/bryne/Pubs/byrneJohnWeb.pdf

Callon, M. (1986). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. In J. Law (Ed.), *Power, action, and belief: A new sociology of knowledge?* London, Boston and Henley: Routledge and Kegan Paul.

Callon, M. (1987). Society in the making: The study of technology as a tool for sociological analysis. In W. E. Bijker, T. P. Hughes & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 83-103). Cambridge, MA and London: MIT Press.

Cantor, J. M., & Cantor, M. G. (1992). *Prime-time television: content and control* (2nd ed.). London: Sage.

Capra, R. G., & Pérez-Quiñones, M. A. (2005). Using web search engines to find and refind information. *Computer Networks and ISDN Systems, 38*(10), 36-42.

Carlson, M. (2007). Order versus access: news search engines and the challenge to traditional journalistic roles. *Media, Culture & Society, 29*(6), 1014-1030.

Carnegie Mellon University. (1995, 20 June). Pittsburgh company established using Lycos internet catalog technology. *News Releases* Retrieved 10 March, 2006, from http://www.cs.cmu.edu/~scsnews/jun20-95.html

Castells, M. (1996). *The information age: economy, society, and culture: The rise of the network society* (Vol. 1). Oxford: Blackwell.

Castells, M. (2001). *The Internet galaxy : reflections on the Internet, business, and society.* Oxford ; New York: Oxford University Press.

Cerf, V. G. (2006, 7 Feburary). Prepared statement of Vinton G. Cerf, Vice President and Chief Internet Evangelist, Google Inc., to the U.S. Senate Committeed on Commerce, Science and Transportation Hearing on "Network Neutrality". Retrieved 12 March, 2008, from http://commerce.senate.gov/pdf/cerf-020706.pdf

China Internet Network Information Center. (2006, January). The internet development in China: 17th statistical survey report. *Statistical Reports on the Internet Development in China* Retrieved 13 May, 2006, from http://www.cnnic.cn/download/2006/17threport-en.pdf

Choros, K. (2005). Testing the effectiveness of retrieval to queries using Polish words with diacritics. *Lecture Notes in Artificial Intelligence, 3528*, 101-106.

Clark, T. (1998, 24 November). AOL buys Netscape for $4.2 billion. *c|Net News.com* Retrieved 21 August, 2006, from http://news.com.com/AOL+buys+Netscape+for+4.2+billion/2100-1023_3-218360.html

CNNMoney. (1998a, 18 June). Disney buys Infoseek stake: Swaps Starwave and $70M for 43% interest; will take option for control. *CNNMoney News - Deals* Retrieved 17 March, 2006, from http://money.cnn.com/1998/06/18/deals/infoseek

CNNMoney. (1998b, 9 June). NBC enters a Web portal: Broadcaster buys 5 percent stake in CNET, forms venture to run Snap! *CNNMoney News - Technology* Retrieved 17 March, 2006, from http://money.cnn.com/1998/06/09/technology/cnet/

CNNMoney. (1999, 12 July). Disney absorbs Infoseek: Entertainment giant, Internet search engine to expand online partnership. *CNN Money News - Deals* Retrieved 17 March, 2006, from http://money.cnn.com/1999/07/12/deals/disney

Coates, K., & Holroyd, C. (2003). *Japan and the internet revolution.* Basingstoke, Hampshire: Palgrave Macmillan.

Cockburn, A., & McKenzie, B. (2001). What do web users do? An empirical analysis of web use. *International Journal of Human-Computer Studies, 54*(6), 903-922.

comScore Networks. (2006, 4 May). 694 million people currently use the internet worldwide according to comScore Networks: comScore announces new worldwide online unverse estimate based on the world's largest, most representative sample and most robust methodology; estimate marks a major step forward for marketers and media properties around the world. *Press Center* Retrieved 3 September, 2006, from http://www.comscore.com/press/release.asp?press=849

Constant, E. W., II. (1989). The social locus of technological practice: community, system, or organization? In W. E. Bijker, T. P. Hughes & T. Pinch (Eds.), *The social*

construction of technological systems: New directions in the sociology and history of technology (pp. 223-242). Cambridge, MA: The MIT Press.

Correa, F. R. (2000, November). Interview with Google's Sergey Brin. Retrieved 1 March, 2004, from http://www.linuxgazette.com/issue59/correa.html

Cothey, V. (2002). A longitudinal study of world wide Web user' information-searching behavior. *Journal of the American Society for Information Science and Technology, 53*(2), 67-78.

Cothey, V. (2004). Web-crawling reliability. *Journal of the American Society for Information Science and Technology, 55*(14), 1228-1238.

Couldry, N. (2004). Theorising media as practice. *Social Semiotics, 14*(2), 115-132.

Cutts, M. (2006, 4 February). Ramping up on international webspam. *Matt Cutts: Gadgets, Google, and SEO* Retrieved 20 March, 2008, from http://www.mattcutts.com/blog/ramping-up-on-international-webspam/

Da Vanzo, P. (n.d.). Ten Questions with: Cindy McCaffrey. Retrieved 25 February, 2004, from http://www.searchengineblog.com/interviews/interview_cindy_mccaffrey.htm

Davidson, E., & Pai, D. (2004). Making sense of technological frames: Promise, progress, and potential. In B. Kaplan, D. Truex, D. Wastell, A. T. Wood-Harperand & J. I. DeGross (Eds.), *Information systems research: Relevant theory and informed practice* (pp. 473-491). Boston: Kluwer Academic Publishers.

de Certeau, M. (1984). *The practice of everyday life*. Berkeley, CA: University of California Press.

Deuze, M. (2007). *Media work*. Cambridge: Polity.

Devlin, A. (1996, 9 September). Michael Mauldin on Lycos. *Ann Online* Retrieved March 10, 2006, from http://www.annonline.com/interviews/960923/

Diaz, A. (2008). Through the Google goggles: Sociopolitical bias in search engine design. In A. Spink & M. Zimmer (Eds.), *Web search: Multidisciplinary perspectives* (pp. 11-34). Berlin: Springer.

Dignan, L. (1999, 29 June). Compaq, CMGi make AltaVista deal official. *c|Net News.com* Retrieved 21 August, 2006, from http://news.com.com/Compaq%2C+CMGi+make+AltaVista+deal+official/2100-12_3-258854.html

Doyle, G. (2002). *Understanding media economics*. London: Sage.

Drèze, X., & Husherr, F.-X. (2003). Internet advertising: Is anybody watching? *Journal of Interactive Marketing, 17*(4), 8-16.

Economist. (2005, 13 August). Imperial bubble? China and the internet. Retrieved 24 May, 2006, from Lexis Nexis Executive database.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (1995). *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.

Euromonitor. (2005). Online adspend 2005. Retrieved 1 September, 2006, from Euromonitor International Global Market Information Database.

Fallhauber, P. (2009). SEO style for search-engine optimisation. *Public Relations Tactics, 16*(2), 19-19.

Feenberg, A. (1999). *Questioning technology*. London: Routledge.

Feenberg, A. (2000). From essentialism to constructivism: philosophy of technology at the crossroads. Retrieved December 1, 2002, from http://www-rohan.sdsu.edu/faculty/feenberg/talk4.html

Feintuck, M. (1999). *Media regulation, public interest, and the law*. Edinburgh: Edinburgh University Press.

Fetterly, D., Manasse, M., & Najork, M. (2004). *Spam, damn spam, and statistics: using statistical analysis to locate spam web pages*. Paper presented at the Seventh International Workshop on the Web and Databases (WebDB 2004), June 17-18, Paris, France.

Finkelstein, S. (2004, 4 May). Jew Watch, Google, and search engine optimization. Retrieved 22 February, 2005, from http://www.sethf.com/anticensorware/google/jew-watch.php

Fischer, C. S. (1992). *America calling : a social history of the telephone to 1940*. Berkeley: University of California Press.

Fleischer, M. H. (2009). *SEO made simple: Strategies for dominating the world's largest search engine*. Totowa, NJ: Lightning Press.

Fortunato, S., Flammini, A., Menczer, F., & Vespignani, A. (2006). *The egalitarian effect of search engines*. Paper presented at the WWW 2006 conference, Edinburgh, UK.

Francisco, B. (2004). China search engine Baidu builds up for an IPO. *CBSMarketWatch.com*

Freiwillige Sebstkontrolle Multimedia-Dienstabieter. (2004, 4 May). Code of Conduct. *FSM Website*   Retrieved 4 September, 2006, from http://www.fsm.de/en/CoC

Fujimura, K., Inoue, T., & Sugisaki, M. (2005, 10-15 May). *The EigenRumor algorithm for ranking blogs*. Paper presented at the WWW 2005, Chiba, Japan.

Galbraith, J. K. (1985). *The affluent society* (4th ed.). London: Andre Deutsch.

Gandy, O. (2004). Audiences on Demand. In A. Calabrese & C. Sparks (Eds.), *Toward a political economy of culture: Capitalism and communication in the twenty-first century* (pp. 327-341). Lanham, MD: Rowman & Littlefield.

Garnham, N. (1990). Media theory and the political future of mass communication. In F. Inglis (Ed.), *Capitalism and communication: global culture and the economics of information* (pp. 1-19). London/Newbury Park/New Delhi: Sage Publications.

Gasser, U. (2006). Regulating search engines: Taking stock and looking ahead. *Yale Journal of Law & Technology, 9*, 124-157.

Gavetti, G., & Rivkin, J. (2004). Rationality and plasticity over time: Toward a grounded theory of the origin of strategies.   Retrieved 20 March, 2006, from http://www.london.edu/assets/documents/PDF/Gavetti_paper.pdf

Geertz, C. (1973). *The interpretation of cultures: Selected essays*. New York: Basic Books.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Cambridge: Polity Press.

Giddens, A. (1989). A reply to my critics. In D. Held & J. B. Thompson (Eds.), *Social theories of modern societies: Anthony Giddens and his critics* (pp. 249-301). Cambridge: Cambridge University Press.

Gieber, W. (1999). News is what newspapermen make it. In H. Tumber (Ed.), *News: A reader* (pp. 218-233). Oxford: Oxford University Press.

Gilbert, G. N., & Mulkay, M. (1984). *Opening Pandora's box: A sociological anlysis of scientists' discourse*. Cambridge: Cambridge University Press.

Goffman, E. (1969). *The presentation of self in everyday life*. London: Allen Lane The Penguin Press.

Goldman, E. (2006). Search engine bias and the demise of search engine utopianism. *Yale Journal of Law & Technology, 9*, 111-123.

Google. (2000, 23 October). Google launches self-service advertising program: Google's AdWrods program offers every business a fully automated, comprehensive and quick way to start an online advertising campaign. *Google Press Center* Retrieved 21 August, 2006, from http://www.google.com/press/pressrel/pressrelease39.html

Google. (2002, 20 February). Google introduces new pricing for popular self-service online advertising program. *Google Press Center* Retrieved 21 August, 2006, from http://www.google.com/press/pressrel/select.html

Google. (2003a, 4 March). Google builds world's largest advertising and search monetization program: Company introduces automated content-targeted ads; Advertising customer base surpasses 100,000. *Google Press Center* Retrieved 21 August, 2006, from http://www.google.com/press/pressrel/advertising.html

Google. (2003b, 18 June). Google expands advertising monetization program for websites: Google AdSense enables sites to maximize revenue potential while enhancing user experience. *Google Press Center* Retrieved 21 August, 2006, from http://www.google.com/press/pressrel/adsense.html

Google. (2004a). An explanation of our search results. Retrieved 22 February, 2005, from http://www.google.com/explanation.html

Google. (2004b). My webpages are not currently listed. Retrieved 19 February, 2004, from http://www.google.co.uk/intl/en/webmasters/2.html

Google. (2005). Frequently Asked Questions *Google AdWords Professionals Support* Retrieved 20 March, 2008, from https://adwords.google.com/support/select/professionals/bin/index.py?fulldump=1

Google. (2006a). Google Annual Report 2005 [online edition]. *Google Investor Relations* Retrieved 21 August, 2006, from http://investor.google.com/pdf/2005_Google_AnnualReport.pdf

Google. (2006b). Where will my ads appear? Google help centre. *Google AdWords* Retrieved 2006, 18 August, from https://adwords.google.com/support/bin/answer.py?answer=6119&hl=en_GB

Google. (2007a). AdWords Advertising Policies. *Google AdWords Help Centre* Retrieved 31 October, 2007, from http://adwords.google.com/support/bin/static.py?page=guidlines.cs

Google. (2007b). What's an SEO? Does Google recommend working with companies that offer to make my site Google-friendly? *Google Webmaster Help Center* Retrieved 20 March, 2008, from http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=35291

Gregson, N. (1989). The (ir)relevance of structuration theory to empirical research. In D. Held & J. B. Thompson (Eds.), *Social theories of modern societies: Anthony Giddens and his critics* (pp. 235-248). Cambridge: Cambridge University Press.

Grehan, M. (2006, 15 May). Google's Matt Cutts: The Big Interview - Parts One and Two. *ClickZ Network* Retrieved 12 August, 2006, from http://www.e-marketing-news.co.uk/audio/mattcutts/part1.mp3 and part2.mp3

Grimmelman, J. (2007). The Structure of Search Engine Law. *New York Law School Public Law and Legal Theory Research Paper Series 23* Retrieved 25 November, 2009, from http://ssrn.com/abstract=979568

Halavais, A. (2009). *Search Engine Society*. Cambridge: Polity.

Hall, S. (1980). Encoding/decoding. In S. Hall, D. Hobson, A. Lowe & P. Willis (Eds.), *Culture, media, language : working papers in cultural studies, 1972-79* (pp. 197-208). London: Hutchinson in association with the Centre for Contemporary Cultural Studies University of Birmingham.

Hanseth, O., & Monteiro, E. (1998). Understanding information infrastructure. Retrieved 29 March, 2004, from http://heim.ifi.uio.no/~oleha/Publications/bok.html

Hård, M. (1993). Beyond harmony and consensus: a social conflict approach to technology. *Science, Technology, & Human Values, 18*(4), 408-432.

Hargittai, E. (2000). Open portals or closed gates? Channeling content on the World Wide Web. *Poetics, 27*(4), 233-254.

Hargittai, E. (2002). Beyond logs and surveys: In-depth measures of people's Web use skills. *Journal of the American Society for Information Science and Technology, 53*(14), 1239-1244.

Hargittai, E. (2007). Special Theme I: The Social, Political, Economic and Cultural Dimensions of Search Engines. *Journal of Computer Mediated Communication, 12*(3).

Harmon, S. (1998, 19 June). Disney buys part of Infoseek; Welcome to TomorrowWeb. *InternetNews.com*   Retrieved 21 August, 2006, from http://www.internetnews.com/bus-news/article.php/21171

Hart, C. (1998). *Doing a literature review : releasing the social science research imagination*. London: Sage.

Heilbroner, R. (1985). *The nature and logic of capitalism*. New York: W.W. Norton & Company.

Heilbroner, R. (1997). Technology and capitalism. *Social Research, 64*(3), 1321-1325.

Hellsten, I., Leydesdorff, L., & Wouters, P. (2006). Multiple presents: how search engines rewrite the past. *New Media & Society, 8*(6), 901-924.

Henzinger, M. R., Motwani, R., & Silverstein, C. (2002). Challenges in Web Search Engines. *SIGIR Forum, 36*(2), 11-22.

Herman, E. S., & Chomsky, N. (1994). *Manufacturing consent: The political economy of the mass media*. London: Vintage.

Herman, E. S., & McChesney, R. W. (1997). *The global media: the new missionaries of corporate capitalism*. London and Washington: Cassell.

Hertz, R., & Imber, J. B. (Eds.). (1995). *Studying elites using qualitative methods*. London: Sage.

Hindman, M., Tsioutsiouliklis, K., & Johnson, J. A. (2003, 28 July). "Googlearchy": How a few heavily-linked sites dominate politics on the web.   Retrieved 22 July, 2004, from http://www.princeton.edu/~mhindman/googlearchy--hindman.pdf

Höchstötter, N., & Lewandowski, D. (2009). What users see - Structures in search engine results pages. *Information Sciences, 179*, 1796-1812.

Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks-the International Journal of Computer and Telecommunications Networking, 33*(1-6), 337-346.

Hsieh-Yee, I. (2001). Research on Web search behavior. *Library & Information Science Research, 23*(2), 167-185.

IAC Search & Media. (2005, 1 August). Ask Jeeves launches Ask Jeeves sponsored listings: Automated system replaces Ask Jeeves premier listings product, enabling access to a broader base of self-services users delivers increased ROI on Ask Jeeves advertising spend. *Press Releases*   Retrieved 21 August, 2006, from http://www.irconnect.com/ask/pages/news_releases.html?d=83045

Infoseek. (1995a, 13 February). Infoseek launches first one-stop internet source for information. *Press Releases*   Retrieved 21 March, 2006, from http://web.archive.org/web/19970216144951/info.infoseek.com/doc/PressReleases/SearchLaunch.html

Infoseek. (1995b, 22 May). New, fast information retrieval service sets the standard for Web searches: Infoseek, Netscape and Sun sponsor free service. *Press Releases* Retrieved 23 March 2006, 2006, from http://web.archive.org/web/19970216144943/info.infoseek.com/doc/PressReleases/FreeNetSearch.html

Infoseek. (1996, 15 April). Infoseek forms strategic partnership with Kanematsu. *PR Newswire*   Retrieved 15 June, 2006, from from Lexis-Nexis Executive database

Infoseek. (1997). Company history. *Company Reference*   Retrieved 21 March, 2006, from http://web.archive.org/web/19970216145001/info.infoseek.com/doc/Reference/History.html

Inktomi Corporation. (1997, 6 March). Inktomi and NTT deliver largest, most powerful commercial search enigne for Japanese market; Goo search site integrates electronic commerce, advanced language analysis. *PR Newswire*   Retrieved 15 June, 2006, from from Lexis-Nexis database

Introna, L. D., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. *The Information Society, 16*(1), 169-185.

iResearch Consulting Group. (2005, 14 June). China internet search market report 2005 (simple version).   Retrieved 4 September, 2006, from http://english.iresearch.com.cn/html/search_engine/detail_free_id_6842.html

ITWeb. (2005, 18 January). Google SA 'A good sign'. *Africa News*   Retrieved 4 September, 2006, from Lexis Nexis Executive database.

Ivins, B. (2006, 31 May - 2 June). *More searchers searching more.* Paper presented at the Search Engine Strategies 2006 Conference & Expo, London, UK.

Jansen, B. J., & Pooch, U. (2000). Web user studies: A review and framework for future work. *Journal of the American Society for Information Science and Technology, 52*(3), 235-246.

Jansen, B. J., & Spink, A. (2005). An analysis of Web searching by European AlltheWeb.com users. *Information Processing & Management, 41*(2), 361-381.

Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management, 42*, 248-263.

Javary, M. (2004). Evolving technologies and market structures: Schumpeterian gales of creative distribution and the United Kingdom internet service providers' market. *Journal of Economic Issues, 38*(3), 629-657.

Jhally, S. (1987). *The codes of advertising : fetishism and the political economy of meaning in the consumer society.* London: Pinter.

Johnson, T. (2005, 17 July). In China, sophisticated filters keep the internet near sterile. *Knight Ridder Washington Bureau*   Retrieved 24 May, 2006, from Lexis Nexis Executive database.

Jones, K. B. (2008). *Search engine optimisation: Your visual blueprint for effective internet marketing.* Indianapolis, IN: Visual Press.

Jones, M. (1998). Information systems and the double mangle: Steering a course between the Scylla of embedded structure and the Charybdis of strong symmetry. Retrieved 16 February, 2004, from http://is.lse.ac.uk/helsinki/jones.pdf

Junnarkar, S. (1999, 19 January). @Home buys Excite in $6.7 billion deal. *c|Net News.com* Retrieved 31 March, 2006, from http://news.com.com/Home+buys+Excite+in+6.7+billion+deal/2100-1023_3-220281.html

Keane, M. T., O'Briend, M., & Smyth, B. (2008). Are people biased in their use of search-engines? *Communications of he ACM, 51*(2), 49-52.

Kezar, A. (2003). Transformational elite interviews: Principles and problems. *Qualitative Inquiry, 9*(3), 295-415.

Klaehn, J. (2002). A critical review and assessment of Herman and Chomsky's 'propaganda model'. *European Journal of Communication, 17*(2), 147-182.

Klein, H. K., & Kleinman, D. L. (2002). The social construction of technology: Structural considerations. *Science, Technology, & Human Values, 27*(1), 28-52.

Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment. *Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* Retrieved 28 August, 2003, from http://www.cs.cornell.edu/kleinber/auth.dpf

Kleinberg, J. M., & Lawrence, S. (2001). The structure of the Web. *Science, 294*, 1849-1850.

Kopytoff, V. (2000, 17 May). Huge deal to buy Lycos: Web portal to sell for $12.5 billion. *San Francisco Chronicle* Retrieved 21 August, 2006, from http://www.sfgate.com/cgi-bin/article.cgi?file=/chronicle/archive/2000/05/17/BU93774.DTL

Kopytoff, V. (2005, 7 May). Net industry eyes China: Business leaders from both nations to meet in Santa Clara. *The San Francisco Chronicle*, from Lexis Nexis Executive database.

Kopytoff, V. (2006, 9 March). Google to settle click-fraud lawsuit. *San Francisco Chronicle* Retrieved 19 March, 2008, from http://sfgate.com/cgi-bin/article.cgi?f=/c/a/2006/03/09/BUGRMHKQTR1.DTL

Langville, A., & Meyer, C. D. (2006). Google's PageRank and beyond: The science of search engine rankings.

Lastowka, G. (2007). Google's Law. *Rutger's School of Law Working Paper* Retrieved 25 November, 2009, from http://ssrn.com/abstract=1017536

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society.* Cambridge, MA: Harvard University Press.

Latour, B. (1991). Technology is society made durable. In J. Law (Ed.), *A sociology of monsters: essays on power, technology and domination* (pp. 103-131). London: Routledge.

Latour, B. (1996). *Aramis or the love of technology* (C. Porter, Trans.). Cambridge, MA: Harvard University Press.

Lave, J., & Wenger, E. (1991). *Situated learning : legitimate peripheral participation.* Cambridge: Cambridge University Press.

Law, J. (1987). Technology and heterogeneous engineering: The case of Portuguese expansion. In W. E. Bijker, T. P. Hughes & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 111-134). Cambridge, MA and London: MIT Press.

Law, J. (1991). Introduction: monsters, machines and sociotechnical relations. In J. Law (Ed.), *A sociology of monsters: essays on power, technology and domination* (pp. 1-25). London: Routledge.

Law, J. (1992). Notes on the theory of the Actor Network: Ordering, strategy and heterogeneity. Retrieved 28 October, 2003, from http://www.comp.lancs.ac.uk/sociology/soc054jl.html

Law, J. (1997). Traduction/trahison - notes on ANT.   Retrieved 21 October, 2002, from http://www.lancaster.ac.uk/sociology/stslaw2.html

Law, J., & Hassard, J. (1999). *Actor network theory and after.* Oxford: Blackwell's for Sociological Review.

Lawrence, S., & Giles, C. L. (1999). Accessibility of information on the World Wide Web. *Nature, 400*, 107-109.

Leiss, W., Kline, S., & Jhally, S. (1990). *Social communication in advertising : persons, products and images of well-being* (2nd ed.). London: Routledge.

Lewandowski, D. (2008a). A three-year study on the freshness of web search engine databases. *Journal of Information Science, 34*(6), 817-831.

Lewandowski, D. (2008b). The retrieval effectiveness of web search engines: considering results descriptions. *Journal of Documentation, 64*(6), 915-937.

Lewis, C. (2005). Negotiating the net: The internet in South Africa (1990-2003). *Information Technologies and International Development, 2*(3), 1-28.

Lichtenberg, J. (2000). In defence of objectivity revisited. In J. Curran & M. Gurevitch (Eds.), *Mass media and society* (3rd ed., pp. 238-254). London: Arnold.

Liquid Africa. (2005, 20 May). New SA search engine promises relevant results. *Comtex News Network, Inc.*   Retrieved 4 September, 2006, from Lexis Nexis Executive database.

Litterick, D. (2005, 31 August). Chirac backs eurocentric search engine. *Telegraph | Money* Retrieved 5 September, 2006, from http://www.telegraph.co.uk/money/main.jhtml?xml=money/2005/08/31/cnsearch31.xml

Livingstone, S. M. (1998). *Making sense of television: The psychology of audience interpretation* (2nd ed.). London: Routledge.

Livingstone, S. M., Van Couvering, E., & Thumim, N. (2004). *Adult media literacy: A review of the research literature.* London: Ofcom.

Lu, W. (2005, 30 August). Survey findings and analysis: China online search market survey report 2005. *China Internet Network Information Center*   Retrieved 18 June, 2006, from http://www.cnnic.cn/download/2005/2005083101.pdf

Machill, M., & Beiler, M. (2009). The importance of the internet for journalistic research: Multi-method study on the research performed by journalists working for daily newspapers, radio, television and online. *Journalism Studies, 10*(2), 178-203.

Machill, M., & Beiler, M. (Eds.). (2007). *The Rising Power of Search Engines.* Cologne: Herbert von Halem.

Machill, M., Beiler, M., & Zenker, M. (2008). Search-engine research: A European-American overview and systematization of an interdisciplinary and international research field. *Media, Culture & Society, 30*(5), 591-608.

Machill, M., Neuberger, C., & Schindler, F. (2003). Transparency on the Net: functions and deficiencies of Internet search engines. *Info - The journal of policy, regulation and strategy for telecommunications, 5*(1), 52-74.

Machill, M., Neuberger, C., Schweiger, W., & Wirth, W. (2004). Navigating the Internet: a study of German-language search engines. *European Journal of Communication, 19*(3), 321-347.

MacKenzie, D. (1987). Missile accuracy: A case study in the social processes of technological change. In W. Bijker, T. P. Hughes & T. Pinch (Eds.), *The social construction of technological systems* (pp. 195-222). Cambridge, Mass.: MIT Press.

MacKenzie, D., & Wajcman, J. (1999). Introductory essay. In D. MacKenzie & J. Wajcman (Eds.), *The social shaping of technology* (2nd ed., pp. 3-27). Buckingham and Philadelphia, PA: Open University Press.

Mansell, R. (2004). Political economy, power and new media. *New Media & Society, 6*(1), 96-105.

Mansell, R., & Javary, M. (2004). New media and the forces of capitalism. In A. Calabrese & C. Sparks (Eds.), *Towards a political economy of culture: Capitalism and communication in the 21st century* (pp. 228-243). Lanham, MD: Rowan & Littlefield.

Mansell, R., & Steinmuller, W. E. (2000). *Mobilizing the information society*. Oxford: Oxford University Pres.

Marcus, G. E. (1995). Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual Review of Anthropology, 24*(-), 95-117.

Marshall, M. (2005, 7 July). Google faces search engine China quandary. *San Jose Mercury News* Retrieved 24 May, 2006, from Lexis Nexis Executive database.

Martin, P. (1996, 18 April). Land-rush in cyberspace. *Financial Times FT.com* Retrieved 20 March, 2006, from http://news.ft.com/msn/s/25dbb3d6-6281-11da-8dad-0000779e2340,dwp_uuid=c2640462-6324-11da-be11-0000779e2340.html

Mattelart, A. (1991). *Advertising international : the privatisation of public space* (M. Chanan, Trans.). London: Routledge.

Mattelart, A. (1994). *Mapping world communication : war, progress, culture*. Minneapolis, MN: University of Minnesota Press.

McLuhan, M. (1964/2001). *Understanding media : the extensions of man*. London: Routledge.

McQuail, D. (1994). *Mass communication theory: An introduction* (3rd ed.). London: Sage.

Meisel, J. B., & Sullivan, T. S. (2000). Portals: The new media companies. *Info - The journal of policy, regulation and strategy for telecommunications, 2*(5), 477-486.

Meiss, M., & Menczer, F. (2008). Visual comparison of search results: A censorship case study. *First Monday, 7*(7).

Melody, W. H. (2003). Policy implications of the new information society. In M. Tool & P. Bush (Eds.), *Institutional analysis and economic policy* (pp. 411-432). Dortrecht, Netherlands: Kluwer.

Morrison, D. E., & Tumber, H. (1988). *Journalists at war: The dynamics of news reporting during the Falklands conflict*. London: Sage.

Mosco, V. (1996). *The political economy of communication: Rethinking and renewal*. London: Sage.

Mosco, V. (1999). New York.com: A political economy of the "informational" city. *The Journal of Media Economics, 12*(2), 103-116.

Mowshowitz, A., & Kawaguchi, A. (2002). Assessing bias in search engines. *Information Processing and Management, 38*, 141-156.

Mowshowitz, A., & Kawaguchi, A. (2005). Measuring search engine bias. *Information Processing & Management, 41*(5), 1193-1205.

Murdock, G., & Golding, P. (1999). Common markets: Corporate ambitions and communication trends in the US and Europe. *The Journal of Media Economics, 12*(2), 117-132.

Nader, L. (1972). Up the anthropologist: Perspectives gained from studying up. In D. H. Hymes (Ed.), *Reinventing Anthropology* (pp. 284-311). New York: Pantheon Books.

Nakamoto, M. (1996, 10 December). Softbank buys into Trend Micro. *Financial Times,* p. 32.

Navarro-Prieto, R., Scaife, M., & Rogers, Y. (1999). Cognitive strategies in Web searching. Retrieved 22 October, 2004, from http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/

New Media Age. (2005, 27 January). Chinese e-commerce: Fortune awaits. Retrieved 24 May, 2006, from Lexis Nexis Executive database.

New Media Age. (2006, 17 August). Profile - Christoph Mohn; Second coming. Retrieved 4 September, 2006, from Lexis Nexis Executive database.

Newcomb, K. (2006, 4 May). Microsoft AdCenter goes live. *ClickZ News* Retrieved 21 August, 2006, from http://www.clickz.com/showPage.html?page=3603746

Nicholson, S., Sierra, T., Eseryel, U. Y., Park, J.-H., Barkow, P., Pozo, E. J., et al. (2006). How much of it is real? Analysis of paid placement in Web search engine results. *Journal of the American Society for Information Science and Technology, 57*(4), 448-461.

Nielsen//NetRatings. (2006a). Germany: Top 10 parent companies month of July 2006. *Top Rankings* Retrieved 9 September, 2006, from http://www.nielsen-netratings.com/resports.jsp?section=pub_reports_intl&report=parent&period=monthly&panel_type=4&country=Germany

Nielsen//NetRatings. (2006b). Online Readership. *Online publishers association South Africa | readership* Retrieved 4 September, 2006, from http://opa.org.za/readership/

O'Brien, T. (1997). The millionaires next door. *Stanford Magazine* Retrieved 22 March 2006, from http://www.stanfordalumni.org/news/magazine/1997/mayjun/articles/excite.html

Odendahl, T., & Shaw, A. M. (2002). Interviewing elites. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context and method* (pp. 299-316). London: Sage.

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science, 3*(3), 398-427.

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science, 11*(4), 404-428.

Orlikowski, W. J., & Gash, D. C. (1994). Technological frames: Making sense of information technology in organizations. *ACM Transactions on Information Systems, 12*(2), 174-207.

Orlowski, A. (2004, 17 May). Google's ethics committee revealed. Retrieved 22 February, 2005, from http://www.theregister.co.uk/2004/05/17/google_ethics_committee/

Orna, E., & Stevens, G. (1995). *Managine information for research*. Buckingham: Open University Press.

Oxford English Dictionary. (1989). Entry for 'bias'. 2nd. Retrieved 29 March, 2004, from http://dictionary.oed.com

Ozmutlu, H. C., Spink, A., & Ozmutlu, S. (2002). Analysis of large data logs: an application of Poisson sampling on Excite web queries. *Information Processing and Management, 38*, 473-490.

Ozmutlu, S., Spink, A., & Ozmutlu, H. C. (2004). A day in the life of Web searching: an exploratory study. *Information Processing & Management, 40*(2), 319-345.

Perez, C. (2002). *Technical revolutions and financial capital: The dynamics of bubbles and golden ages*. Cheltenham: Edward Elgar.

Perkins, A. (2003). The classification of search engine spam. Retrieved 1 July, 2004, from http://www.ebrandmanagement.com/whitepapers/spam-classification/

Pinch, T., & Bijker, W. E. (1989). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. In W. E. Bijker, T. P. Hughes & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge, MA and London: The MIT Press.

Pinkerton, B. (2001). WebCrawler Timeline. *WebCrawler Facts* Retrieved 22 March, 2006, from http://www.thinkpink.com/bp/WebCrawler/History.html

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour*. London: Sage.

PriceWaterhouseCoopers. (2006, 6 April). IAB internet advertising revenue report: 2005 full-year results. *Internet Advertising Bureau* Retrieved 3 September, 2006, from http://www.iab.net/resources/adrevenue/pdf/IAB_PwC_2005.pdf

Rainie, L., & Shermak, J. (2005). Data memo: Search engine use November 2005. *Pew Internet and American Life Project* Retrieved 1 December, 2005, from http://www.pewinternet.org/pdfs/PIP_SearchData_1105.pdf

Reuters. (2004, 3 August). South Korean company buys Lycos. *Wired News* Retrieved 21 August, 2006, from http://www.wired.com/news/business/0,1367,64431,00.html

Rogers, R. (2000). Introduction - towards the practice of Web espistemology. In R. Rogers (Ed.), *Preferred placement - knowledge politics on the Web* (pp. 11-24). Maastricht: Jan van Eyck Akademie.

Rogers, R., & Zelman, A. (2002). Surfing for knowledge in the Information Society. In G. Elmer (Ed.), *Critical perspectives on the internet* (pp. 63-88). Lanham, MD: Rowman & Littlefield Publishers.

Röhle, T. (2007). Desperately seeking the consumer: Personalized search engines and the commercial exploitation of user data. *First Monday, 12*(9).

Rose, F. (2002, January). The $7 Billion Delusion. *Wired Magazine* Retrieved 4 April, 2006, from http://www.wired.com/wired/archive/10.01/excite_pr.html

Schiller, D. (1999). Deep impact: The Web and the changing media economy. *Info - The journal of policy, regulation and strategy for telecommunications, 1*(1), 35-51.

Schiller, H. (1992). *Mass communication and American empire* (2nd ed.). Boulder, CO: Westview Press.

Schiller, H. (1998). Striving for communication dominance: A half-century review. In D. K. Thussu (Ed.), *Electronic empires: Global media and local resistance* (pp. 17-26). London: Arnold.

Schudson, M. (1984). *Advertising, the uneasy persuasion: Its dubious impact on American society*. New York: Basic Books.

Schudson, M. (1989). The sociology of news production. *Media, Culture & Society, 11*(-), 263-282.

Schwartz, B. (2004, 5 August). Search memories: Live from SES San Jose. *SearchEngineWatch* Retrieved 17 March, 2006, from http://forums.searchenginewatch.com/showthread.php?t=949

Schwartz, B. (2009, 29 December). Hitwise: Facebook (sort of) more visited than google on Christmas. *Search Engine Land* Retrieved 25 April, 2010, from http://searchengineland.com/hitwise-facebook-more-visited-than-google-on-christmas-32554

Scoble, R. (2005, October 5). Andy Edmonds and Erik Selberg - Frank talk about MSN Search. *The Videos.* Retrieved 7 February, 2005, from http://channel9.msdn.com/showpost.aspx?postid=128677

Seekport. (2006). Mission. *Seekport UK Company Website.* Retrieved 5 September, 2006

SEMPO. (2006). The state of search engine marketing 2005. *Learning Centre* Retrieved 19 March, 2008, from http://www.sempo.org/learning_center/research/sempo_research/state_of_sem_2005/SEMPO-Market-2005-final-all.ppt

SEMPO. (2007, 8 February). Search engine marketing is a rocket; Spending is up 62%; Advertisers spent a total of $9.4 billion in 2006, according to SEMPO survey of SEM industry. *SEMPO Press Releases* Retrieved 4 March, 2007, from http://www.sempo.org/news/releases/02-08-07

SEMPO. (2008, 17 March). Search engine marketing shows strength as spending continues on a growth track against doom and gloom economic background. *Press Releases* Retrieved 18 March, 2008, from http://www.sempo.org/news/releases/03-17-08

Sewell, W. H. (1992). A theory of structure: Duality, agency and transformation. *American Journal of Sociology, 98*(1), 1-29.

Sherman, C. (2006, 28 June). Yahoo settles clickfraud lawsuit. *SearchEngineWatch Blog* Retrieved 20 March, 2008, from http://blog.searchenginewatch.com/blog/060628-202403

Sherman, C. (2008, 31 January). Report: click fraud up 15% in 2007. *Search Engine Land* Retrieved 20 March, 2008, from http://searchengineland.com/080131-132519.php

Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Engineering Bulletin, 24*(4), 35-43.

SinoCast China IT Watch. (2003, 27 October). Google, Baidu, Huicong dominate Chinese searhc engine market. *Global News Wire - Asia Africa* Retrieved 24 May, 2006, from Lexis Nexis Executive database.

SinoCast China IT Watch. (2004, 26 April). Baidu.com cautious of going public. *Global News Wire - Asia Africa* Retrieved 24 May, 2006, from Lexis Nexis Executive database.

Smythe, D. (1977). Communications: Blindspot of Western Marxism. *Canadian Journal of Political and Social Theory, 1*(3), 1- 27.

Spink, A. (2002). A user-centered approach to evaluating human interaction with Web search engines: an exploratory study. *Information Processing and Management, 38*, 401-426.

Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: Web search changes. *IEEE Computer, 35*(3), 107-109.

Spink, A., Wolfram, D., Jansen, M. B. J., & Saracevic, T. (2001). Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology, 52*(3), 226-234.

Spink, A., & Zimmer, M. (Eds.). (2008). *Web Search: Multidisciplinary Persepctives*. Berlin: Spinger.

Star, S. L., & Greisener, J. R. (1989). Translations and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*, 387-420.

Star, S. L., & Ruhleder, K. (1994). Steps towards an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. *Information Systems Research, 7*(11), 111-134.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). London: Sage.

Stuttgarter Nachrichten. (2004, 9 June). Web.de startet "SmartSearch" [Web.de launches "SmartSearch"]. *Stuttgarter Nachrichten Online - Computer* Retrieved 4 September, 2006, from http://www.stuttgarter-nachrichten.de/stn/page/detail.php/746567

Sullivan, D. (2000). Search engine alliances chart. Retrieved 27 April, 2006, from http://searchenginewatch.com/_subscribers/article.php/0006-alliances.mht

Sunday Times (South Africa). (2005, 26 June). Search and ye shall not find. *Grapevine* Retrieved 4 September, 2004, from Lexis Nexis Executive database.

Tartakoff, J. (2010). The (short) history of Twitter's plans to make money. *Guardian.co.uk News - Media - Digital Media* Retrieved 27 April, 2010, from http://www.guardian.co.uk/media/pda/2010/mar/29/twitter-making-money

The Nutch Organization. (2004). Nutch: faq. Retrieved 6 July, 2004, from http://www.nutch.org/docs/en/faq.html

Thelwall, M., & Hasler, L. (2006). Blog search engines. *Online Information Review, 31*(4), 467-479.

Thomas, R. J. (1995). Interviewing important people in big companies. In R. Hertz & J. B. Imber (Eds.), *Studying elites using qualitative methods* (pp. 3-17). London: Sage.

Thompson, R. (2008, 4 March). Blown away. *Financial Post: Business Magazine* Retrieved 19 March, 2008, from http://www.financialpost.com/magazine/story.html?id=324817

Thussu, D. K. (2000). *International communication: Continuity and change*. London: Arnold.

Total Telecom. (2004, 10 March). China's Web search engines set to take on Google. Retrieved 24 May, 2006, from Lexis Nexis Executive database.

Tuchman, G. (1973). Making news by doing work: Routinizing the unexpected. *American Journal of Sociology, 79*(1), 110-130.

Tuchman, G. (1978). *Making news: A study in the construction of reality*. New York: The Free Press.

Useem, M. (1995). Reaching corporate executives. In R. Hertz & J. B. Imber (Eds.), *Studying elites using qualitative methods* (pp. 19-39). London: Sage.

Van Couvering, E. (1998, 30 November). AOL/Netscape: Deal now to avoid distribution price hike. *Jupiter Communications European Internet Strategies Analyst Note* Retrieved 13 June, 1999, from http://www.jup.com/sps/eis/1998/57/

Van Couvering, E. (2004, 8 December). Literature searching and maps. Retrieved 20 October, 2005, from http://groupblog.workasone.net/archives/2004/12/literature-searching-and-maps/

van den Belt, H., & Rip, A. (1987). The Nelson-Winter-Dosi model and synthetic dye chemistry. In W. E. Bijker, T. P. Hughes & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 135-158). Cambridge, MA and London: MIT Press.

Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. *Information Processing & Management, 40*(4), 677-691.

Vaughan, L., & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes. *Information Processing & Management, 40*(4), 693-707.

Vise, D. A., & Malseed, M. (2005). *The Google Story.* New York: Delacorte Press.

Wallack, T. (2001, 17 December). Who killed Excite@Home: The suspects inlcude AT&T, the Excite and @Home merger, and cable companies. *San Francisco Chronicle* Retrieved 21 August, 2006, from http://www.sfgate.com/cgi-bin/article.cgi?file=/chronicle/archive/2001/12/17/BU23049.DTL

Walsham, G., & Sahay, S. (1996). GIS for district-level administration in India: Problems and opportunities. *MIS Quarterly, 23*(1), 39-66.

Wasko, J. (1994). *Hollywood in the information age : beyond the silver screen.* Cambridge: Polity Press.

Wasko, J. (2004). Show me the money: Challenging Hollywood economics. In A. Calabrese & C. Sparks (Eds.), *Toward a political economy of culture: Capitalism and communication in the twenty-first century* (pp. 131-150). Oxford: Rowman & Littlefield.

Webster, F. (2002). *Theories of the information society* (2nd ed.). London: Routledge.

Wenger, E. (1998). *Communities of practice: learning, meaning, and identity.* Cambridge: Cambridge University Press.

Werner, B., & Helft, M. (2000, 1 May). Portals Start to Feel the Heat. *The Industry Standard.*

White, D. M. (1999). The 'gatekeeper': A case study in the selection of news. In H. Tumber (Ed.), *News: A reader.* Oxford: Oxford University Press.

Wildman, S. S. (1994). One-way flows and the economics of audiencemaking. In J. S. Ettema & D. C. Whitney (Eds.), *Audiencemaking: how the media create the audience* (pp. 115-141). Thousand Oaks, London, and New Delhi: Sage.

Williams, M. (1997, 14 October). Excite launches Japan subsidiary. *Newsbytes* Retrieved 15 June, 2006, from from Lexis-Nexis database

Williams, M. (1998, 15 April). Lycos Japan established. *Newsbytes* Retrieved 5 June, 2006, from Leixs Nexis Executive database.

Winner, L. (1993). Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology. *Science, Technology, & Human Values, 18*(3), 362-378.

Wolfram, D., Spink, A., Jansen, B. J., & Saracevic, T. (2001). Vox populi: The public searching of the Web. *Journal of the American Society for Information Science and Technology, 52*(12), 1073-1074.

Xiang, Z., Wober, K., & Fesenmaier. (2008). Representation of the online tourism domain in search engines. *Journal of Travel Research, 47*(2), 137-150.

Yahoo! (1995, 29 June). Yahoo! forms marketing powerhouse to design and manage its new look: Yahoo! takes innovative approach to building powerful online advertising mechanism. *Yahoo! Media relations* Retrieved 20 March, 2006, from http://web.archive.org/web/20060320014607/http://docs.yahoo.com/docs/pr/release2.html

Yahoo! (1996, 22 April). Yahoo! Inc and Softbank unveil Yahoo! Japan: Online guide for Japanese Internet users goes live today. *Yahoo! Media relations* Retrieved 12 June, 2006, from http://yhoo.client.shareholder.com/press/ReleaseDetail.cfm?ReleaseID=173427

Yahoo! (2001, 13 November). Yahoo! forms alliance with Overture (formerly GoTo) to launch sponsor matches program for search results: Overture as its pay-for-performance search provide until at least April 2002. *Yahoo! Media Relations* Retrieved 20 August, 2006, from http://yhoo.client.shareholder.com/press/ReleaseDetail.cfm?ReleaseID=173810

Yahoo! (2002). Yahoo! to acquire Inktomi: Create the most comprehensive search offering on the Web with largest global audience, unmatched deadth and depth of online services and world class technology. *Yahoo! Media Relations*   Retrieved 21 August, 2006, from http://yhoo.client.shareholder.com/press/ReleaseDetail.cfm?ReleaseID=98489

Yahoo! (2003a, 7 April). Yahoo! introduces new Yahoo! Search: Faster, easier search results mark latest step toward Yahoo!'s goal of providing highest-quality search experience on the Internet. *Yahoo! Media Relations*   Retrieved 21 August, 2006, from http://yhoo.client.shareholder.com/press/ReleaseDetail.cfm?ReleaseID=105839

Yahoo! (2003b, 14 July). Yahoo! to acquire Overture: Acquisition positions Yahoo! as the largest global player in the rapidly growing internet advertising market, combines leading web and commercial search services with internet's largest global audience. *Yahoo! Media Relations*   Retrieved 21 August, 2006, from http://yhoo.client.shareholder.com/press/ReleaseDetail.cfm?ReleaseID=113537

Zhou, Y. (2006). *Historicizing online politics: Telegraphy, the internet, and political participation in China*. Stanford, CA: Stanford University Press.

Zimmer, M. (2008). The externalities of Search 2.0: The emerging priacy threats when the drive for the perfect search engine meets Web 2.0. *First Monday, 13*(3).

Zook, M. A. (2005). *The geography of the internet industry*. Oxford: Blackwell.