



**HAL**  
open science

## Search Engine: a Synthesis, Innovation & Regulation Chair, Workshop, 16 Mai 2008.

Pierre-Jean Benghozi

► **To cite this version:**

Pierre-Jean Benghozi. Search Engine: a Synthesis, Innovation & Regulation Chair, Workshop, 16 Mai 2008.. 2008. hal-00403509

**HAL Id: hal-00403509**

**<https://hal.science/hal-00403509>**

Preprint submitted on 10 Jul 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## **SEARCH ENGINE : A SYNTHESIS**

WORKSHOP - 16 May 2008

***Pierre-Jean Benghozi***

Recent information of a very different nature has shown the challenge posed by search engines in the internet economy.

- 60% of people seeking health-related information consult a search engine.
- On average, Americans spend 17 minutes a day on Google.
- 40% of online advertising revenue is directed towards search engines.
- Google captures 95% of European requests to search engines (75% in the USA).

### **Introduction**

In a digital economy which is increasingly becoming an information and attention-based economy, search engines are occupying a central role shaped by the symbolic status of Google both in terms of its renown and its dominance of internet traffic.

Despite their importance, economic models and competition structures for search engines are not widely understood. Thanks to their top-heavy IT infrastructures, they come under the network economy. And because they supply service applications, they can also be classified as part of the content economy. Google's disruptive success is down to the very fact that it has been able to innovate in the technical, service, use and business model spheres, all at the same time.

This diversity explains why it is so difficult to grasp the search engine phenomenon on a global scale. The aim of this workshop is to describe the market by systematic consideration of all the problems posed by search engine development, in terms of regulation, economic models and relations with access suppliers.

## **Content access tools**

Search engines are at the crossroads of information technology, telecommunications, marketing and the cultural sector. On the one hand, they orchestrate information exchanges on reading uses in return for information on readers, whilst on the other hand they orchestrate information exchanges on readers in return for advertising. Their technical dimension has considerable consequences on the range of content offered. In the first place, search engines have physical constraints linked to the time required to scan the entire web (3 or 4 days minimum). This means that they are not particularly suitable for "red hot" information, unlike portals based on publishing and selection activities (like Yahoo in its early days). Then again, despite the size of the web which they process, in order to be effective good general search engines mustn't surpass the expectations of internet users. They all tend to supply the same results, which isn't the case with specialized search engines, as their users expect them to come up with unexpected results.

Search engines are thus hard to pin down to an intermediary role between supply of content and demand for information. They are tools generating reading contents and part of an "access industry" based on the computerization of reading media. In this light, we could say that search engines represent a new stage of library industrialization, based on the same tool structures and organized according to library techniques, the index system, responses to specific requests and reading interfaces.

In economic terms, search engines differ from libraries in their relationship with the additional costs of little-used content matter: the "long tail". The negligible nature of these costs in the digital economy makes it possible to create a commercial model which is able to break free from the subsidies on which libraries are dependent. Search engines bring together supply and demand for content in a fractal sense; they can insert advertising – both charged and specific – at each interaction node because there is no pre-defined referencing system as you would find in a traditional library, with an existing encyclopedic knowledge organization system.

Against this background, relationships between libraries and search engines form an interesting topic for study and in certain cases, it has even been suggested that readers' rights should be established to respond to this industrialization of the reading environment. These relationships exist in large-scale digitization projects (Google Library, Quaero) or more global schemes within digital archives (accessibility, diversity, exclusion, property rights, fair use, investment costs, etc.).

The example set by a library such as the one in Lyons is an interesting case in point. The benefits that Google can offer the largest library in France's second city show that even now the largest libraries don't have the resources to digitize their catalogues. This process seems however crucial in order to reach a broader environment, as well as enhancing the availability, access tools and quality of the reading material supplied locally. This outsourcing of specific IT resources plays a part in adding value to document collections, but also creates new problems in its own right. Every web application turns now to be conceived as a 'library' these days, and traditional libraries have to compete with other operators. With digital technology, it is no longer enough for a library to hold external documents for consultation by internal readers; the trend is increasingly towards finding internal documents for distribution/promotion in the outside world.

### **New market forms**

Search engines have many different business models and present a wide variety of economic configurations. Most are based on a "free" business model which makes the search engine market very similar to the media market. It strives to reach the widest possible audience and places great importance on the creation of commercial brands to guarantee a good reputation and appeal.

Search engines do not restrict themselves to redefining the supply and economy of contents. They also design new market forms for information and communication where values have shifted from the applications and the information to the audience. In other words, search engines are "third party" prescribers who influence both the content producers and the consumers. They form part of a fragmented system comprising a primary market in which consumers choose the content, a prescribers' market in which consumers select a portal or a search engine and a referral market in which search engines "sell" their promotion capacity to advertisers.

Search engines are thus causing large-scale changes in the intermediary economy and multi-sided platform by extending the limits of their economic model. Some disruptive innovations concerning organization or business models thus have the ability to challenge established competitive positions: e.g. when existing software publishers are challenged by online applications. This then creates more competition between alternative business models than between rivals within the same economic model.

So whilst traditional internet participants target primary product markets by encouraging the integration of different functionalities within a specific technology to enable them to control consumers better, a large part of Google's success is down to the fact that it bases its economic model and

remuneration system on the web referencing market. Supplying sponsored contextual information alongside results is the cornerstone of Google's income and this is thus very different from an everyday advertising support function and more akin to an advertising consultancy role and a "placement" service. Customization and contextual adverts provide the "perfect market" in some ways for advertising by adapting the advertising level and cost precisely to each market niche.

This search engine business model, primarily advertising-based (98% of Google's income comes from Adwords), can also be regarded as one of its weaknesses. Others should not go unmentioned. The relevant search engine market is hard to define because there is considerable interaction with other markets (the advertisers' market) and it combines a number of advertising forms, online advertising, banners, sponsored links and search engines. It also has to contend with language barriers (note importance of Chinese, Russian or Arab markets). Finally, it has to deal with the emergence of information and communication forms which define new ways of distributing value, and especially with the development of RSS flows, widgets or closed social networks. From an economic standpoint, however, market analysis reveals that entry barriers are starting to be formed, based especially on IT storage capacity and information processing within servers.

So we need to ask ourselves whether the concentration trend we are noticing nowadays will still benefit Google and, in general, other search engines in the future, rather than other portal forms. In the history of search engines, the dominant players have always been in a virtual monopoly situation, but these leaders changed very regularly. Operators such as Exalead are an interesting case in point, as they illustrate a number of alternative dynamics. The company (website creator, consultant to technology suppliers) sells its search engine to companies and the Exalead search engine available on the web is simply a "showcase" which is not based on an economic model in its own right.

## **Regulations – what regulations?**

The considerable growth in online advertising and the dominant role played by search engines in this sector have led to financial issues regarding the bottlenecks represented by the advertisers' market and audience control. Competition thus tends to arise not only between other digital players, but also in associated markets centred on advertising (e.g. the press). These economic and legal pressures lead to new questions on the regulation methods which apply in this sector: analysis of the search engine market, effects on associated sectors both upstream (content suppliers) and downstream (advertisers), new forms of competition emerging in the software and infrastructure market, monitoring competition and current financial trends (Yahoo/Microsoft merger talks). A number of major issues also seem to

be affected: vertical relations between content publishers and access suppliers, copyright and intellectual property, trademark law, distribution and protection of personal data, the efficiency of the process of creating cultural programs and innovation.

The main question we need to ask is whether or not the search engine market should come under a generic set of regulations, and how market criteria can be applied to non-market objectives (freedom of expression, protection of privacy, etc.). Opinions often differ dramatically on this subject. First, we should remember that specific sector regulations (especially those of the telecoms sector) do not cover the search engine market, even though certain concepts might be useful such as the notion of a Minimum Quality of Service. Consequently, some believe there are no real problems with competition in this sector and that the regulatory authorities therefore don't need to get involved; even if problems do arise, a specific regulation would not be needed if traditional competition law was able to be applied.

On the other hand, others feel that the internet economy accentuates and reformulates some traditional questions on the subject of competition such as selectivity and discrimination between content suppliers (more or less well ranked and referenced by the search engines) or the risk of the main players blocking innovation. The Open Source strategy is sometimes quoted by way of example, as those players who get involved in publishing standards may seek to benefit from the community of developers to control radical innovations and prevent them from emerging.

From a competition viewpoint, relations between downstream and upstream of the value chain, between electronic communications operators and content suppliers (audiovisual in particular) are still the major source of potential conflict. Just as sites such as Google, eBay, Yahoo and Skype defend the creation of "open" platforms, including mobiles, to offset the control of subscribers by telecoms operators, the latter are simultaneously advocating the notion of passing on some of the investment needed in infrastructures to these sites by invoicing them for use of the network according to the applications used. This question has come directly to the fore in the USA, especially thanks to the debate on 'netneutrality' or on exclusivity clauses or tariff discrimination.

## **Intellectual property**

Compared to traditional information aggregators and databases (directories, television guides, etc.), search engines raise specific legal issues regarding the protection of intellectual property rights. However, these questions arise in a radically different way because of the technology of search engines. Questions are no longer just about referencing and aggregation, but also the new status of tags and hyperlinks, the existence of "cache" files or the

responsibilities of hosts. The justice system has already turned its attention to these legal issues surrounding search engines, but the debate extends way beyond a number of significant recent decisions such as the one involving Belgian press publishers.

Collecting and indexing information from the web represents the first source of conflict because search engines develop their market activities by exploiting content which they haven't paid for. They are then in direct competition with the information suppliers as regards advertisers, readers and audiences.

A second major source of conflict, which is both little known and often underestimated, concerns trademark law (affecting 20% of Google search requests). This trademark and advertising law is little used nowadays, but search engines could be attacked more than they are in this respect, since the referencing strategies used by many sites lead to illegal situations involving deception or overly similar advertising on the part of competitors. In this case, the responsibility and the role – active or passive – of the search engine have yet to be determined, especially if they are directly involved in suggesting key words.

### **Protection of personal data**

Search engines need to know the background of internet users' requests in order to provide pertinent and personalized responses and, at the same time, target consumers by advertising using sponsored links. This is meant to improve the traceability of search requests and consultation processes by diversifying within applications so as to be able to enhance the process of personalizing the supply of information. Such investments are only justified if the personal data collected can be exploited and used for added value which will inevitably be at odds with the rights of individuals: the more or less anonymous nature of the collected information, transparency and length of time for which the information is kept, marketing of the acquired data.

The individual protection policies in force conflict, in material terms, with the diverse approaches and creation of public data in various countries, even though the networks are international and transcend borders, and despite the fact that the European Data Protection Directive does not apply to search engines.

In terms of keeping search requests on file, systematic scanning of e-mails by search engines, and indeed the exploitation of identity profiles as put forward in social networks, consumer protection requirements have led to heated controversies concerning the best solution to adopt. Some defend strict regulations imposing very short limits and data storage times with a view to defending individuals and their right to keep certain pieces of information

confidential. Others feel that freedom of expression and personal choices can be more successfully defended by improving service performance within a transparent context, whilst all public documents should be available by free access, without any censorship as a general principle.

## **Conclusion**

To date, the pressures described above have only come to light occasionally, since search engines cover very different areas of the legislative system such as competition, personal data and intellectual property in particular. Against the background of a developing and highly innovative internet, the originality of search engines calls for new approaches which the regulatory bodies seem poorly equipped to handle. This is the topic of current debates, especially in Europe.