IRIS Special

Published by the European
Audiovisual Observatory

OBSERVATOIRE EUROPÉEN DE L'AUDIOVISUEL
EUROPEAN AUDIOVISUAL OBSERVATORY
EUROPÄISCHE AUDIOVISUELLE INFORMATIONSSTELLE

Searching for Audiovisual Content

COUNCIL OF EUROPE CONSEIL DE L'EUROPE

IRIS *Special:* Searching for Audiovisual Content

European Audiovisual Observatory, Strasbourg 2008
ISBN: 978-92-871-6559-6
EUR 89

Director of the Publication:
Wolfgang Closs, Executive Director of the European Audiovisual Observatory
E-mail: wolfgang.closs@coe.int

Editor and Coordinator:
Dr. Susanne Nikoltchev, LL.M. (Florence/Italy, Ann Arbor/MI)
Head of the Department for Legal Information
E-mail: susanne.nikoltchev@coe.int

Contributing Partner Institution:

Institute for Information Law (IViR)
Rokin 84
NL-1012 KX Amsterdam
Tel.: +31 (0) 20 525 34 06
Fax: +31 (0) 20 525 30 33
E-mail: website@ivir.nl
URL: http://www.ivir.nl

Editorial Assistant:
Michelle Ganter

Marketing:
Markus Booms
E-mail: markus.booms@coe.int

Translation/Proof-reading:
Caroline Bletterer, Alison Hindhaugh, Amélie Lépinard, Bernard Ludewig, Stefan Pooth, Erwin Rohwer, Roland Schmid, Dorothée Seifert-Willer, Nathalie Sturlèse, Candelaria Van Strien-Reney, Anne-Lise Weidmann

Typesetting/Print:
Pointillés, Hoenheim (France)

Publisher:
European Audiovisual Observatory
76 Allée de la Robertsau
F-67000 Strasbourg
Tel.: +33 (0)3 88 14 44 00
Fax: +33 (0)3 88 14 44 19
E-mail: obs@obs.coe.int
http://www.obs.coe.int

Please quote this publication as:
Susanne Nikoltchev, Ed., IRIS *Special:* Searching for Audiovisual Content
(Strasbourg, European Audiovisual Observatory, 2008)

IRIS *Special*

Searching for Audiovisual Content

In 1989/90 Tim Berners-Lee developed the World Wide Web (WWW) for the European Organization for Nuclear Research (CERN). Barely 20 years old, it has made an unprecedented impact on the way we deal with information. The WWW has become a key resource for private life and business. To each of us it offers an unimaginable variety and quantity of content and it calls on each of us to also use it in order to share our own creations with others. The WWW nurtures our expectations of what information we can find, it influences our habits of how and where to look for content, and it generates fears as to whether we master the machine or the machine masters us.

CERN was not only the Web's midwife, it might also serve as its role model for generating gigantic amounts of information. CERN expects to produce more than 15 million Gigabytes of data each year as a result of accelerating hundreds of millions of subatomic particles in the Large Hadron Collider (LHC), the world's largest particle accelerator. This bears some resemblance with the swelling mass of information floating on the WWW. What drives both systems is our desire to gather information.

When CERN launched its recent experiment to reconstruct the big bang that gave birth to our planet, critics reminded us that not all that technology can do is without risk. Certainly the responsible managers and scientists at CERN will do all they can to make their experiment safe, yet they also admit that a single mind can no longer control and overlook an experiment of that magnitude. Accordingly, supervision is carried out by a team with each of the many team members handling only a small part of the project.

Are there similar developments to be noticed for the WWW? Are we still in a position to handle it or have we created an information monster that a single human being can no longer control? Are we gradually being washed away by a deluge of information? These questions must lead us to reflect about the principal role or even several principal roles that the electronic information supply plays in our lives:

I. The amount of information electronically processed is mind-boggling. One of the contributions to this publication mentions that in 2008 electronically created, stored and replicated information reached the equivalent of three million times the information in all books ever written and that the amount will multiply by six over the next two years. No doubt somebody looking for relevant audiovisual content faces a big challenge. The fact that we still manage to find what we are looking for is largely owed to electronic helpers such as search engines and electronic programming guides.

In short: the abundance of electronic information compels us to develop and trust in new search facilitating tools.

II. The more content from an increasing number of sources is stored on a rapidly growing number of platforms and made available to more and more users with increasingly diverse backgrounds, the more difficult it is to provide a coherent system that will enable everybody to achieve satisfactory search results. The successful search for audiovisual content is a heightened challenge because most existing search tools are text-based.

A coherent approach to render data searchable requires a coherent methodology for identifying and classifying them. We need systematic ordering and tagging just like we did when we could still tidy away printed material in bookshelves.

In short: the abundance of electronic information might necessitate universal content identification systems.

III. The ease with which anybody can put anything about any subject online has a significant impact on how we can intrude upon each others lives. We may find online the most irrelevant and even manipulated but also the most pertinent information on any subject and any person. Moreover, we ourselves might be the target of such information without even knowing that it exists.

In short: the abundance of electronic information renders any control about personal information impossible.

IV. The need to channel or filter the vast array of information available on the Internet so that each may find the relevant pieces may clash with the right to freedom of expression. When search engines and electronic programming guides list relevant sources they also always decide what information to exclude from that list.

In short: Should the user be content that the selection among the abundance of electronic information is made by someone else and with the help of an automated system?

V. Given the importance of search engines and electronic programming guides for finding and accessing information, isn't there a need to control their functioning? Who, if anybody, steps in if information is excluded through the algorithm of a search engine for no apparent reason? According to which rules do search engines and electronic programming guides produce their results?

In short, does the European regulatory framework address audiovisual search tools?

VI. Not by chance has the World Wide Web also been referred to as the World Wild Web. It seems that regulators are necessarily limping behind technical developments and cannot move as fast as the World Wide Web expands. What can we, indeed, what *must* we do in order to ensure

that the online information supply is in line with the European value system? How do we, for example, secure the delicate balance between the freedom of information and the right to privacy? Isn't there a strong need to regulate? Could self- or co-regulatory mechanisms achieve what traditional regulation may not?

In short: Who, if anybody, can adequately delineate the legal framework for searching audiovisual content in an electronic world of abundance?

These six aspects are further developed in the six chapters of this IRIS Special emphasising the legal context in which they become relevant. The highly complex matter of this publication also provoked an intense exchange of views during a workshop co-organised by the Institute for Information Law (IViR) of the University of Amsterdam and the European Audiovisual Observatory in April 2008. The gist of the discussion is summarised in the Workshop Report with which this IRIS *Special* begins.

We would like to thank all participants to the workshop for their committed contributions to the discussion. Their names are listed at the end of this publication. Especially, we would like to express our gratitude to our colleagues at IViR who were instrumental in putting together and leading us through the workshop as well as to the authors of the six articles and the summarising report. Nico van Eijk and Christina Angelopoulos of IViR provided additional support by pre-editing the original (English) texts. And we like to point out once again that the success of the IRIS series is also the success of highly qualified and motivated translators and proof readers. We thank them for their valuable work on this edition.

Strasbourg, November 2008

Wolfgang Closs
Executive Director

Susanne Nikoltchev
Head of Department for Legal Information

IRIS *Special*

Searching for Audiovisual Content

CONTENTS

Workshop on Audiovisual Search
Summary of the Discussion

Christina Angelopoulos, Joris van Hoboken
Institute for Information Law (IViR), University of Amsterdam

Introduction

On 12 April 2008, the Institute for Information Law of the University of Amsterdam (IViR) and the European Audiovisual Observatory held a joint expert workshop in Amsterdam on the topic of "Audiovisual Search - Regulatory Challenges for Audiovisual Abundance". The aim of the workshop was to stimulate an exchange of ideas on the future of audiovisual search and the relevant regulatory issues. To this end, eight separate presentations were made, each followed by a round table discussion. The first set of presentations elucidated the current practical challenges facing the content navigation business. The second concentrated on an analysis of the normative and regulatory framework within which this business operates. This workshop report provides a summary of the opinions expressed and conclusions reached during the discussion rounds. A thematic, rather than a chronological, approach has been taken.

What became apparent from the presentations and subsequent discussions on audiovisual search services, the market and the laws governing the audiovisual content value chain was that this is a rapidly developing and extremely dynamic environment. One of the greatest challenges for the future will be to keep existing laws and regulation up to date and to facilitate the still developing market in audiovisual search technology and services. Fundamental questions about the free and effective flow of information between information providers and end-users and the important issue of privacy protection should inform the debate about the possible need to regulate audiovisual search engines. As search services move to the centre of the information environment, future information law and policy will have to address these new entities more adequately.

1. Current Practical Issues in Audiovisual Search

1.1. The Business Models of Search Engines and Evolving Search Technology

One of the first topics to catch the attention of participants was that of the current trends in audiovisual search, as well as in general search services, and the operation, in particular, of the relevant market. Special attention was paid to the question of entry barriers preventing new players from joining. It was agreed that the main barriers are the operational costs of crawling and indexing, on the one hand, and query answering, on the other, as suggested by presenter Ramón Compañó.[1]

1) See the article by Ramón Compañó in this publication.

Following on from this, the theory was put forth that, apart from financial costs, other elements, such as detailed information to better target audiences, might also create added hurdles. Established search engines are compiling what have been called "databases of intentions",[2] that enable them, through profiling, to match an individual's search patterns with those of others, helping them to enhance their efficiency and popularity and offer a compelling way of targeting audiences for the benefit of advertisers. If this emerges as an essential item for providing a successful search tool, newcomers who have not had the chance to gather similar data will find themselves at a disadvantage. Consequently, it was suggested that personalised services might therefore result in locked-in effects: at the moment, changing service provider is only a click away. If, however, a search engine holds personal data on its users, this can entrench them, creating obstacles for competitors to provide comparable services.

Does this mean that the Internet search market is one of natural monopoly? Participants were inclined to reject the idea, particularly in view of the following:

1. Google itself, now the predominant player, was in fact a late-comer. The first search engines, such as AltaVista, were eventually squeezed out of the market.
2. Google is indeed the market leader in the US and Europe, but this is not true for other parts of the world. It is possible that language might be an obstacle preventing expansion to certain regional markets.
3. Finally, subject-specific search engines are currently appearing, indicating niches in the market able to accommodate additional players.

The presentation of Thomas Roukens introduced the business model of the largest Belgian cable network provider Telenet and thus provided a more concrete basis for analysing existing business practices. An initial observation concerned the practice of vertical integration that Telenet has been pursuing. Although some participants were critical of such integration, it was conceded that, due to the linguistically-bounded Flemish market for which the company caters, the case in point is exceptional. It was noted that initially local broadcasters and content producers were wary of such digital services, believing that the Internet would wipe away their viewers and revenue. Telenet tried to overcome this attitude by creating a collaborative model with content providers and broadcasters and highlighting the complementary nature of the additional services it provided. As a result, not only did the transactions for Telenet's Video on Demand (VoD) service, iDTV (Integrated Digital Television), increase, but revenue for local broadcasters and content providers was also boosted and their position safeguarded. It was asserted that, in a small language community, vertical integration through collaboration can be the only way of surviving the competition of bigger broadcasters.

Another topic of discussion highlighted the evolving interactive design of Electronic Programme Guides (EPGs). Before launching its own service, Telenet was sceptical, questioning whether the modern consumer would not be more attracted to something more flashy and elaborate. Yet the simplicity of the service offered seemed to be precisely what drew consumers to it. In addition, consumers made enthusiastic use of features enabling the recording of programmes. The EPG eventually became the centre point of Telenet's interactive platform. Participants were puzzled by this development, asking whether it does not in fact transform the EPG into a recording device. Telenet, however, has not gained this impression, since the additional information on programmes is not simply still available through the EPG, but also widely used.

1.2 The Importance of Metadata

Janet Greco's presentation on the need for a consistent supply of metadata for use in EPGs turned the conversation to the relevant intellectual property rights. To what extent do intellectual property rights exist over the metadata, it was asked, and who holds them? Participants quickly pointed to the *Magill* case.[3] The case upheld the imposition of a compulsory license on television companies to remedy the exercise of their exclusive rights under national copyright legislation that prevented publishers of weekly guides from copying their listings. The case indicated that there was indeed

2) See the article by Ramón Compañó in this publication.
3) Joined Cases C-241/91 P and C-242/91 P *RTE and ITP v. Commission (Magill)* [1995] ECR I-743.

copyright over metadata in the form of TV listings, but established that, in exceptional circumstances, the exercise of an exclusive right by a proprietor may constitute abusive conduct. As one participant put it, this reasoning amounted to a conclusion that competition law can in some cases "override" intellectual property.

The subject gave rise to an exchange of opinions: on the one hand, it was felt that there are strong incentives to protect intellectual property rights over, for example, descriptions and ratings of audiovisual content. A reference to a programme is valuable since it reinforces society's capacity to compare and choose. The generator of this added value, therefore, should indeed be able to request control over how it is used. Hence, legislation such as the EU's Database Directive, which in essence protects the sweat of the brow that is invested in compiling such information, is justified. On the other hand, however, it was also asserted that from a business perspective, this approach creates complexities. One participant, accordingly, was of the opinion that copyright over such information as the title of a film or the name of its director, as opposed to e.g., a synopsis, should not be recognised, as these are mere points of fact. In an environment where copyright is acknowledged, an aggregator struggling to put together consistent and correct metadata has to juggle the management of business relations with countless diverse publishing information sources. Concerning ourselves with the ownership of the content, ran this argument, is missing the whole point: the goal of enabling consumers to find content that meets their interests and has all the required labelling from the regulatory point of view.

Yet other participants viewed the very idea of consistent metadata as unattainable in an interconnected world. Organising programme producers or distributors to the extent of enabling systematic information input and classification according to international standards is difficult. Moreover, the contents of the relevant databases are constantly changing and the rate of change is increasing. Yet the emergence of the Internet could make these problems irrelevant. Search engines on the Internet operate on an entirely different basis, but are arguably better equipped to deal with the disarranged state in which metadata currently finds itself. Through the use of algorithms, information retrieval is effective, if messy. Small inconsistencies, such as the use or not of an actor's middle initial, no longer compromise search results significantly. As one of the participants explained, search results do not have to be 100% accurate to be serviceable: They simply have to be good enough to enable a robust search tool. In this setting, it was asked whether EPGs are not an artefact of a pre-Internet world.

Ralph Traphöner's presentation, after all, which analysed the workings of the THESEUS project,[4] demonstrated that the ability of new search technology to deal with inconsistent data sets and extract information and metadata from an uncontrolled decentralized information environment is one of the core innovations in modern information retrieval technology. Furthermore, audiovisual data, from a technological point of view, is not fundamentally different to text: in order to obtain metadata that describes text, keywords are extracted from the strings. When it comes to audiovisual material, the paradigm is the same, although more sophisticated algorithms and computer power might be necessary. The irrelevance of traditional EPGs will become even greater, if, as one of the participants believed will happen, it becomes possible not only to search for (moving) images through the input of metadata, but also through the comparison of the images themselves.

The preceding observation turned the discussion to the role consumers themselves can play in categorizing and valuing audiovisual content. The examples of the Compact Disk Database (CDDB)[5] and the Internet Movie Database (IMDb)[6] were cited. In such models, the collection of data is partly placed

4) The THESEUS project is a research programme initiated by the German Federal Ministry of Economy and Technology with the goal of developing a new Internet-based infrastructure, so as to better use and utilize the knowledge available on the Internet. At the current time, 30 reasearch institutions, universities and companies have joined the programme under the coordination of empolis GmbH.

5) The Compact Disc Database (CDDB) is a Internet-accessible database that enables software applications to look up information on audio compact disks. The original software behind CDDB was released under the GNU General Public License and was based on voluntary contributions from users. The project was eventually incorporated as CDDB LCC in 1998. It was then bought by high-tech multimedia electronics manufacturer Escient and, in 2000, renamed Gracenote.

6) The Internet Movie Database (IMDb) is an online database of information and photos related to film, television shows, actors, production crew personnel and video games. The IMDb was launched on 17 October 1990 and was acquired in 1998 by Amazon.com.

in the hands of the consumers, who have a great incentive to provide information and the best knowledge of their needs. Such systems could therefore prove highly operable. Again, however, the question of ownership surfaces: it is indicative that CDDB was eventually incorporated, sold and relaunched as Gracenote, a wholly owned subsidiary of Sony Corporation of America, while the IMDb is now owned by Amazon.

In any case, it was suggested that the solution can only arise from the market. Database companies are likely to take the lead and evolve into the gatekeepers of metadata. One participant even posited that the solution could finally be found in competition: in contrast to the single pan-European PEGI classification system for video-games, in the realm of film Europe boasts 27 different classification boards. Possibly this is an advantage that gives consumers an insight into a more balanced view, instead of obliging them to accept a single perspective as written in stone.

2. Fundamental Rights Perspectives on Audiovisual Search

2.1. Audiovisual Content Online and Privacy

After Michael Zimmer's presentation on the Faustian Bargain we currently face, as audiovisual search technology simultaneously both enhances reach and recall of information and jeopardises users' privacy,[7] an animated discussion ensued. During the discussion, average users' knowledge of the threats to their privacy and their interest in protecting themselves was questioned. The objection raised to this line of thinking, however, was that privacy violations can be difficult for individuals to perceive on an abstract, de-contextualized level. On the contrary, if the problem affects them in a personal manner, i.e., when their own sense of privacy is compromised in reality, privacy concerns are much better understood. After all, search engines have become so integrated into our lives that the average user's sensitivity to the privacy threat they might pose has been dulled due to overexposure. The trust with which search engines inspire us is the key to their success.

It was further suggested that this problem is aggravated in the case of the so-called "information have-nots". The concern was expressed that those who are not connected and who therefore do not enjoy the benefits of technology might find themselves, in addition, in a weaker position vis-à-vis privacy on the net. The notice and take down procedure adopted by the relevant regulatory authorities of most countries quickly reaches its limit where one does not have or cannot master the technical means that enable awareness of invasions into privacy.

Finally, the efficacy of the current trend for user education as a form of defence was debated. It was pointed out that this presupposes a user not only aware of the existence of the problem, but also capable of retaining vast volumes of highly specialised information on everything from copyright and privacy legislation to consumer protection.

Turning to the possible solutions available, certain participants discerned an obligation for both academia and public advocacy to intervene. It was suggested that public advocates ought to raise issues and defend rights that concern all, but offer to no one individual a personal motive for immediate action. Their role could be especially beneficial for the weaker members of the information society. Additional regulatory solutions were also envisioned: the imposition of obligations such as face-blurring for services like Google Street View[8] forms a case in point; such measures, of course, take the process to the opposite pole, establishing protection as the default. The provision of educational information as a public service, to the end of facilitating user self-edification, was also advanced.

It was pointed out that many of the privacy issues that arise from the increasingly sophisticated audiovisual search services relating to people are, in one way or another, covered by EU data protection law. In this context, the recent opinion on data protection issues relating to search engines issued by

7) See the article by Michael Zimmer in this publication.
8) The Guardian, "Google Blurs the Privacy Issue" (May 2008), available at http://www.guardian.co.uk/business/2008/may/13/google.digitalmedia (accessed 15 June 2008).

the Article 29 Data Protection Working Party was mentioned, as having reached noteworthy conclusions, for instance, with regard to the use of facial recognition software by search engines: "Search engine providers that specialise in the creation of value added operations, such as profiles of natural persons (so called 'people search engines') and facial recognition software on images and audiovisual content must have a legitimate ground for processing, such as consent, and meet all other requirements of the Data Protection Directive, such as the obligation to guarantee the quality of data and fairness of processing"[9]. This again places the burden of protection on the service provider. On a relevant note, another solution proposed would involve private companies crawling the web in search of privacy-endangering data. The drawback of this solution, it was noted, is that it might only benefit the financially robust.

In a more technical realm, solutions for control over personal data and its processing can also be anticipated. Along this line of thought, a personal type of encryption and DRM that enable control of uploaded content without limiting the ability of sharing information with a community of online contacts was proposed. In the meantime, the application of machine readable licenses that protect data may also emerge (in the same way, for example, that the Creative Commons licensing suite is designed to protect copyrighted content). Nevertheless, multiple challenges were foreseen. Enforcement represents one stumbling block (how to oblige the end-user to respect designated restrictions?), while achieving the necessary level of regularity also seems difficult (how can one imagine every context in which an image might be used in order to encode that into a protective information system?). In the long run of course, if the idea is that privacy is becoming increasingly contextualised, then patterns might eventually be discerned that allow for some degree of automation.

On a different note, excessive privacy protection should also be avoided and the free flow of information respected, given that online content constitutes part of our public sphere. As a matter of fact, the view was also posited, during the discussion, that no type of privacy-protecting design is warranted: search engines may enable insight into another's private life, but, at the same time, users should be aware that information posted online is posted into a public realm. Detailed regulation might be an excessive reaction with only minimal gains for actual privacy rights, where a simple behavioural adjustment could achieve more effective results.

Finally, it is worth noting that the aforementioned concept of a "database of intentions"[10] also caused disquiet around issues of privacy. The reply offered was that trade-offs of this type are a necessary feature of the information society. At the end of the day, a balancing act has to take place between the value of effective information provision and the invasion of privacy that this might involve. In mobile search, for example, at the moment, the most popular queries relate to the weather, other local information and maps, which obviously have the capacity to reveal the person's location, yet equally provide valuable information on an immediate basis. The suggestion, therefore, was that our concept of privacy will in the future evolve, in order to accommodate possibilities that technology is only now enabling.

In fact, it seemed apparent to the majority of participants that the protection of privacy and data protection in particular should not focus on a block mode of protection, but rather a more flexible approach. An over-arching legal or even technological regime that accommodates all different kinds of interaction between people is difficult to envision, especially in the face of search functions which ignore the traditional barriers that allow for this differentiated interaction, thus doing away with "privacy through obscurity".[11]

2.2. Audiovisual Content Online and Freedom of Expression

In view of the above, the need for a clarification of search engines' legal position as concerns freedom of expression provisions becomes all the more pressing. Joris van Hoboken's presentation opened the debate on these issues.

9) Article 29 Working Party, Opinion on data protection issues related to search engines, WP 148, 4 April 2008.
10) See above, page 1.
11) See the article by Michael Zimmer in this publication.

Firstly, questions were raised as to the current legal framework's suitability to deal with freedom of expression issues in the online context. After all, Art.10 ECHR[12] has often been accused of being anachronistic; As much as the European Court of Human Rights has striven to maintain a dynamic and evolutive approach, it has not always been able to keep up with technological changes. With regard to the question of the protection of access for information providers to search engines, it was pointed out that Art. 10 does not entail a general right to reach an audience.[13] Yet this approach arguably could exclude some information providers from equitable access to some of the most effective means of communication, i.e., search platforms. After all, access to a willing audience is, in the final analysis, a premise for the exercise of freedom of expression: without it, your voice falls into the void. As an alternative to strict legal rules on access, the idea was put forth that value might be had in considering various soft law sources, such as Council of Europe recommendations.[14] Finally, the conjecture that the duties and responsibilities with which freedom of expression is coupled in Art. 10 might provide a basis for establishing search engine liability, at least in a co- or self-regulatory manner, was proffered.

Secondly, attention was paid to the recent Council of Europe's recommendation on Internet filters.[15] The suggestion, at the beginning of the very first guideline listed, that "users must be informed that a filter is active and, where appropriate, be able to identify and to control the level of filtering the content they access is subject to" caused some perturbation. One participant felt that such a requirement would be excessive, particularly in view of the widespread use of, for instance, spam filtering. It was asserted that much of such filtering is undertaken without the user's awareness or consent. Nevertheless, for the most part, it is considered a beneficial service. No reasonable need could therefore be discerned for user briefing on the operation of such filtering.

Finally, from the user side of the equation, the right to access information freely gains added importance in an online environment. In fact, Article 19 of the Universal Declaration of Human Rights[16] provides for a right to "seek [...] information and ideas through any media", an underdeveloped element that is often overlooked. Obviously, the user has an interest in freely searching the information that is available online. What is important to understand, however, is that what Art.19 guarantees, i.e., the freedom to seek information without hindrance, differs significantly from an actual right to access information. As one of the participants observed, the importance of this distinction becomes clear when one contemplates the main elements of the privacy discussion outlined above: when a society has developed such sophisticated search tools so as to enable logging of its members' every move, a right to actual access can emerge as a powerful and harmful weapon. Tensions between freedom of information and privacy thus emerge: each right, in an optimal situation, providing the necessary counter-balancing mechanisms to rein in the excesses of the other.

3. Regulatory Aspects of Audiovisual Search

3.1. The Place of Audiovisual Search in Current Laws and Regulation

As far as the regulatory treatment of search and navigation tools is concerned, a number of proposals were put forward as to the appropriate method of approaching the subject. Peggy Valcke's presentation elucidated the fragmented treatment that is currently in force: The current EU Regulatory Framework for Electronic Communications (ECNS – Electronic Communications Networks and Services

12) Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (signed 4 June 1950, entered into force 3 September 1953) (hereinafter: ECHR) Art 10.

13) See *Stafford v. the United Kingdom*, Judgment of the European Court of Human Rights (Grand Chamber) of 28 May 2002, para. 68; *Appleby and Others v. the United Kingdom*, Judgment of the European Court of Human Rights (Fourth Section) of 6 May 2003, paras. 47 and 48; In addition, similar issues arise in respect of broadcasting in, *inter alia*: *VgT Verein gegen Tierfabriken v. Switzerland*, Judgment of the European Court of Human Rights (Second Section) of 28 June 2001; *Haider v. Austria*, Decision of inadmissibility of the European Commission of Human Rights (First Chamber) of 18 October 1995, Application No. 25060/94.

14) See the article by Joris van Hoboken in this publication.

15) Recommendation CM/Rec(2008)6 of the Committee of Ministers to member states on measures to promote the respect for freedom of expression and information with regard to Internet filters, available at https://wcd.coe.int/ViewDoc.jsp?Ref=CM/Rec(2008)6 (accessed 1 July 2008).

16) Universal Declaration of Human Rights (adopted 10 December 1948) UNGA Res 217 A (III), Art. 19.

Directives) is inappropriate for the regulation of audiovisual search engines and search engines in general. Art. 2 (c) of the Framework Directive states categorically that the term "electronic communications services" excludes information society services. Search engines, therefore, are not covered. Likewise, they cannot be considered associated facilities, which Art. 2 (e) Framework Directive currently defines as "facilities associated with an electronic communications network and/or an electronic communications service which enable and/or support the provision of services via that network and/or service."[17] EPGs, on the contrary, are specifically mentioned by the Directive as constituting associated facilities. In any case, the ECNS, in terms of its broadcasting aspects, was intended for the regulation of vertically integrated TV platforms, where one operator controls a series of facilities. In the case of the Internet, on the contrary, the value chain is structured differently.

One place in the current regulatory framework was identified where room for search engines might exist, at least as concerns consumer protection. The proposed text for a revised Art. 20 (5) of the Universal Service Directive[18] as it stands would read as follows: "Member States shall ensure that where contracts are concluded between subscribers and undertakings providing electronic communications services and/or networks, subscribers are clearly informed in advance of the conclusion of a contract and regularly thereafter of any limitations imposed by the provider on their ability to access or distribute lawful content or run any lawful applications and services of their choice." The suggestion was that, with the opportunity of the ongoing reform of the ECNS, legislative changes could be made to the effect of broadening the notion of associated facilities, as defined above, so as to include search engines. In addition, the proposed Art. 20 (5) should be rewritten to apply also to the situation where no contract exists as such, but an individual is making *de facto* use of a service. The way would then be open to oblige search engines to provide added transparency concerning e.g., ranking methods or sponsored links. Reactions to this suggestion were hesitant, especially in view of the fact that the same effect could be achieved through search engine self-regulation.

Finally, the extent to which the terms of the E-Commerce Directive[19] can influence search engines was also investigated. As was remarked, however, although this Directive does in general apply to search engines, in fact it offers very little in terms of actual obligations. Articles 12 to 14 of the E-Commerce Directive provide safe harbours for three types of intermediaries: caching, hosting and mere conduit.[20] Art. 21 of the Directive makes it clear that these provisions do not cover "the liability of providers of hyperlinks and location tool services". On the contrary, the Article dismisses search engines to further examination in a biannual report on the application of the Directive, in which particular consideration shall be paid to "the need for proposals concerning the liability of providers of hyperlinks and location tool services, 'notice and take down' procedures and the attribution of liability following the taking down of content." [21]

At this point it was put forward that, notwithstanding the E-Commerce Directive's safe harbour for caching in Art. 13, some of the caching by search engines could involve liability, especially considering the habit of search engines of storing caches of webpages for the exact purpose of offering them when access to the original page is cut off. Analogies were made at this point in the discussion to similar provisions in the Copyright Directive, which contains a provision on caching as well.[22] In this context, reference was made to the *Copiepresse* case, currently winding its way through the Belgian courts. It was pointed out that the Belgian Court of First Instance[23] has held that caching on the part of Google amounts to unauthorised reproduction and communication to the public and is, therefore, a violation

17) Directive 2002/21/EC of the European Parliament and of the Council of 7 March 2002 on a common regulatory framework for electronic communications networks and services (Framework Directive) [2002] OJ L 108/33.
18) Directive 2002/22/EC of the European Parliament and of the Council of 7 March 2002 on universal service and users' rights relating to electronic communications networks and services (Universal Service Directive) [2002] OJ L 108/51.
19) Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce) [2000] OJ L178/1 (hereafter: E-Commerce Directive).
20) E-Commerce Directive, Arts. 12-14.
21) E-Commerce Directive, Art. 21.
22) Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L 167/10, Art 5.
23) *Google Inc v. Copiepresse SCRL* (RB (Brussels)) Tribunal de Première Instance (Brussels) 13 February 2007; [2007] E.C.D.R. 5 2007 WL 1623283.

of copyright. The court did find that, to the extent that caching by a search engine consists of the "automatic, intermediate and temporary storage of that information, performed for the sole purpose of making more efficient the information's onward transmission to other recipients of the service upon their request"[24], the E-Commerce Directive's safe harbour in general does apply. Nevertheless, the ultimate conclusion was that, in the particular case, the issue at stake was not the temporary storage of cached pages as part of the indexation process, but the visibility of the cached pages for users, which was held to be a breach of copyright.

Of course, the obvious place for audiovisual search engine regulation, if necessary, would be as part of the Audiovisual Media Services (AVMS) Directive's regulation of audiovisual services.[25] After all, it is this that governs cultural and content issues, as opposed to the ECNS framework that deals with questions of transmission and competition. The problem here is that the AVMS Directive currently does not cover audiovisual search engines,[26] has only very recently, in December 2007, been reviewed and is unlikely to be revisited by the Commission for some time to come. Nevertheless, a theoretical discussion on the topic ensued. This mainly revolved around the interpretation of the term "editorial responsibility".

Editorial responsibility is that requirement for an audiovisual media service, as defined by Art. 1 of the AVMS Directive, which was mentioned as providing the main stumbling block for including search engines within its remit.[27] The same article defines editorial responsibility as "the exercise of effective control both over the selection of the programmes and over their organisation either in a chronological schedule [...] or in a catalogue [...]". During the course of the discussion, it was established that the intention at least of the law-maker when drafting the Directive was to exclude search engines from this definition. The preparatory works, it was felt, made this clear: what was aimed at was the activities of broadcasters, i.e., the selection of programmes and their inclusion in a chronological scheme or catalogue. This is distinct from the automatic generation of a list through the use of an algorithm. Recital 19 of the AVMS Directive states that "the definition of media service provider should exclude natural or legal persons who merely transmit programmes for which the editorial responsibility lies with third parties." "Carriers", therefore, including search engines, were argued to be unavoidably excluded.

Nevertheless, some of the participants held the view that this definition should be sufficient to encompass search engines. According to this position, the use of an algorithm is in fact a demonstration of effective control: after all, it was pointed out, an algorithm that does not take into account the exclusion of incitements to hatred would have to be modified accordingly. As opposed to hosting providers, who need not take heed of the content of the information stored on their servers to the extent that it is legal,[28] search engine algorithms do take an interest in the nature and content of the information contained on the websites to which they provide references. If not for any other reason, then this would be because it is precisely their mission to furnish their users with information relevant to their query.

The opposition to this train of thought turned on a technical understanding of the way in which search engines actually function. The observation was thus made that it is actually the content that controls the search engine, rather than the other way around. Consequently, a search engine is different from a catalogue or directory behind which a human editorial team, which definitely does exercise editorial responsibility according to Art. 1 AVMS Directive, can be found. Similarly, a distinction should be made with regard to vertical search platforms, as, for example, Google News, seeing as there the engine operator makes a decision as to what particular type of content the search engine is searching. Accordingly, it was suggested that a case can be made for the position that platforms such as YouTube do exercise editorial responsibility. Once again, an analogous perception in

24) E-Commerce Directive, Art. 13.
25) Directive 2007/65/EC of the European Parliament and of the Council of 11 December 2007 amending Council Directive 89/552/EEC on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the pursuit of television broadcasting activities [2007] OJ L 332/27.
26) See the article by Peggy Valcke in this publication.
27) See the article by Peggy Valcke in this publication.
28) E-Commerce Directive, Art. 14.

the field of copyright could have severe implications for search engine liability. The *Viacom* case,[29] for example, was mentioned: if it is accepted that a specific service provider actually does exert editorial control, then it would be eliminated from safe harbour provisions in the DMCA or similar legislation in Europe, making it responsible for the prevention of possible copyright infringement by, and the illegal or harmful nature of, the content to which it links.

Finally, from the business perspective, Telenet's approach to editorial responsibility was also discussed. To begin with, it was noted that, due to the broad Belgian definition of broadcasting in the law, on-demand services are indeed included; meaning that for Telenet editorial responsibility can in any case not be avoided. Within this legal framework, Roukens also explained that Telenet has tried to balance the expectable desire of a company to minimize accountability with customer expectations. Research indicates quite clearly that people are not familiar with on-demand services. They expect the same kind of service, therefore, on their IDTV platform that they receive from a film rental. Telenet has responded to such expectations, through, e.g., systems fitted into their set-top boxes limiting access to films rated as unsuitable for children.

3.2. Self- and Co-regulation and Search Engines

As a result of the current lack of centralized and clear regulation on the part of the state, search engine providers are obliged to set a standard for themselves. Yet, as Wolfgang Schulz asserted in his presentation, this self-regulatory practice is far from transparent and of questionable efficacy. Especially from the point of view of the smaller providers, the resulting legal uncertainty can be difficult to manage. In general, it could produce a chilling effect.[30] The observation was made during the workshop that search engines themselves have actually expressed a desire for more precise regulatory stipulations that would provide them with a more stable footing for questions such as whether or not, for example, picture previews in search results can constitute copyright infringement in and of themselves.

In any case, to the extent that it exists, search engine self-regulation mainly involves notice and take down procedures. It is worth noting that no put back provision is at the moment in effect, something that could prove problematic. However, during the discussion, the desirability of this notice and take down method as a defence against objectionable material on the Internet was brought into question. After all, one participant remarked, if the target is limited to the mere reference, which is automatically generated using third party information, while the original material remains online, what we are engaging in remains at best an exercise in futility. Targeting the original website would surely be more effective. In this context, the proposal of approaching search engines as being similar to caching services, rather than hosting ones, was put forth.

These reflections aside, one question raised concerned the German *Freiwillige Selbstkontrolle Multimedia-Diensteanbieter (FSM)* code.[31] This involves a self-regulatory code of conduct, which the major search engines in Germany have adopted since 2004. Some confusion seemed to exist among participants as to the voluntary nature of the scheme. The search engines themselves appear to claim the contrary in certain contexts, i.e., that the self-regulatory scheme is State-imposed. So, although a number of the websites deemed inappropriate under the scheme are to a great extent still accessible

29) The Viacom case (*Viacom International Inc, vs. YouTube, Inc,* No. 07 Civ. 2103 (S.D.N.Y., 13 March 2007)) is a closely-watched case currently pending before the U.S. courts, which deals with the USD 1 billion lawsuit for massive copyright infringement filed, in March 2007, by the media conglomerate Viacom ("Video & Audio Communications") against YouTube and its parent company Google. Viacom claims that YouTube hosts numerous unauthorized copyrighted clips of its entertainment programming uploaded by users. YouTube, on the other hand, asserts applicability of the safe harbour provision of the Digital Millennium Copyright Act (17 U.S.C. § 512(c)(1) (2000)), under which ISPs are not required to monitor their sites for infringing material, but must remove such material promptly once its existence has been brought to their attention. Viacom maintains that the requirements for the exercise of the safe harbour exception are not met.

30) See the article by Joris van Hoboken in this publication.

31) Subcode of Conduct for Search Engine Providers ("VK-S") of the Association of Voluntary Self-Regulating Multimedia Service Providers (*Verhaltenssubkodex für Suchmaschinenanbieter,* "VK-S", *der Freiwilligen Selbstkontrolle Multimedia-Diensteanbieter*) of 21 April 2004, available in English at:
http://www.fsm.de/en/Subcode_of_Conduct_for_Search_Engine_Providers (accessed 1 July 2008) and in the original German at: http://www.fsm.de/de/Subkodex_Suchmaschinenanbieter

via some participating search engines, Google, on the other hand, states on its search result page that the removal of search results is as the consequence of a legal request. The participants' conclusion was that the answer lies in the interpretation of the word "voluntary": When providers filter on a self-voluntary basis, measures such as the founding of self-regulation initiatives, the adoption of a code of conduct and the operation of a complaints office (as happens in the case of the FSM) serve to externalise the responsibility. In this manner, compulsion to conform would seem to derive from an external body, which the parties themselves experience as external pressure. That, however, is not to say that an actual legal obligation exists as such.

3.3. The Actual Need for Regulatory Intervention

In the final analysis, however, the conclusion seemed to be that the very need for regulation of search engines should not be taken for granted. When faced with Wolfgang Schulz's detailed list of the risks triggered by search engines,[32] one participant objected, claiming that this might constitute an overly gloomy approach. For example, the precise meaning of, and dangers presented by, the fragmentation of the public sphere was inquired into. The answer referred to studies that have indicated that the use of so-called "research media" shortens the agenda of issues upon which an individual gathers information. Traditional media, on the other hand, bring readers face to face with subject matter that they would not necessarily have pre-selected themselves, but which might, in effect, be of interest to them or of relevance to their needs.[33] The counter-argument presented was that search media could be viewed as facilitating the creation of communities of common interests, thereby increasing social cohesion.

What was agreed upon by all participants was that the way in which people approach information is changing in a radical manner. Jane Buckingham, founder of the Intelligence Group, has reported a college student's conviction that, "if the news is that important, it will find me."[34] Some participants therefore felt that the problem with regulating search engines could be that, at the moment, we reside only in an interim phase: the consequences and effects new media will have and the pitfalls they may contain are as yet unclear. In a similar vein, the idea was also introduced that the use of search engines could actually help reduce the overall need for regulation. For example, in the field of consumer protection, search engines can help strengthen the position of users through easy access to knowledge about the products presented to them. Likewise, when it comes to media pluralism, search engines have the power to guide their users precisely to a pluralistic supply of services. Transparency is thus increased and the need for regulatory intervention abated. Such benefits, however, although real, should not blind one to the need to address the separate question of the possible reasons for regulatory intervention presented by search engines themselves.

Accordingly, other participants saw a rosy, not a gloomy, picture in the aforementioned list of risks. According to this view, most of the dangers it includes are either already covered by existing legislation, by self-regulation or by self-organisation on behalf of users. So, the danger of distortion of competition is a question for competition law and access to harmful content can be limited through self-regulatory codes of conduct. Finally, issues such as the fragmentation of the public sphere can be counterbalanced through the potential of the Internet for bringing common interest groups together. In the final conclusion, it was suggested that a much better use of regulators' time and effort would be, in relation to both broadcasters and any associated facilities (such as EPGs) that are available on a given platform, to make them deliver complete information about their programmes, properly tagged. In this way, the information accessed through audiovisual search tools would be correct and no call for an *ex ante* regulation of these tools themselves would present itself.

---

32) These would include the following: Access to harmful content, Access to illegal content, Discrimination of content, Misleading consumers, Influence on opinion making, Fragmentation of the public sphere, Exploiting protected works, Exploiting personal data, Distortion of competition, including transfer of market power to other markets (e.g., advertising). See the article by Wolfgang Schulz in this publication.

33) Klaus Schönbach et al., *Online Newspapers: A Substitute for Print Newspapers and Other Information Channels?* 6th World Media Economics Conference, Centre d'études sur les médias and Journal of Media Economics, HEC Montréal, Montréal, Canada, 12-15 May 2004, available at http://cf.uba.uva.nl/nl/handl/googlescholar/ (accessed 9 July 2008).

34) Brian Stelter, "Finding Political News Online, the Young Pass It On", available at http://www.nytimes.com/2008/03/27/us/politics/27voters.html?_r=1&adxnnl=1&oref=slogin&adxnnlx=1215003699-5hhJsU3ewQ+YzGMGjWskOQ (accessed 2 July 2008).

Techno-economic Challenges of Audiovisual Search Engines

Ramón Compañó (*)
European Commission - Joint Research Centre (JRC)
Institute for Prospective Technological Studies (IPTS)

Acknowledgements:

This paper is dedicated to my colleague and friend Boris Rotenberg, who passed away on 23 December 2007 in a tragic skiing accident. Many of the ideas discussed here were conceived during debates with him about the prospects for search engines. I will always remember him as a dynamic and bright professional and an enthusiastic person.

This work has been supported by the European Commission through the IST project "Chorus" (Contract number 2006-045480).

(*) Disclaimer

The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission. Neither the European Commission, nor any person acting on behalf of the Commission, is responsible for the use that might be made of the following information.

Introduction

Audio-visual (AV) search engines promise to become key tools for the audio-visual world, as text search did for the present text-based digital environment. AV search applications will enable us to reliably index, sift through, and "accredit" (or give relevance to) all forms of audiovisual creation (both individual and collaborative). Moreover, AV search will become central to predominantly audiovisual file-sharing applications. It will also lead to innovative ways of handling digital information. For instance, pattern recognition technology will enable users to search for categories of images or film clips. Likewise, AV search could be used for gathering all the past voice-over-IP conversations in which certain keywords were used.

However, if these kinds of applications are to emerge, search technology must develop rapidly in scale and type. There will be a growing need to investigate new AV search techniques built, for instance, around user behaviour. Therefore, AV search is listed as one of the top priorities of the three major US-based search engine operators – Google, Yahoo! and Microsoft. In Europe, the setting up of the French Quaero initiative,[1] for the development of a top-notch AV search portal, or the German Theseus research programme,[2] provide further evidence of the important policy dimensions of AV search.

There are a variety of AV search engines for different types of applications. They differ by type of audio-visual content (text, audio, video), by technological platform (mobile, PC), or by application (e.g., publicly available information on the web versus proprietary databases). From the general public's point of view, the best known ones are "web search engines". The term "web search engine" refers to a service available on the public Internet that helps users find and retrieve content or information[3] from the publicly accessible Internet.[4] Well-known examples of web search engines are Google, Yahoo!, Microsoft and AOL's search engine services. Web search engines may be distinguished from search engines that retrieve information from sources that are not publicly accessible. Examples of the latter include those that only retrieve information from companies' large internal proprietary databases (e. g., those that look for products in eBay or Amazon, or search for information in Wikipedia), or search engines that retrieve information that cannot be accessed by web search engines.[5] Similarly, we also exclude from the definition those search engines that retrieve data from closed peer-to-peer networks or applications that are not publicly accessible and do not retrieve information from the publicly accessible Internet. They can be for general purpose search, like Google, Yahoo! or MS Live Search, or specialized for particular themes. Thematic or vertical search portals include those specifically designed for children,[6] for people search,[7] those focused on health-related topics,[8] for retrieving software codes,[9] for radio search, music search and for science, etc. A regularly updated list of search engines by domain is available at Pandia.[10]

---

1) http://www.quaero.org/ (last visited: 6 June 2008).
2) http://theseus-programm.de/ (last visited: 6 June 2008).
3) It is more accurate to refer to search results as "content" or "information", rather than web pages, because a number of search engines retrieve information other than web pages. Examples include search engines for music files, digital books, software code and other information goods.
4) For a similar definition, see James Grimmelmann, *The Structure of Search Engine Law* (draft), 13 October 2006, p.3, at http://works.bepress.com/james_grimmelmann/13/ (last visited: 6 June 2008). It is acknowledged that many of the findings of this paper may be applicable to different kinds of search engines.
5) Part of the publicly accessible web cannot be detected by web search engines, because the search engines' automated programmes that index the web, crawlers or spiders, cannot access them due to the dynamic nature of the link, or because the information is protected by security measures. Although search engine technology is improving with time, the number of web pages is increasing drastically too, rendering it unlikely that the "invisible" or "deep" web will disappear in the near future. As of March 2007, the web is believed to contain 15-30 billion pages (as opposed to sites), of which one-fourth to one-fifth are estimated to be accessible by search engines. See and compare
http://www.pandia.com/sew/383-web-size.html (last visited: 6 June 2008) and
http://technology.guardian.co.uk/online/story/0,,547140,00.html (last visited: 6 June 2008).
6) E. g., http://www.fragfinn.de/
7) E. g., http://www.spock.com/
8) E. g., http://www.healia.com/
9) E. g., http://www.koders.com/ or http://www.krugle.org/
10) http://www.pandia.com/powersearch/index.html (last visited: 6 June 2008).

While web search engines scan publicly available information, this is not the case for search on private *desktops* or in *closed proprietary multimedia databases*. The former is popular with individuals who download the freely-available tools provided by major search engine providers to search for data on their personal computers. Although there are some privacy concerns, they are popular because they are free. Companies with sensitive data make use of business solutions that take care of their specific needs. National libraries are an example. Another type is *search in peer-to-peer networks*, which is aimed at finding audio-visual content on the sites of private PCs (for example, search tools operating on BitTorrent exchange platforms). Another emerging trend is *mobile search*.

In the following section, some fundamental technological aspects will be introduced, in order to understand the major technological challenges.

1. Technological Challenges

1.1. Search Engine Basics

The state-of-the-art in search engine technology depends on the specific retrieval tool (audio, images or video) and the platform (e. g., distributed computer architecture, peer-to-peer network, etc.). The purpose of this section is to present technological elements in order to understand important technological challenges, rather than to offer an exhaustive description.

Figure 1: Search Engine Processes

In essence, a search engine is composed of a number of technical components that fulfil four basic exchanges of information: information gathering and indexing, user querying, information provision and user information access. The gathering of information is undertaken by automated software agents called robots, spiders, or crawlers.[11] Once the crawler has downloaded a page and stored it on the search engine's own server, a second programme, known as the indexer, extracts various bits of information regarding the page. Important factors include the words the web page contains, the location of these key words (e. g., title), the weight attributed to specific words (e. g., rare words have a higher search value than frequently-used ones), the proximity of words, or the hyperlinks on the page. Importantly, the index is not an actual reproduction of the page or something a user would want to read. The index is further analysed and cross-referenced to form the runtime index that is used in the interaction with the user.

When a user sends a query, the engine browses the index on the server. The query algorithm is the "soul" of the search engine. This algorithm undertakes two major processes. First, it defines the matching process between the user's query and the content of the index. Second, the algorithm sorts and ranks the various hits. Therefore, the degree of relevance to the user depends on the algorithm. The user obtains content either from the link to the original source or from the search engine's cache, i. e., the "temporary archive" on the search engine's own server. The "cache" is a copy of the last time the search engine's crawler visited the page in question. Thus, it is not necessarily up to date, but may be useful for the user, if the server or page is temporarily unavailable, or if the user wishes to find out what the latest amendments to the web page were.

Current search engines are predominantly text-based, even for AV content. This means that non-textual content like images, audio and video files are indexed, matched and ranked according to textual

11) There are also non- or semi-automated alternatives on the market, such as the open directory project whereby the web is catalogued by users, or search engines that tap into the wisdom of crowds to deliver relevant information to their users, such as Wiki Search, the wikipedia search engine initiative (http://search.wikia.com/wiki/Search_Wikia) (last visited: 6 June 2008), or http://www.chacha.com/ (last visited: 6 June 2008). See Wade Roush, *New Search Tool Uses Human Guides*, Technology Review, 2 February 2007, at http://www.techreview.com/Infotech/18132 (last visited: 6 June 2008).

clues such as filenames, tags, text near images or audio files (e. g., captions) and even the anchor text of links that point directly at AV content. Truveo is an example of this for video clips,[12] and SingingFish for audio content.[13]

While text-based search is efficient for text-only files, this technology has important drawbacks when it has to retrieve information from content other than text. For instance, images that are very relevant for the subject of enquiry will not be listed by the search engine if the file is not accompanied by the relevant tags or textual clues. For example, although a video may contain a red mountain, the search engine will not retrieve this video when a user inserts the phrase "red mountain" in his search box. The same is true for any other information that is produced in formats other than text. In other words, a lot of relevant information is systematically left out of the search engine rankings, and is inaccessible to the user. This, in turn, affects the production of all sorts of new information.[14]

Thus there is a huge gap in our information retrieval process. This gap is growing with the amount of non-textual information that is being produced at the moment. Researchers across the globe are currently seeking to bridge this gap. One strand of technological developments revolves around improving the production of meta-data that describes the AV content in text format. A solution could be found by, for instance, developing "intelligent" software that automatically tags audio-visual content.[15] However, though technology is improving, automatic tagging is still very inefficient due to complex algorithms and high processing or computational requirements. Another possibility is to create a system that tags pictures using a combination of computer vision and user-inputs.[16] However, manual tagging is time consuming and thus costly.

1.2. Challenges

Technical challenges are multiple and complex, as they concern both fundamental theoretical principles and implementation issues. For the sake of simplicity, these challenges can be grouped into the modules presented in Figure 1, namely: collecting the maximum amount of information (data gathering), understanding user queries, adjusting to the user's context (user interaction), supplying the most relevant results, supplying results which help user interaction (returning results). Another challenge is related to how to manage the multiplicity and interoperability of devices and platforms. This will be discussed briefly in the following section.

1.2.1. Gathering Information: Dealing with the Data Explosion

The amount of information created, stored and replicated in 2000 was 3 billion gigabytes (Exabyte), 24 Exabyte in 2003, and 161 Exabyte in 2008.[17] For the sake of comparison, the last figure is equivalent to three million times the total information contained in all books ever written. This figure is expected to reach 988 Exabyte by 2010.[18] We can fairly say that we are witnessing a data explosion trend.

This huge amount of data comes in a variety of formats, and content has evolved far beyond pure text description. In fact, though data was still largely text-based in the early days of digitization, it is believed that now 93% of the data on the Internet is in multimedia format and this will approach 99% in 2010. It is also interesting that 70% of this data is created or achieved by non-professional

12) http://www.truveo.com/ (last visited: 6 June 2008).
13) SingingFish was acquired by AOL in 2003, and has ceased to exist as a separate service as of 2007. See http://en.wikipedia.org/wiki/Singingfish (last visited: 6 June 2008).
14) See Matt Rand, *Google Video's Achilles' Heel*, Forbes.com, 10 March 2006, at http://www.forbes.com/2006/03/10/google-video-search-tveyes-in_mr_bow0313_inl.html (last visited: 6 June 2008).
15) In this regard, see James Lee, *Software Learns to Tag Photos*, Technology Review, 9 November 2006, at http://www.technologyreview.com/Infotech/17772/. See Chris Sherman, *Teaching Google to See Images*, Search Engine Land, 5 April 2007, at http://searchengineland.com/070405-172235.php (last visited: 6 June 2008).
16) See Michael Arrington, *Polar Rose: Europe's Entrant Into Facial Recognition*, Techcrunch, 19 December 2006, at http://www.techcrunch.com/2006/12/19/polar-rose-europes-entrant-into-facial-recognition (last visited: 6 June 2008).
17) See and compare http://www.pandia.com/sew/383-web-size.html (last visited: 6 June 2008) and http://technology.guardian.co.uk/online/story/0,,547140,00.html (last visited: 6 June 2008).
18) See Andy McCue, *Businesses face data 'explosion'*, ZDNet, 23 May 2007, at http://news.zdnet.co.uk/itmanagement/0,1000000308,39287196,00.htm (last visited: 6 June 2008), referring to IDC/EMC Study *The expanding Digital Universe*.

users.[19] This underlines the view that in the coming years, users will gradually shift from being mere consumers to becoming producers, providers and mediators of content. It is fair to assume that search engines, in order to cope with the growing amount of audiovisual content, will increasingly become AV search engines. Future audiovisual search engines will have to deal with two major challenges. First, they will need to be scalable to deal with more and more data. The amount of data is due not only to the aforementioned creation of original data to store, but also, and to an even greater extent, to the huge adjacent processing and meta-data necessary to make an effective retrieval of AV content possible. In fact, the indexing of audio, images and video requires extensive resources in terms of computer processing power and storage capacity. Second, they will need to be able to process different kinds of audiovisual material. This will possibly require less deterministic search concepts, e. g., moving from exact to fuzzy matching. The need for an increasing scalability and less determined content will lead to new computer architectures and search concepts, which are sketched in the following figure and will be briefly discussed in the following section.

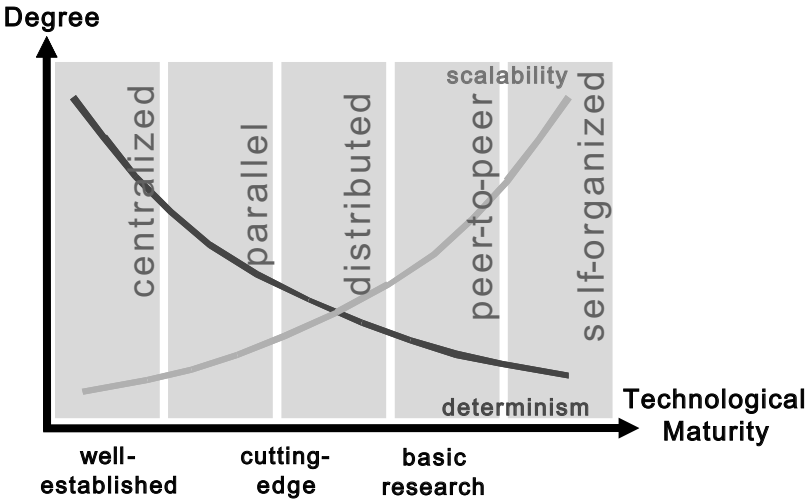


Figure 2

The multimedia content produced daily by Internet users will exceed the processing power of today's search engines to index and search AV content within associated text and meta-data. This has implications for computer architecture (from centralised to future self-organised clusters) and the way searches are made – from very to less deterministic searches, e. g., from exact to fuzzy matching, or from precise to approximate searches.[20]

In early times, content was stored in centralised servers. As long as the amount of information was reasonably large and content mostly text-based, search was reasonably good. As content size grew, centralized computer systems could not respond to users within an acceptable time. Centralised systems had to be replaced, first by parallel computer architectures and later by distributed systems. In fact, many scientists believe that state-of-the-art distributed processing will not be sufficient in the future to cope with the increasing amount of data. They think that future web search engines will be peer-to-peer (P2P) based, and in the long-term – given the complexity of operating such P2P computer architectures efficiently – a degree of self-organization will be adopted to deal with complexity autonomously.

Today's state-of-the-art computer architectures for search engines are distributed systems. P2P systems are in an advanced phase of development, research prototypes have shown their potential[21] and some companies, like Faroo,[22] are already offering beta versions of P2P web search engines with distributed index, crawler and ranking. P2P retrieval could become a scalable alternative to clustered

19) Fauto Rabitti, Sapir presentation at Concertation meeting Vilamoura (Faro, Portugal) 16 April 2008.
20) http://www.sapir.eu/papers/sapir-poster.pdf (last visited 6 June 2008).
21) http://www.sapir.eu/ (last visited 6 June 2008).
22) http://www.faroo.com/ (last visited 6 June 2008).

retrieval engines if state-of-the-art P2P search techniques can be shown to operate reliably in a large number of peer machines, i. e., if they are supported by a large community. It is expected that for full-text retrieval, P2P approaches will contribute to document partitioning (unstructured overlay network for search, e. g., Gnutella) and term partitioning (structured overlay network for search e. g., Chord, P-Grid), while for audiovisual search, P2P will mostly contribute to the indexing task and also to similarity research. It is unclear when - or indeed if - P2P architectures will be ready to replace distributed systems. Some (isolated) solutions for P2P similarity search for single AV features (e. g., colour or shape) are already available, but systems combining text and multiple features (e. g., colour AND shape) are not yet efficient.

From an operational point of view, it is expected that future search engines will be able to perform better when accessing the deep web and retrieve any sort of audiovisual content. The increased amount of data poses challenges with respect to data storage and processing, particularly for indexing and rapid response to user queries. It has to be noted that multimedia formats (images, video, audio, 3D, etc.) make the indexing task far more complex than text–based formats do. Until new architectures are in place and solutions for non-indexed content are available, it is possible that there will be a widening AV information retrieval "gap". It is likely that search engines will rely for a considerable period on two complementary AV search approaches, namely annotation-based technologies, where indices are generated automatically or manually, (e. g., tagging) and content-based technologies, e. g., similarity search. To help the search process at a later stage, indexing right from the source level would be beneficial. For instance, commercial devices would have a range of data already built in, such as authorship, production process, geographical data, etc. In addition, there would be the option for user-supplied indexing, introducing tagging, comments, etc. Such AV-related meta-data would then help the finding of relevant content.

1.2.2. Understanding User Queries: Bridging the Semantic Gap

Typing keywords is not the ideal way to interact in a search for information. Getting closer to a kind of natural speaking language would help overcome "keywordese"; but this solution is a long way off. While "simple phrase" search is already possible today, slightly more complicated sentences are difficult for machines to handle as contextual search is not possible. For instance, if we introduce the term "Jaguar" the search engine is unable to distinguish whether we mean the automobile or the animal, as processing the user's query is linked to the machine's ability to understand the semantics of the query. As mentioned before, progress in machine learning and artificial intelligence has been steady, but the semantic web as a reality is still a long way off.

Retrieval performance decreases with the "complexity" of content, as will be illustrated in the case of images. From a retrieval point of view, the simplest images are those with clear boundaries, colours, geometry (e. g., no deformation), shape, etc. Examples are "artificial" images, like logos or symbols. They are relatively easy to retrieve by search engines because "similarity" approaches work very well. Therefore, search engines are used in applications to detect trademark infringements, as trademarks are often images with low-level features.

The next group of images are those with specific similarity measures. They are more complex than the previously mentioned "artificial" images, but they have particular features that render their retrieval fast. One example is frontal pictures of faces,[23] because their individual features (eyes, mouth and nose) are well defined, as are the spatial relationships between the features. Another example is fingerprints. Their colour contrast is pronounced and the matching shape similarity is done by comparing them with fingerprints already stored as pictures in a database.

23) Software is able to detect faces based on the high contrast areas such as the eyes, nose and mouth. In optimal situations - near frontal face portraits in high-quality photographs - the systems perform at over 95% accuracy with minimal false positives. For side (non-frontal) views, low-quality pictures or older images from cultural heritage collections the performance is still far above average. For an overview, see: Yang, M.H., Kriegman, D.J., and Ahuja. N. 2002. *Detecting Faces in Images: A Survey*. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(1), 34-58.

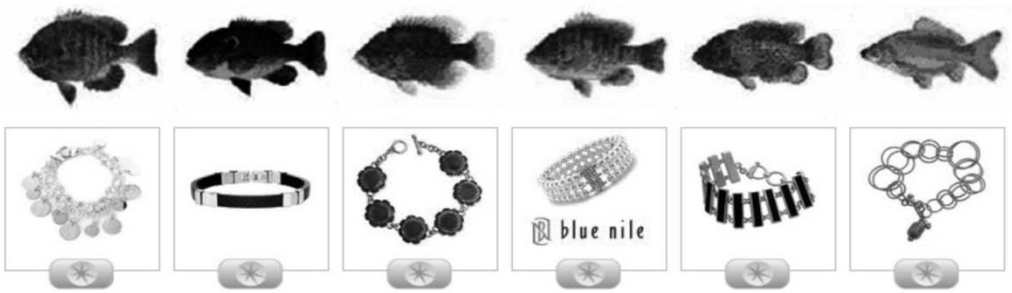


Figure 3: "Semi-artificial" pictures

Semi-artificial pictures are more complex from a retrieval point of view. These are images taken under well-defined conditions, such as one single object per picture against a homogeneous background. The sets of fish and bracelets in Figure 3 are examples that fulfil these conditions. Searching for this kind of picture is useful for educational purposes, particularly for encyclopaedias, and also for commercial purposes (i. e., searching for particular products).[24] The most difficult pictures to retrieve are real images, like those in photo albums or on the Internet. They contain many different types of objects, (e. g., landscapes, people, art, artificial objects, etc.), rendering them "unpredictable" and complex and, thus, difficult to analyse and process.

Technologically speaking, machines can retrieve low-level feature images, like artificial images, fairly well by analysing basic elements like the object's colour, texture and shape. Therefore, similarity search engines have proven useful in specific contexts like searching trademark databases,[25] detecting copyright infringement,[26, 27] finding video shots with similar visual content and searching for music with similar rhythms. In all these applications, the basic features such as colour and texture in images and video, or dominant rhythm, melody or frequency spectrum in audio, are tightly correlated to the search goals of the particular application. For instance, in SongTapper,[28] a user can search for a song by typing the rhythm of the song on the space bar of the keyboard or, in Midomi,[29] by humming the song into a microphone.

High-level features are far more complex, as an "understanding" of the content is often required. For instance, humans can immediately identify a two-dimensional projection of a three-dimensional object, i. e., we can easily recognize a different picture as just another projection (from a different angle) of the very same object. Machines, however, are unable to recognise the 2D representation of 3D objects. They lack the understanding of the image content and therefore of the way a human would describe an object. The difference between two descriptions of an object by different representations is called the semantic gap. *Closing the semantic gap is one of the biggest challenges in computer science*.

This is not just an academic challenge, as bridging the semantic gap would have practical consequences which would allow more user-friendly interfaces. These would be able to translate contextual knowledge in natural language (high-level) into an elementary and reproducible operation of a computing machine (low-level). Designing such user-friendly search engines could bring the vast amounts of multimedia knowledge from libraries, databases and collections closer to citizens. These systems would need to understand the semantics of a query, and not just the underlying low-level computational features alone. ImageScape[30] was an early content-based retrieval system addressing semantic gap issues in the query interface, indexing and results. This early search engine allowed a

24) E.g., http://www.riya.com/ (last visited 6 June 2008).
25) Eakins, J.P., Riley, K.J., and Edwards, J.D. 2003. *Shape Feature Matching for Trademark Image Retrieval.* CIVR, 28-38.
26) Jaimes, A and Chang, S-F. 2002 *Duplicate Detection in Consumer Photography and News Video,* ACM Int. Conf. on Multimedia, 423-424.
27) Joly, A., Buisson, O., and Frelicot, C. *Robust content-based copy detection in large reference database*, Int. Conf. on Image and Video Retrieval, 2003.
28) http://www.songtapper.com/ (last visited 6 June 2008).
29) http://www.midomi.com/ (last visited 6 June 2008).
30) Lew, M.S. 2000. *Next Generation Web Searches for Visual Content.* IEEE Computer, November, 46-53.

user to query multiple visual objects such as sky, trees, water, etc., using spatially positioned icons in an index containing more than 10 million images and videos. Automatically detecting all semantic content within an image such as a face, tree, animal, etc., with emphasis on the presence of complex backgrounds, is still a huge challenge. Researchers trying to classify whole images often get stuck on technological challenges and on practicalities, such as the picture's granularity being too coarse to be useful in real world applications.

Progress in research is steady, although the semantic gap problem is far from being solved. Current approaches towards learning semantics are multiple, and include concepts like examining hidden associations during image indexing, developing visual dictionaries which group similar colours and textures, or introducing learning approaches, and many others.

1.2.3. Returning More Relevant Results: Getting the User into the Loop

The level of relevancy depends on the quality of the search algorithm. One difficulty search engine developers have to face is pollution and negative externalities - for example, Spamdexing or link-bombing are well-known examples of unauthorized search engine optimization tools which search engine providers have already had to face up to.

One idea about how to increase the quality of similarity-based search systems is to enable them to learn through use, by integrating continuous feedback from the user query. This interactive process is called "relevance", "query refinement", or "interactive search". Here, a list of candidate images is presented to a user who is asked to rank the relevance of these images to his search. One of the major problems in relevance feedback is how to address small training sets, as a typical user may want to label far fewer images than the algorithm needs. One option for better user feedback is to offer an adequate representation of the information retrieved. Some search engines are starting to introduce visual interfaces in the structure of the information, and the link to other pieces of information is visualized. Clusty[31] is an example of a search engine that clusters results, and Kart00[32] offers a visual interface of the links and relationships of the retrieved results. Another way of increasing relevance is to make use of social bookmarking, like Mister-Wong,[33] or getting user's feedback into search process like the engine ChaCha.[34]

Search services would become more efficient and relevant if likely queries could be anticipated, at least to some extent. This would require the personalisation and customisation of the search engine to the user. In extreme cases, high-levels of personalization would mean that systems would have to store equally large amounts of personal information on individuals. Consequently, the aggregate of all the user's searches would get us closer to what John Battelle calls the "Database of Intentions".[35]

The motivation for personalisation and customization depends on the actor involved. The user expects a reduction of irrelevant results, and hopes to receive few but relevant links to information. From the search engine provider's point of view, personalization permits more targeted advertising, permitting advertisement subscribers to sell "personalised eyeballs". It would increase the click-through rate and possibly reduce the risk of click-fraud. From a technological point of view, personalization represents a trade-off between processing performance and speed of response. On the one hand, more user-feed generated information contributes to narrowing the search domain, and on the other, more information has to be processed. In view of the fact that audiovisual search and pattern recognition is highly process intensive, being able to introduce a query in natural language may slow down the response.

Customization may be at the level of individual search, like A9, Jeeves, Google, following collective search engines, like Eurekster, or by domains (search history, tagging, notes and diary sharing, etc.

31) http://www.clusty.com/ (last visited 6 June 2008).
32) http://www.kartoo.com/ (last visited 5 June 2008).
33) http://www.mister-wong.com/ (last visited 5 June 2008).
34) http://search.chacha.com/ (last visited 9 June 2008).
35) Battelle, John (Sept. 2005). *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture.* New York, ISBN 1-59184-088-0.

Practically all major search engines use a number of approaches to optimize search. A prominent way is to use cookies with unique IDs linked to the browser. It permits the server to keep records on the log of keywords, IP address, language, time and date, type of browser, URL of page requested, redirection, etc. Other relevant data could be drawn from various sources, namely desktop search, toolbar or location-based search. One risk is that this information may be linked and crossed with data originating from other applications and sources - for instance, combining the user's traces in Google Search with the information extracted from his Gmail account.

1.2.4. Towards Content-based Retrieval

AV search often refers specifically to techniques better known as content-based retrieval. These search engines retrieve audio-visual content relying mainly on pattern or speech recognition technology to find similar patterns across different pictures or audio files.[36] Pattern or speech recognition techniques make it possible to consider the characteristics of the image itself (for example, its shape and colour), or of the audio content. In the future, such search engines would be able to retrieve and recognise the phrase "red mountain" in a song, or determine whether a picture or video file contains a "red mountain", despite the fact that no textual tag attached to the files indicates this.[37] Content-based methods are thus indispensable when text annotations are non-existent or incomplete. They could also improve retrieval accuracy even when text annotations are present. The research challenge is how to improve multimedia retrieval using content-based methods. Here the performance of the retrieval process and the quality of the results depend on the type of content, as Lew et al indicate in an overview paper.[38]

Many concepts (i.e., algorithms) and systems have been developed for automatically structuring, indexing and finding *speech and audio content*. Possibly one of the first applications for spoken document retrieval was the "THISL broadcast news retrieval" system in the 90s.[39] Here, speech was transformed into text with the help of a large vocabulary continuous speech database. For each word of the automatically transcribed word sequences, a time code was attached. This allowed a search based on the usage of a standard text retrieval mechanism. Since then, progress has been steady as the benchmarking exercises of the US National Institute of Science and Technology (NIST) show. NIST carries out periodic benchmark exercises to gauge the quality of spoken document retrieval evaluations. Comparing the performance of indexing for broadcast news during the subsequent text retrieval conferences (TREC, 1997–2000) it became clear that it has become increasingly challenging to improve significantly the system quality. One reason is that the retrieval of speech recordings is very complex: The overall performance depends on a considerable number of factors, including the number of speakers, background noises, type of speech (broadcast or colloquial), etc., and each of them is a challenge in itself.

Equally difficult is the case of music retrieval. One of the first *music* indexing and retrieval systems was based on low-level features of audio processing.[40] These were mainly spill-overs invented for MPEG7 Audio standards and Audio-ID systems developed by several groups. The Audio-ID technology generates a fingerprint of a segment of music and provides fast matching algorithms to find this fingerprint in a large pre-processed archive. Recently, the focus of music retrieval has been changed to genre and mood classification.

36) Pattern or speech recognition technology may also provide a cogent way to identify content, and prevent the posting of copyrighted content. See Anick Jesdanun, *Myspace Launches Pilot To Filter Copyright Video Clips, Using System From Audible Magic*, Associated Press Newswires, 12 February 2007.

37) See Dr. Fuhui Long, Dr. Hongjiang Zhang and Prof. David Dagan Feng, Fundamentals of Content-Based Image Retrieval, at http://research.microsoft.com/asia/dload_files/group/mcomputing/2003P/ch01_Long_v40-proof.pdf (last visited: 6 June 2008).

38) Lew, M. S., Sebe, N., Djeraba, C., and Jain, R. 2006. *Content-based multimedia information retrieval: State of the art and challenges*. ACM Trans. Multimedia Comput. Commun. Appl. 2, 1 (Feb. 2006), 1-19. DOI= http://doi.acm.org/10.1145/1126004.1126005

39) D. Abberley, D. Kirby, S. Renals and T. Robinson, *The THISL broadcast news retrieval system*, Proc. of ESCA ETRW Workshop on Accessing Information in Spoken Audio, Cambridge (UK), April 1999.

40) Foote, J. 1999. *An Overview of Audio Information Retrieval.* ACM Multimedia Systems 7(1), 42-51.

For *video retrieval*, the research focus in the mid 90s was to find a way to detect object boundaries, and the distance between colour histograms corresponding to two consecutive frames in a video.[41] Since then, progress has been made in automatizing the process of boundary detection in an objective way, and a way has been found of using the motion within the video to determine the shot boundary locations, and performing semantic classification of the video shots into categories such as zoom-in, zoom-out, pan, etc.[42]

The search engine sector is currently thriving, and examples of beta versions across those various strands are available, both for visual and audio information. Tiltomo[43] and Riya[44] provide state-of-the-art content-based image retrieval tools that retrieve matches from their indexes based on the colours and shapes of the query picture. Pixsy[45] collects visual content from thousands of providers across the web and makes these pictures and videos searchable on the basis of their visual characteristics. Using sophisticated speech recognition technology to create a spoken word index, TVEyes[46] and Audioclipping[47] allow users to search radio, podcasts, and TV programmes by keyword.[48] Blinkx[49] and Podzinger[50] use visual analysis and speech recognition to better index rich media content in audio, as well as video, format. However, the most likely scenario for the near future is a convergence and combination of text-based search and search technology that also indexes audio and visual information.[51] For instance, Pixlogic[52] offers the ability to search not only the meta-data of a given image, but also portions of an image that may be used as a search query.

The following table summarizes some current technological approaches and their challenges in multi-media analysis together with current and future promising applications. It is split into different media formats, namely text, audio and video analysis. The audio domain is divided into speech and music, given the different application ranges and the related challenges. Some important research challenges cut across different domains, such as the previously discussed semantics and the information overflow.

41) Flickner, M. Sawhney, H. Niblack, W. Ashley, J. Qian Huang Dom, B. Gorkani, M. Hafner, J. Lee, D. Petkovic, D. Steele, D. Yanker, P. 1995. *Query by image and video content: the QBIC system,* IEEE Computer, September, 23-32.
42) Lienhart, R. 2001. Reliable Transition Detection in Videos: A Survey and Practitioner's Guide. International Journal of Image and Graphics 1(3), 469-486.
43) http://www.tiltomo.com/ (last visited: 6 June 2008).
44) http://www.riya.com/ (last visited: 6 June 2008).
45) http://www.pixsy.com/ (last visited: 6 June 2008).
46) http://www.tveyes.com/ (last visited: 6 June 2008); TVEyes powers a service called Podscope (http://www.podscope.com/) (last visited: 6 June 2008) that allows users to search the content of podcasts posted on the Web.
47) http://www.audioclipping.de/ (last visited: 6 June 2008).
48) See Gary Price, *Searching Television News*, SearchEngineWatch, 6 February 2006, at http://searchenginewatch.com/showPage.html?page=3582981 (last visited: 6 June 2008).
49) http://www.blinkx.com/ (last visited: 6 June 2008).
50) http://www.podzinger.com/ (last visited: 6 June 2008).
51) See Brendan Borrell, *Video Searching by Sight and Script*, Technology Review, 11 October 2006, at http://www.technologyreview.com/read_article.aspx?ch=specialsections&sc=personal&id=17604 (last visited: 6 June 2008).
52) http://www.pixlogic.com/ (last visited: 6 June 2008).

Table 1: Overview of the approaches, applications and challenges for content-based retrieval technologies by type of format.

Text	Speech	Music	Images	Video
Technological Approach				
- Named entity recognition, SVM, PLSI Bayesian semantic reasoning - Caption augmentation	- Speech recognition: spoken document retrieval, subword indexing - Speech segmentation: speaker clustering and recognition speech-to-video transcoding	Music segmentation: spectral flatness, genetic algorithms, music retrieval and recommendation	- Low-level image processing (histograms, shapes, textures, MPEG7-visual, SIFT), image similarity measurements - Relevance feedback	- Shot detection, keyframe generation - Object tracking based on motion-based features, closed captions recognition, etc. - Object detection and recognition - Video annotation and summarization - Video event detection
Applications				
- Classification of news and documents in companies - Email filtering - Text-based search engines, semantic analysis of multimedia (automatic) annotations	- Indexing of broadcast news/archives - Podcast/ videocast search (Potzinger, Blinkx) audio archives (Parliament data, historical archives)	- Indexing of music collections - Query by humming - Audio-music identification recommendation engines	- Content-based retrieval in image collections - Object recognition, face recognition (security, photo collections), automatic annotation of image collections with keywords and textual descriptions	- Indexing broadcast material, media observation, indexing of videocast material, recommendation engines, video fingerprinting, logo detection, security, etc. - 3D video
Challenges				
- Semantics, ontologies design (e.g., enrichment)	- Variability of content (e.g., background noise) domain and language dependency, scalability of subwords approaches	- Genre classification - Polyphonic instrument recognition, - Affective analysis	- Semantic gap, image segmentation, sensory gap	- Detection of complex concepts, thousands of different objects. Segmentation into more semantic-based units (i.e. complex scenes), multimodality, fusion

Source: Adapted from Sebe/Koehler[53]

53) N. Sebe, J. Koehler, "State of the art in audio-visual content indexing and retrieval technologies" in Chorus Report 2.1 "State of the Art on Multimedia Search Engines", p.29 *et seq.*, available at http://www.ist-chorus.org/documents/D21_final.pdf (last visited 10 June 2008).

2. Economic Considerations

The data explosion in the digital realm has rendered search a critical functionality. Data abundance is at the origin of search engine developments, but the fact that search is context dependent makes a difference in applications and their markets. The well-known web search engines offer a service to users to find publicly-available content on the Internet. The business model is predominantly supported by advertising. This is very different from search engines for business solutions, for instance. Here, search is mostly carried out in closed, proprietary, multimedia databases (e.g., national libraries). System providers deliver a retrieval tool specifically tailored to the needs of the customer company, adapted to their security restrictions, access rights and content securities, etc.

Although these retrieval tools do share similar technological challenges, the nature of their applications differs and so do their markets. Therefore, the four most prominent domains, namely web search, search for business solutions, peer-to-peer search and mobile search, will be discussed in the following paragraphs.

2.1. Web Search

Although close to one hundred search engines are operational, the bulk of the searches are performed by only a few service providers. According to the consultancy firm Nielsen/Netratings, the top three operators control more than 80% of the market. In particular, online searches by engine performed in the US in August 2007 were executed by Google 53.6%, Yahoo! 19.9%, MSN 12.9%, AOL 5.6%, Ask 1.7% and the rest 6.3%. These searches include local searches, image searches, news searches, shopping searches and other types of vertical search activity. More than 5.6 billion searches were carried out in that month alone (August 2007).[54] The ranking of the top three players is undisputed. According to comScore Networks, in December 2006 Google sites captured 47.3% of the US search market, Yahoo! 28.5% and Microsoft 10.5%. Americans conducted 6.7 billion searches in December 2006. With respect to the same month last year, this represents an annual increase in search query volume of 30%. This rate is considerable and explains why online advertising on search engines is expected to be a promising growth market.

European Internet users make as much use of search engines as their counterparts on the other side of the Atlantic. The intensive use of search engines explains why they are amongst the most visited pages on the Internet and attract a lot of traffic. Google is the most visited search engine in practically all countries of the European Union. For instance in June 2007, Google reached 88.8% of the UK, 69.5% of the French and 69% of the German online population. Google's audience is notably larger than for the Microsoft sites (83.3% UK, 62.3% France, 54% Germany) and Yahoo! (65.9% UK, 39.6% France and 36% Germany), according to the Internet audience measuring company comScore.[55] The search engine market consolidation becomes evident when observing the evolution of hits over longer time periods.

Figure 4 and Figure 5 show the evolution of the share for Germany and France, respectively. The evolution of Germany and France is similar to other European Member States. In particular, less than a handful of search engine providers have a market share of over 90% and Google has a much bigger share than its competitors.

54) See http://www.nielsen-netratings.com/
55) comScore Press releases, available at http://www.comscore.com/

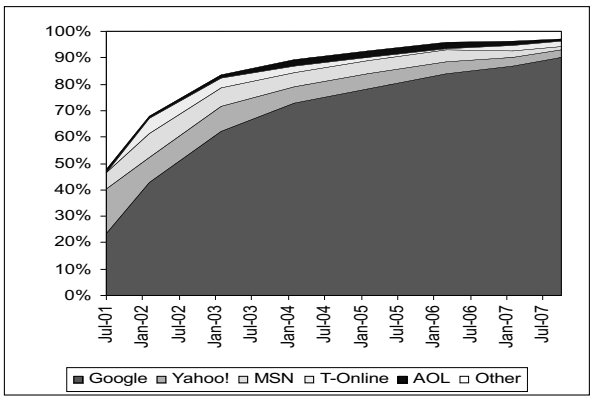


Figure 4 Evolution of WebHits for search engines in Germany in the period 2001-2007.
Source: WebBarometer,[56] H Speck[57] and own calculations.

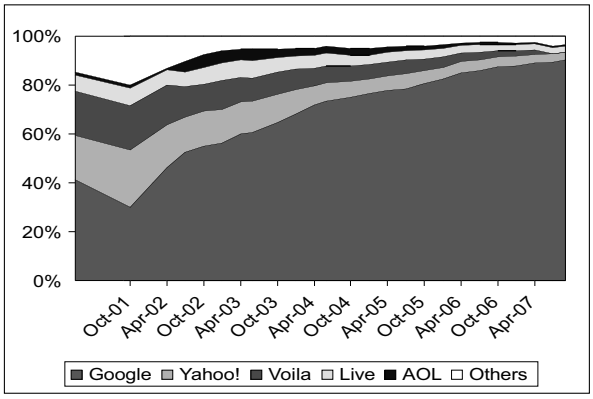


Figure 5: Evolution of WebHits for search engines in France in the period October 2001-September 2007.
Source: Baromètre Secrets2Moteurs[58] and own calculations.

These data highlight the fact that the search engine market is highly concentrated and also that search engines have become part of our lives. The average German, for instance, uses Google more than forty times a month[59] and three-quarters of Internet users access Internet offers through search engines.[60] Although the audience for the search engine providers may vary from one country to another, the user experience is similar for most western countries.

Consultancy firms calculating the market share, such as Nielson/NetRatings, Compete, Hitwise or comScore, measure the search behaviour of internet users by installing real-time meters on their computers (Nielsen states 500,000 people worldwide). The market share retrieved by these consultancy firms may differ to a certain extent for each of the search engines, due to the fact that they employ different metrics for measurement and the accuracy of the data is not sufficiently clear. This may partly explain why comScore's traffic data for Germany and France differs from the hit counts recorded by the German WebHits and the French Secrets2Moteurs.fr.

The concentration of the web search engine market appears to be a general trend in the USA and most EU Member States. There may be a number of reasons why Google is more dominant in Europe

56) http://webhits.de/
57) Hendrik Speck. *Search.Market*. 10th InetBib Conference. 9-11 April 2008, Würzburg (Germany) available at http://www.egs.edu/faculty/speck/files/presentation2008inetbibsearchenginemarkets.pdf (last visited 5 June 2008).
58) http://www.secrets2moteurs.fr/
59) comScore German data June 2007.
60) *Internetverbreitung in Deutschland: Potenzial vorerst ausgeschöpft?* Birgit van Eimeren, Heinz Gerhards and Beaste Frees, Media Perspektiven, Vol 8, pages 350 – 370.

than in the USA, including stronger branding, national marketing strategies, better adaptation to market size, better technological adaptation to language, lack of powerful national search engines, etc. An interesting case is Russia, even though it is not part of the European Union. Here, Google is only third by market share after Yandex[61] and Rambler. Yandex claims to have superior technology as it masters better the declensions and conjugations of the Russian language than other search engines. Other Slavic search engines, like the Czech Morfeo[62] or the Polish NetSprint,[63] also claim to have better technology. How much the Yandex high market share of over 55% in Russia can be attributed to better linguistic performance is, however, not obvious, as the same search engine provider achieves only 16% in the Ukraine, although the Russian and Ukrainian languages are, linguistically speaking, very close.

One factor that has favoured Google's dominant position today is the rate at which innovative services have been introduced. Many of these have been proposed to the audience at development phase (beta versions), rather than offering finished services to the users. This user involvement in the development stage is part of the company's culture of learning-by-doing. The company has benefited from using the dominant Internet language, English, when testing services and applications in the huge Anglo-Saxon environment, before introducing and adapting these to other cultures.

Operational Costs

The operational cost of providing web search is immense. Basically, these costs have two major components. The first is the cost of content crawling and indexing, which is practically proportional to the volume of the content. Digital data has grown immensely and there are no signs that the number of sites and data will not continue to grow immensely in the future. Consequently, the costs related to the gathering and indexing of content will also further increase. Crawling is a resource consuming task, but not a huge distinguishing factor amongst search engine providers having their own indexing system. As of 1 June, for instance, the 'modest' web search engine, Exalead, had crawled more than 8,064 million sites, similar to the number of sites crawled by the big search engines. The second, more important, component of the operational costs is related to the operational costs of processing user queries. Today's users are very demanding and expect quasi-immediate responses to their queries. The processing power needed to answer queries is proportional to the traffic, i.e., the user population. So too are the costs. For the big search engines, with a lot of traffic, these costs are immense. In a well-designed business, not only the costs, but also the revenues should be proportional to traffic and user population. Therefore, the traffic to a search engine could be gradually ramped up by balancing income and expenditure. However, this kind of organic growth would need to be supported by other factors like having the top search technology, a good brand name, well-designed advertising campaigns, financial stability etc., rendering it very difficult for a newcomer to catch up with the market leaders. This entry barrier will probably become more pronounced the more search moves from text-based search to audio-visual content retrieval, as it is technologically more demanding and the operational costs will further increase.

Progress in technology can contribute to reducing operational costs. These challenges include quicker algorithms, more efficient server architectures, cheaper hardware and less energy consuming processors. In fact, minimizing the power consumption of server farms is a real challenge. State-of-the-art server farms consume so much energy that cheap electricity and access to cooling (e.g., rivers, lakes) are among the most important criteria for their location.

Centrality of Search between Specialization and Integration of Services

The more search has become a central technology for citizens and businesses, the more stringent are the requirements in terms of performance in retrieving relevant content for the user. The need to deliver increasingly good relevance partly explains the concentration of general search engines and the emergence of thematic search engines. In the first case, only a few providers can maintain the pressure of high operational costs. These costs comprise the creation of server farms, their maintenance and operational costs, the development costs for better performance and also new adjacent services.

61) http://www.yandex.ru/
62) http://morfeo.centrum.cz/
63) http://www.netsprint.pl/serwis/

Therefore only a few general search engines are successfully operating world-wide. In view of the fact that even more processing power for audiovisual content is needed than for text, the entry barrier is unlikely to be lowered for new incumbents of future audiovisual search engines. A consolidation around a few general web search engines is likely. An alternative is specialisation in a niche area. This is the case for the thematic search engines. They provide relevant retrieval in specific domains and the entry barrier is more modest that for general web search engines.

Search engines are becoming increasingly integrated with other applications and services. Major search engines offer a package of free services, such as email, chatting, instant messages, voice services, etc. Many of these services are not profitable individually, but make sense at a corporate level to keep users within the "realm" of the service engine operator. As search engine providers have an interest in increasing the "stickiness" of their users, they try to personalise services. Through personalization, they are more likely to be able to anticipate queries and this will enable search engines to operate proactively. Rather than just "pulling" information, they could also "push", information and thus target advertisements even better.

2.2. Business Search Solutions

In early times, company investments were largely devoted to the establishment of an efficient IT infrastructure. With time, companies have gathered the necessary resources and technologies to capture, store and transfer the information the enterprise needs for its operation. One remaining bottleneck is to provide a consolidated user-centred view for employees to make their jobs easier and allow them to be more efficient. This shift from a basically storage-oriented infrastructure to information consumption goes along with a user-centric model rather than a technology-based one. Providing an efficient, interactive and secure way to present user-specific content is complex, because it has to take into account different operational systems, file formats, schemas, etc. One example of business search solutions is intelligent tools for providers of yellow (and white) telephone pages services. These business search tools first retrieve information from the companies' databases, but can cross-check them with other sources. Through analysis and comparison, data that is missing or inconsistent can be detected. Another example is in the healthcare sector. Medical records, images, radiographies, laboratory analyses, computer tomography and other relevant medical information is increasingly "born" digital. There is a need for automatic tools to retrieve, analyse and interpret such medical information.

Therefore, tailored search solutions for business and enterprise are an emerging field. Identifying and enabling specific content across an enterprise to be indexed, searched and displayed to authorized users gives added value to companies. Following a report by the consulting firm IDC, the worldwide market for enterprise search and retrieval software in 2005 was USD 976 million. This had grown by 32% with respect to the previous year. The size of this sector is notably smaller than the aforementioned web search advertisement market. The three big players, Google, Yahoo! and Microsoft have some activity in this field, but their revenues from licensing technology are modest compared to their core business. In contrast with the web search market, the business solution market is fragmented and there are a number of highly specialised companies on the market. There are also some European companies: among the most prominent of these are the Norwegian company FAST (recently acquired by Microsoft), Autonomy in the United Kingdom or Expert System SpA in Italy. These companies sell knowledge management tools where the search function is an increasingly important software module. With these tools companies hope to effectively retrieve information and also to uncover meaning arising from any enterprise information including documents, e-mails, entries in relational databases, etc.

Today, the market for 'knowledge management' tools is very distinct from the web search engine market. The future Internet, however, will become less text-based and more audio-visually driven. The more web search engines need to deal with audio-visual content, the more the technological interest in the need to develop solutions for conceptual search, document classification, text mining, and information analysis and correlation will overlap. This may drive current web search engines to penetrate the "knowledge management" market further. The fact that Microsoft recently acquired FAST may be an indicator of this trend.[64]

64) http://www.01net.com/editorial/368946/microsoft-s-achete-la-place-de-numero-un-de-la-recherche-en-entreprise/

2.3. Mobile Search

Mobile Search refers to information retrieval services accessible through mobile devices like phones or PDAs. European telecom operators already provide some search options for their 2G, 2.5G and 3G services. For these, they rely on technology provided by external companies like Google or FAST and many search portals offering a dedicated interface for handheld services, like MetaGer.

Although prospects for business models are not yet consolidated, the mobile search market is likely to differ significantly from the web search engine market. The technological context (e.g., small screens, limited bandwidth), the reduced amount of suitable content for mobile devices, the role of the market players (e.g., as telecom operators to the Internet by mobiles have a more powerful role as providers than Internet service providers have for accessing the internet via a computer), user behaviour (e.g., type of search requested on the move), may demand a different search engine business model. Walled-garden markets seem to be the prevailing model, but the market may become more open in the future. The possibility of flat-rate pricing is being discussed and bandwidth restrictions may force payment by bit download. This would make a difference not only for bandwidth-intensive downloading such as video (e.g., there may be pricing by video per resolution), but also for location-based services, which are regarded as being very promising and would allow us to find the nearest restaurant by typing the question, or simply speaking, into our mobile telephone.

Although the market is in its infancy and fragmented, the trends are highly promising and closely linked to the number of subscribers to mobile broadband connections. Although still tiny in absolute terms, the growth rates are impressive. The GSM Association reported 32 million mobile broadband connections in March 2008, up from only 3 million broadband connections in March 2007![65] More importantly, mobile Internet and search is an expanding market. Its main attractions are local information, like weather, maps or directions. It is estimated that in 2008 there are 2.5 billion mobile users world-wide, of whom roughly 40% have 2.5 G and 10% 3G technology. As the number of mobile subscribers is still increasing world-wide (particularly in highly populated, less developed countries where the penetration rate is modest) and the share of 2.5G mobiles or higher will be increasing, the number of subscribers will also increase. Today, 489 million people have access to mobile Internet (not necessarily broadband), and this may double by 2011. As the vast majority of the mobile Internet users will also be searching, they are potential customers for search engine providers. In fact, eMarketer considers that the mobile search advertisement market has reached USD 221 mio. in 2008 and will increase to USD 2,361 mio. by 2011.[66]

For this to come about, however, a number of challenges, which arise from the fact that mobility imposes specific requirements with regard to user interaction, retrieving data and displays, need to be resolved. Another issue is adapting or creating content suitable for mobile devices. Content search particularly will have added value when content is adapted to the user and combined with other technologies, e.g., location-based services. This will render the search experience more personalised.

Expanding the Search Experience

Search tools are tailored to retrieve information from specific platforms. Current web search is mostly performed on personal computers or laptops which access one of the more than 70 web search engines world-wide. One challenge is to expand the search capability from other platforms, particularly for mobile devices. Technical possibilities for platforms other than computers are still in their infancy. For instance, for mobile search, devices have technical limitations and business models are not consolidated. The ideal from the users' point of view would be full interoperability, as their sole interest is to retrieve relevant content irrespective of the technology.

65) www.gsmworld.com
66) www.eMarketer.com

3. Summary

Current retrieval tools (mostly text-based) are key technologies for the Information Society. Given the explosion of audiovisual data, future AV search engines will become even more central to society than they are today. The provision of relevant results in AV search is far more complex than in text-based search, and progress in AV search will depend on new innovations and also on improving existing concepts. For instance, retrieving relevant audio-visual material will benefit from new content-based search (audio, video, images) but it will also rely largely on meta-data (text) concepts. The technological challenges range from basic science to development tasks. Semantic approaches for search or novel technological concepts to master the generation and flow of huge amounts of data are examples of fundamental research tasks. Interactive search concepts, relevance feedback systems, multi-modal analysis or improved retrieval algorithms are examples of bottlenecks that must be removed in order to improve current applications.

Technology and business considerations go hand in hand. How to deal with the processing, storage and traffic of huge amounts of audiovisual data is a technological challenge that will require new computer architectures and distributed search solutions. There are also important financial implications that could influence considerably the entry barriers for newcomers onto the AV search market. The deployment of AV search technology is, therefore, likely to reinforce many of the current techno-economic trends of the web search market. Examples are the concentration effect of general-purpose web search engines, or the emergence of thematic search networks. At the moment, there is a clear distinction in the AV search market with regard to web search, business solutions or mobile search, due to the nature of the industry, the client structure and the business models. In the future, these differences may become blurred. Web search engine providers are already starting to acquire companies offering business solutions and non-walled garden business models may emerge in the mobile search sector.

The Future of Content Navigation

*Janet Greco**
Broadcast Projects, London

Introduction

As consumers are increasingly faced with a multitude of audio-visual and other content choices vying for their attention, how can they be sure they are not missing the content that really matches their interests? In this increasingly complex scenario content navigation mechanisms such as Electronic Programme Guides and search engines are key.

1. Electronic Programme Guides at a Crossroads

It used to be simple to find out what was on television: a quick check of the relatively few channels listed in the daily newspaper or the local weekly TV guide sufficed. Printed guides provided the titles and details of scheduled programmes in a simple to read, column format, organized by time. Additional editorial commentary and a photo flagged noteworthy programmes of the day. As the number of channels began to increase, printed "grid" format TV listings were introduced, and by the mid 1980s they became the preferred and ultimately dominant format. Since the introduction of Electronic Programme Guides (EPGs) for digital television systems in the mid-1990s, the grid format has continued to be the main structural building block for content navigation in the pay-TV environment. Over the last ten years EPGs have become the essential tool for navigating content in the mass market TV landscape. But with so many channels, insufficient "screen real estate" and no way to convey much of the useful detail that could better inform decisions, grid-format guides have reached their limitations.

More recently, Internet-based tools have become commonplace for locating content beyond the television set. With the proliferation of Tivo-like devices and Personal/Digital Video Recorders (PVRs), we have changed our habits to navigating what is "always on" or "always available". Choices have increased many times over, ranging from professionally produced mainstream entertainment to the avalanche of "user-generated content", thanks to video-sharing phenomena like YouTube. The key challenge has now fundamentally changed: how will consumers locate the content relevant to their personal tastes? Might more active participation and comments from users of social networks fundamentally affect how people share and receive recommendations?

Most people probably have not thought much about the effort it takes for specialized intermediaries (e.g., TV "data aggregators" and press agencies) to obtain regularly and update the programme schedules that they receive from broadcast TV channels. Today, this information is still disseminated

* Janet Greco is founder and principal consultant of Broadcast Projects (www.broadcastprojects.com).

by the marketing departments of broadcasters and channels in a variety of data formats. It is useful to think of these "TV listings" in terms of how you might import names and addresses into your contact management software: the field names have to be compatible and the field characteristics have to match (e.g., you cannot import characters into a "numeric" field, nor would it be of much use to import a person's name into the "country" field). Most of the time, only narrative documents are sent, so these cannot be easily imported into a database. Moreover, publishers need one stream of information for a multitude of channels, in a single data format that supports their specific system requirements. They cannot manage multiple inbound formats coming from each and every channel.

So even though international "metadata standards" exist, it is extremely difficult to ensure that multiple providers would spell, classify and verify this descriptive information in a consistent way. As a result, the data aggregation intermediaries that perform this vital data processing, standardization and verification role remain important players in the dissemination of this information to professional publishers (online and in print).

The main critical flaw in this process is that the intermediary is not directly linked (except in rare cases) to the transmission and broadcast play-out centres of the channels. Therefore, there is a natural disconnect between the data disseminated by the data aggregators and the actual broadcast transmission, which can give rise to errors. Nevertheless, the data harmonization functions they perform will continue to ensure that these companies will play an important role. The complicating factors are, of course, the proliferation of online and on-demand content, without any further need for "scheduled programmes", and finding a way to offer consumers a more accurate and personalized navigation system, covering all forms of available content, no matter what the source.

Addressing how consumer-facing metadata can be better organized requires significant organizational change. But it is not often easy to change work habits within the professional broadcast environment, which has been making a huge effort to migrate to digital file-based systems. Moreover, the international metadata standards (such as TV Anytime), that have been so diligently and comprehensively created, have resulted in a very large number of possible "tags" that could be used to annotate the content that would facilitate search and navigation.

2. The Problem with Metadata Standards

With such standards, there is an expectation that multiple content producers, TV channels, broadcasters and others would tag the information in a consistent way, enabling the seamless interchange of information amongst a vast array of services and devices. In this context it is interesting to note that where "tagging" occurs most frequently and consistently is online and by consumers, who originated for example the Internet Movie Database,[1] and the Compact Disc Database[2] (the system that drives iTunes metadata). User-tagged content, often referred to as "folksonomies", offers an interesting angle to consider in the context of this huge job of tagging content. Such systems are often self-correcting, in that intelligent search systems can determine similar groupings of words to interpret the intention of a search request. Metadata will remain a key driver of search engine results. Video search innovators are attempting to address and work around these issues in many varied ways. But the simple reality is that they are attempting to fill in, with very advanced technology, what is missing in the simple form of consistent metadata. As an extension of the social-networking focused Web 2.0, the semantic web and beyond, it may be more likely that consumers help with the job of creating metadata once, to be consistently maintained in perhaps a "wikipedia"-like way.

The challenge today is that major content owners do not agree with (or perhaps do not understand) the concept that the data that describes their content should be freely available to consumers and inter-operable with other systems. This is mainly a point of education: metadata is like product packaging. But product packaging in the digital marketplace is more a matter of data distribution and processing (an Information Technology issue) than it is about "marketing" or "advertising" in the traditional splashy way it has always been understood.

1) http://www.imdb.com
2) http://www.cddb.org

International standards that deal with the unique identification of audio-visual works (e.g. ISAN: International Standard Audiovisual Number, www.isan.org) exist and it is probably in the public interest that works are registered once and correctly. ISAN, however, requires payment for the registration of all audio-visual works, whereas the music recording industry took an entirely different approach and, like ISBN numbers for books (which is also a much older, but also paid-for system) it is unheard of for major labels not to uniquely identify recordings. Under the right conditions, the application of ISAN (or any other unique identification system) could considerably un-complicate this story. But even if the system amounted to the equivalent of a Universal Product/"bar code" with which we are all familiar, the challenge persists of changing work habits to integrate those unique codes into the large number of existing databases!

Many individual silos of information remain today that replicate the same descriptive information on audiovisual content over and over again. These include the data aggregators, content owners, copyright societies, audience research companies, TV guides and many others. While programme synopses may differ according to the language, writer or promoter of a piece of content (such as the descriptive summary of an episode of "Grey's Anatomy'), other "static" details need not be maintained in separate databases. These include items of metadata that never alter: the original language title and episode title, the names of the actors, the director, producer, and date of production. So a unique identifier makes economic sense, as there is so much duplication of effort in terms of maintaining these databases.

So the challenges are many, and new legal issues are now added into the mix. With the pending implementation of the EC Audiovisual Media Services Directive (by December 2009), and considering the vast array of other European regulatory and policy initiatives, new requirements for mandatory content labelling for Television and "Video on Demand" services are emerging. New child protection initiatives such as the UK's Byron Review will result in the creation of a UK Council on Child Internet Safety (by September 2008). Its recommendations, which are wholly supported by the UK government, already provide one clear example of how new content labelling requirements will affect businesses. In this case, video games in the UK must now carry the British Board of Film Classification's (BFFC) labelling system. The BBFC also provides another interesting piece of "metadata" in the form of a content advisory synopsis, which describes in quite some detail the content of a video game (or movie, etc). While other self-regulatory initiatives by video games producers such as PEGI[3] sufficed until now as the labelling mechanism to be employed in the UK, the introduction of the BBFC system under UK law represents just one example of how conflicting labelling systems will ultimately affect the implementation of effective and coherent content navigation systems. The implementation of this will also become increasingly complex for content distribution businesses working across multiple territories and that will be required to adhere here to multiple legal regimes.

3. The Importance of Consistent Data Input

In the meantime, consumers increasingly struggle to identify content of real interest across a variety of platforms. Media companies and consumer electronics firms now seek to fill in the gaps with new "advanced guide" products and "content navigation" services. But today these only work well in a walled-garden environment, where the data set for the EPG is clearly defined (e.g., 500 channels on BSkyB, plus VOD and Sky Anytime – this is a manageable set of data).

However, merging such "walled gardens" with Internet content is an entirely different matter. On the Internet, video search engines employ many technologies to identify content – e.g., those engines identify what is *not* there in the form of consistent metadata (by analysing audio, subtitle tracks, face recognition, etc). And where watermarking/fingerprinting technology is used, these are not yet fully linked up with a robust content metadata service (linking to any form of Unique IDs) – e.g., these represent additional silos of information.

It would be much easier if every piece of content were identified - once - and could then be digitally accessible by all. Audio-visual content could be described once in much the same way as a word is

3) Pan European Game Information (www.pegi.info).

identified and described once each time in a dictionary. Static elements do not change (e.g., the grammatical form: noun, verb, etc) in much the same way as for audiovisual content: title, director's name, country of origin, date of production, etc, do not change either. Descriptive (subjective) text could be handled on the basis of links to the main record (different language versions, different synopsis lengths, different editorial perspectives, etc).

The semantic web promises a new generation of information services. XML[4] mark-up formats like TV Anytime[5] require a great deal of effort to implement. This is but one of many metadata specifications. For it all to work well everyone has to do it and the data input has to be consistent (imagine someone in Germany tagging a programme using the word "*Komödie*"[6], while using the term "sitcom" in the US to describe the same programme).

For every good argument to support metadata, tagging and content navigation moving in this direction, there will also be corresponding counter-arguments. The potential commercial motivation for content owners to open up suddenly their databases freely to the world would have to imply:

- a better understanding of the product packaging role of metadata in an IT environment (as opposed to its traditional marketing function);
- a willingness and an ability to find an interoperable solution for the protection of content (e.g., interoperable digital rights management or DRM).

For now, the acquisition and use of professional – robustly searchable – consumer-facing metadata is a service that has to be specifically commissioned and paid for today, unless we are content to leave the job to the search engines.

4. Future Content Navigation Systems

In "search", the online world continues to have the upper hand. Search engines are far more advanced and able to quickly (if sloppily) pinpoint content by the mere entry of a single word or phrase. And even more pertinent is the fact that the younger generation is accustomed to and highly adept at manipulating both software and services to find what they want online and multitask while doing it. The challenge is plain for the traditional pay-TV operators and, when you compare the interfaces that exist in the Set-top-box-driven pay TV world, the gap is staggering.[7]

Advanced "guide" products are coming, but accessing consistent metadata across both online and broadcast sources remains a huge challenge. The future of content navigation is heavily dependent upon gaining access to a rich source of accurate and consistent metadata. Especially important are:

- One master record
- Uniquely identified
- Accessible by all (industry and the public)
- Correct in spelling, grammar and punctuation
- Multi-lingual
- Extensively tagged
- Extensible - so that user generated comments and descriptions (and photos, videos, etc) can be appended using the unique ID as anchor.

It remains to be seen how the market for metadata will evolve to enable new content navigation services to emerge. Advanced guide products have already been designed, but have not yet been deployed because the issue of accessing robust metadata across both broadcast and Internet sources is still so difficult.

4) XML (Extensible Markup Language) is a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets and elsewhere.
5) www.tv-anytime.org
6) In English "comedy".
7) For references see www.epgmetadata.com

Even the most professional metadata aggregation companies cannot cope with the increasing requirements that the industry now makes (DVD cover art, photos, video clips, keywords, international film ratings systems, content advisories and access information for the disabled, etc). And labelling requirements will increase as regulatory policies evolve. The way forward relies upon education, incentives, tools and public sector support to create an environment where the digital economy can thrive. In such a scenario, a unique array of new products and services can find their way onto the market.

Privacy and the Drive for the Perfect Audiovisual Search Engine

Michael Zimmer
School of Information Studies
Center for Information Policy Research
University of Wisconsin-Milwaukee

Introduction

Web search engines have emerged as ubiquitous and vital tools for the successful navigation of the growing online information sphere. The world's largest search engine, Google, captures the goal of all general-purpose search engines in its mission statement: to "organize the world's information and make it universally accessible and useful".[1] To this end, Google, like most search providers, strives to create "a perfect search engine [that] will process and understand all the information in the world"[2] and understands "exactly what you mean and give back exactly what you want".[3] The drive for such a perfect search engine has resulted in the rise of powerful search tools that offer access to tens of billions of online information sources, processing millions of queries every day, and providing relevant and often personalized results – all for free.

With the rise of increased bandwidth speeds, cheaper digital technologies, the growing ubiquity of webcams and cellphone cameras, and innovative websites for content storage and social interaction (such as Flickr or Facebook), the Web is increasingly becoming a platform for the creation, storage and sharing of user-generated multimedia content. As a result, search engine providers strive to expand the power and reach of their crawlers to include image, audio and video content in their searchable indexes, offer their own platforms for the storage, retrieval and sharing of images and videos, and perhaps specialize in the delivery of multimedia content exclusively (such as the European search engine project, Quaero). In short, the perfect search engine is increasingly becoming audiovisual.

The growing power and ubiquity of search engines have not escaped controversy, including concerns about search engine bias, censorship (or its circumvention), copyright and trademark infringement, and free speech online. Perhaps the most potent, however, are the privacy concerns related to power search tools.[4] By striving to make audiovisual content as "universally accessible and useful" as its

1) Google. (2005). *Company overview*. Retrieved 3 May 2006, from http://www.google.com/corporate/index.html
2) Google Press Center. (7 June 1999). *Google Receives $25 million in equity funding* [press release] (June 1999) http://www.google.com/press/pressrel/pressrelease1.html (accessed 18 August 2006)
3) Google. (2007). *Our philosophy*. Retrieved 27 March 2007, from http://www.google.com/intl/en/corporate/tenthings.html
4) See, for example, Sullivan, D. (2 April 2003). *Search privacy at Google & other search engines*. Retrieved 31 March 2007, from http://searchenginewatch.com/showPage.html?page=2189531; Associated Press. (17 July 2005). *Google Growth yields privacy fear*. Retrieved 31 March 2007, from http://www.wired.com/politics/security/news/2005/07/68235; Mills, E. (3 August 2005). *Google balances privacy, reach*. Retrieved 7 January 2007, from http://news.com.com/Google+balances+privacy,+reach/2100-1032_3-5787483.html; Zimmer, M.(2008). *Privacy on planet Google: Using the theory of "contextual integrity" to clarify the privacy threats of Google's quest for the perfect search engine.* Journal of Business & Technology Law, *3*(1), 109-126; Zimmer, M. (2008) *The gaze of the perfect search engine: Google as an infrastructure of dataveillance*. In A. Spink, & M. Zimmer (Eds.), *Web searching: Multidisciplinary perspectives*. (pp. 77-99). Dordrecht, The Netherlands: Springer.

textual cousins, the emergence of powerful audiovisual search tools has renewed concerns over these issues among academics and regulators alike.[5] Because of the unique nature of audiovisual content, the privacy-related issues that stem from the drive for a perfect audiovisual search engine deserve particular attention. This contribution will describe some of the latest trends in audiovisual search technologies – such as the integration of user-generated images into searchable indexes, rich photo tagging and metadata, and user-targeted face recognition tools – and bring into focus the threats to privacy they engender.

1. The Drive for the Perfect Search Engine

Since the first search engines started to provide a way of interfacing with the content on the Web, there has been a drive for the "perfect search engine," one that indexes all available information and provides fast and relevant results.[6] A perfect search engine would deliver intuitive results based on users' past searches and general browsing history.[7] For example, the perfect search engine would know whether a search for the phrase "Paris Hilton" is meant to help a user locate hotel rooms in the French capital or obtain the latest celebrity gossip about the young socialite, and deliver only the appropriate results in each case. Search engine companies have clear financial incentives for achieving the "perfect search": receiving personalized search results might contribute to a user's allegiance to a particular search engine service, increasing exposure to that site's advertising partners as well as improving chances that the user would use fee-based services. Similarly, search engines can charge higher advertising rates when advertisements are accurately placed before the eyes of users with relevant needs and interests (i.e., someone seeking travel information rather than celebrity gossip).[8]

Journalist John Battelle[9] summarizes how such a perfect search engine might work:

> "Imagine the ability to ask any question and get not just an accurate answer, but your perfect answer – an answer that suits the context and intent of your question, an answer that is informed by who you are and why you might be asking. The engine providing this answer is capable of incorporating all the world's knowledge to the task at hand – be it captured in text, video, or audio. It's capable of discerning between straightforward requests – who was the third president of the United States? – and more nuanced ones – under what circumstances did the third president of the United States foreswear his views on slavery?"

This perfect search also has perfect recall – it knows what you have seen, and can discern between a journey of discovery – where you want to find something new – and recovery – where you want to find something you have seen before.

When asked what a perfect search engine would be like, Google's Sergey Brin replied, "like the mind of God".[10] To attain such an omnipresent and omniscient ideal, the perfect search engine must have both "perfect reach" to provide access to all available information on the Web, and "perfect recall" to deliver personalized and relevant results that are informed by the previous habits of that particular searcher.

---

5) For example, in September 2007, the Jan van Eyck Academie (Maastricht, The Netherlands) held a critical forum on the European audiovisual search project Quaero, see http://www.janvaneyck.nl/0_2_3_events_info/arc_07_quaero_thinktank.html. This was followed in April 2008 by an expert workshop on the regulatory challenges faced by audiovisual search engines, hosted by the Institute for Information Law at the University of Amsterdam.

6) See Kushmerick, N. (23 February 1998). *The search engineers*. The Irish Times, p. 10; Andrews, P. (7 February 1999). *The search for the perfect search engine*. The Seattle Times, p. E1; Gussow, D. (4 October 1999). *In search of…* St. Petersburg Times, p. 13; Mostafa, J. (24 January 2005). *Seeking better Web searches*. Retrieved 30 January 2005, from http://www.sciam.com/print_version.cfm?articleID=0006304A-37F4-11E8-B7F483414B7F0000

7) Pitkow, J., Schütze, H., Cass, T., Turnbull, D., Edmonds, A., & Adar, E. (2002). *Personalized search*. Communications of the ACM, 45(9), 50-55; Teevan, J., Dumais, S. T., & Horvitz, E. (2005). *Personalizing search via automated analysis of interests and activities*. Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 449-456.

8) Hansell, S. (26 September 2005). *Microsoft plans to sell search ads of its own*. The New York Times, pp. C1, C8.

9) Battelle, J. (8 September 2004). *Perfect search*. Searchblog. Retrieved 16 May 2006, from http://battellemedia.com/archives/000878.php

10) Ferguson, C. (2005). *What's next for Google?* Technology Review, 108(1), 38-46.

1.1. Perfect Reach

To achieve the reach necessary for the realization of Search 2.0, web search engines amass enormous indexes of the Web's content. Expanding beyond just HTML-based Web pages, search engines providers have indexed a wide variety of media found on the Web, including images, video files, PDFs and other computer documents. For example, in 2005 Yahoo! claimed to have indexed over 20 billion items, including over 19.2 billion Web documents, 1.6 billion images and over 50 million audio and video files.[11] Furthermore, search engines are increasingly digitizing the "material world," adding the contents of popular books, university libraries, maps and satellite images to their growing indexes. In addition to their expansive and diverse searchable indexes, today's search engines also obtain a "perfect reach" by developing various tools and services to help users organize and use information in contexts not considered traditional Web searching. These include communication and social networking platforms, personal data management, financial data management, shopping and product research, computer file management and enhanced Internet browsing.

Combining these two aspects of the perfect reach – expansive searchable indexes and diverse information organization products – the perfect search engine enables users to search, find and relate to nearly all forms of information they need in their everyday lives. The reach of the perfect search engines allows users to search and access nearly all content on the Web, and also enables them to communicate, navigate shop, and organize their lives, both online and off.

1.2. Perfect Recall

Complementing the perfect *reach* of the perfect search engine is the desire of search engine providers to obtain perfect *recall* of each individual searcher, allowing the personalization of both services and advertising. To achieve this perfect recall, web search engines must be able to identity and understand searchers' intellectual wants, needs and desires when they perform information seeking tasks online. In order to discern the context and intent of a search for "Paris Hilton," for example, the perfect search engine would know if the searcher has recently shown interest in travel to France, or whether she spends significant time online searching for celebrity gossip.

The primary means for personalizing search results is to rely on users' search habits and history.[12] To gather users' search histories, most Web search engines maintain detailed server logs recording each Web search request processed through their servers, the IP address making the request, the pages viewed and the results clicked.[13] Search engines also rely heavily on Web cookies to help differentiate users and track activity from session to session, and increasingly push the creation of user accounts to help associate particular users with their online activity. The motivation behind gathering this user information is explained to the users in terms of improving their search experience. Google, for example, states, "We use this information to improve the quality of our services and for other business purposes",[14] while the search engine Ask.com also presents its economic motivations fueling the need for this perfect recall in pursuit of the perfect search: "We collect…anonymous information to improve the overall quality of the online experience, including product monitoring, product improvement, targeted advertising, and monetizing commercially oriented search keywords".[15] Given the expansive array of products and services offered by search engine providers, their server logs potentially contain much more than simply a user's Web search queries.[16] Search queries for various sorts of online content – images, news stories, videos, books, academic research, products, music, and so on – are almost

11) Mayer, T. (8 August 2005). *Our blog is growing up – and so has our index*. Yahoo! Search Blog. Retrieved 25 November 2006, from http://www.ysearchblog.com/archives/000172.html

12) See, for example, Speretta, M. (2000). *Personalizing search based on user search histories.* University of Kansas; Pitkow, J., Schütze, H., Cass, T., Turnbull, D., Edmonds, A., & Adar, E. (2002). *op.cit.*; Teevan, J., Dumais, S. T., & Horvitz, E. (2005). *op.cit.*

13) See, for example, Google. (2005). *Google Privacy faq*. Retrieved 3 May 2006, from http://www.google.com/privacy_faq.html; IAC Search & Media. (13 July 2005). *Privacy policy for Ask.com*. Retrieved 6 January 2007, from http://sp.ask.com/en/docs/about/privacy.shtml ; Yahoo! (11 November 2006). *Yahoo! Privacy policy*. Retrieved 6 January 2007, from http://info.yahoo.com/privacy/us/yahoo/details.html

14) Google. (2005) *Google Privacy faq. Op. cit.*.

15) IAC Search & Media. *Op.cit*.

16) Wikipedia lists over 100 different products and services offered by Google, and over 50 by Yahoo!

certainly logged, as well as links clicked and related usage statistics from within non-search related services provided by search properties, such as e-mail, discussion groups, mapping services and the like.

In summary, the drive for the perfect search engine has led search providers to expand their reach to index and make searchable a wide variety of information sources (both online and off), as well as enhancing their recall of particular users' habits, needs and wants. Considering the steps taken to achieve the perfect reach and perfect recall noted above, search engines seem to be approaching this omniscient and omnipresent ideal and are poised to "become the perfect, all-seeing, all-knowing, all-powerful force of the 21st century".[17]

2. Privacy and the Perfect Search Engine

The drive to create the perfect search engine – one that promises breadth, depth, and relevancy – is not completely benign. The path leading to the perfect search engine has been fraught with concerns, including bias, censorship (or its circumvention), copyright and trademark infringement, and free speech.[18] Perhaps most potent, however, are the privacy concerns about the drive for the perfect search engine.[19] Both the reach and recall necessary to achieve the perfect search engine have particular implications for personal privacy.

2.1. Privacy and the Perfect Reach

Achieving the "perfect reach" requires search engines to index as many Web pages and other online sources as are necessary to provide the largest possible database of potential search results. Among the billions of pages indexed by search engines are Web pages containing personal information about individuals, such as personal homepages (both active and abandoned), discussion forum postings, online resumes, minutes of public meetings, property tax records and court records. Few people are not affected by the "long arm of Google's Web crawler," explains journalist Neil Swidey, referring to the broad scope of Google's Web index:

> "Maybe it was a stupid fraternity prank or a careless posting to an Internet newsgroup in college. Perhaps you once went on a rant at a selectmen's meeting or signed a petition without stopping to read it. Or maybe you endured a bitter divorce. You may think those chapters are closed. Google begs to differ.
>
> While most of your embarrassing baggage was already available to the public, it was effectively off-limits to everyone but the professionally intrepid or supremely nosy. Now, in states where court records have gone online, and thanks to the one-click ease of Google, you can read all the sordid details of your neighbor's divorce with no more effort than it takes to check your e-mail."[20]

Engaging in a "vanity search" – a Web search for one's own name – can reveal a surprising amount of personal information: tax assessments, court documents, marriage licenses, deeds and voter registration information, for example. The notion of "Googling" someone before a blind date has

17) Ayers, C. (1 November 2003). *Google: Could this be the new God in the machine?* The Times, p. 4.
18) See, for example, Grimmelmann, J. (forthcoming). *The structure of search engine law.* Iowa Law Review; Introna, L. & Nissenbaum, H. (2000). *Shaping the Web: Why the politics of search engines matters.* The Information Society, 16(3), 169-185; Elkin-Koren, N. (2001). *Let the crawlers crawl: On virtual gatekeepers and the right to exclude indexing.* University of Dayton Law Review, 26, 180-209; Gasser, U. (2006). *Regulating search engines: Taking stock and looking ahead.* Yale Journal of Law & Technology, *9*, 124-157; Goldman, E. (2006). *Search engine bias and the demise of search engine utopianism.* Yale Journal of Law & Technology, 188-200.
19) See, for example, Sullivan, D. *op.cit.*; Associated Press. *op.cit.*; Mills, E. *op.cit.*; Zimmer, M.(2008). *Privacy on planet Google. Op.cit.*; Zimmer, M. (2008) *The gaze of the perfect search engine. Op.cit.*.
20) Swidey, N. (2 February 2003). *A nation of voyeurs: How the internet search engine Google is changing what we can find out about one another - and raising questions about whether we should.* The Boston Globe Sunday Magazine, p. 10.
21) Lobron, A. (5 February 2006). *Googling your Friday-night date may or may not be snooping, but it won't let you peek inside any souls.* The Boston Globe Magazine, p. 42.

become common practice.[21] Almost one in four Web users have searched online for information about co-workers or business contacts,[22] and employers are Googling prospective employees before making hiring decisions.[23] In less than an hour, one reporter uncovered a variety of personal details of Google CEO Eric Schmidt's life:

- Schmidt doesn't reveal much about himself on his home page. But spending 30 minutes on the Google search engine lets one discover that Schmidt, 50, was worth an estimated USD 1.5 billion last year. Earlier this year, he pulled in almost USD 90 million from sales of Google stock and made at least another USD 50 million selling shares in the past two months as the stock leaped to more than USD 300 a share.

- He and his wife Wendy live in the affluent town of Atherton, Calif., where, at a USD 10,000-a-plate political fund-raiser five years ago, presidential candidate Al Gore and his wife Tipper danced as Elton John belted out "Bennie and the Jets."

- Schmidt has also roamed the desert at the Burning Man art festival in Nevada, and is an avid amateur pilot.[24] [25]

The expansion of both the depth and breadth of their searchable indexes to obtain the "perfect reach" has reduced users' sense of "privacy through obscurity".[26] A massive amount of information about individuals, which was once difficult to find and even more difficult to cross-reference, is now readily accessible and collectible through the use of search engines. As a result, there is a growing ability to exploit the increased publication of personal information online in order to engage in cyberstalking,[27] perform aggregation and data mining of personal information across otherwise disparate databases,[28] and to build digital dossiers of individuals.[29] Ethicist Herman Tavani has written specifically about the ease with which personal information can be routinely collected, aggregated and analyzed by Web search engines:

> "Because the various news groups contain links to information posted by a person, they can provide search-engine users with considerable insight into that person's interests and activities.[30] So it would seem to follow that not all of the personal information currently included on Web sites accessible to search engines was necessarily either placed there by the persons themselves or explicitly authorized to be placed there by those persons."[31]

An individual might not be aware that her name is among those included in one or more of those databases accessible to search engines, let alone fluent in how the search engines themselves work and their ability to retrieve personal information from a variety of online sources.

---

22) Sharma, D. (21 October 2004). *Is your boss Googling you?* Retrieved 6 January 2007, from http://news.com.com/Is+your+boss+Googling+you/2100-1038_3-5421210.html

23) Weiss, P. (19 March 2006). *What a tangled Web we weave: Being googled can jeopardize your job search*. New York Daily News. Retrieved 7 January 2007.

24) Mills, E. (3 August 2005). *Op.cit*.

25) Ironically, Google punished CNET for publishing the personal information about Schmidt – found via its own search engine – with a one-year boycott against answering any inquirers from the news service. Amid public criticism, Google ended the boycott two months later.

26) Swidey, N. (2 February 2003). *op. cit.*; Ramasastry, A. (12 May 2005). *Can we stop zabasearch – and similar personal information search engines? When data democratization verges on privacy invasion*. Retrieved 12 June 2006, from http://writ.news.findlaw.com/ramasastry/20050512.html

27) Tavani, H. & Grodzinsky, F. (2002). *Cyberstalking, personal privacy, and moral responsibility*. Ethics and Information Technology, 4(2), 123-132.

28) Garfinkel, S. (2000). *Database nation: The death of privacy in the 21st century* (1st ed.). Sebastopol, CA: O'Reilly.

29) Solove, D. (2004). *The digital person: Technology and privacy in the information age* (Ex machina). New York: New York University Press.

30) A search for my name uncovers (admittedly forgotten) posts to Usenet discussion forums from the early 1990s on topics ranging over abortion rights, Catholicism, feminism, marketing, and Lotus 1-2-3 spreadsheet software.

31) Tavani, H. T. (2005). *Search engines, personal information and the problem of privacy in public*. International Review of Information Ethics, 3, 39-45.

John Battelle best summarizes the anxiety that emerges from the perfect reach of search engines:

> "What do we do when information that we know, by law, should be public, becomes, well…really public?...What happens when every single thing that's ever been publicly known about you – from a mention in your second-grade newsletter (now online, of course) to the vengeful ravings of a spurned lover – trails your name forever?"[32]

2.2. Privacy and the Perfect Recall

Accompanying these privacy concerns with the increasingly perfect reach of the perfect search engine are growing concerns about the ability of search engines to collect and aggregate a wide array of personal and intellectual information about their users in pursuit of the perfect recall. The most obvious concern involves the tracking of individual search queries and the ability to associate them with particular IP addresses. This practice of collecting and retaining search query data in support of attaining "perfect recall" gained widespread public attention in January 2006 when it was revealed that, as part of the government's effort to uphold an online pornography law, the U.S. Department of Justice had asked a federal judge to compel the Web search engine Google to turn over records of millions of its users' search queries[33]. Google resisted, but three of its competitors, America Online (AOL), Microsoft and Yahoo!, complied with similar government subpoenas of their search records[34]. Later that year, AOL released over 20 million search queries from 658,000 of its users to the public in an attempt to support academic research on search engine query analysis[35]. Despite AOL's attempts to make the data anonymous, individual users remained identifiable based solely on their search histories, which included search terms matching users' names, social security numbers, addresses, phone numbers and other personally identifiable information[36].

These cases brought search query retention practices into a more public light, creating anxiety among many searchers about the presence of such systematic monitoring of their online information-seeking activities,[37] and leading news organizations to investigate and report on the information search engines routinely collect from their users.[38] In turn, various advocacy groups have criticized the extent to which Web search engines are able to track and collect search queries, often with little knowledge on the part of the users themselves,[39] while both European and U.S. government regulators have begun to investigate search engine query retention practices and policies.[40]

Along with the growing public attention focused on the practice of archiving users' web search queries in server logs, new concern is emerging regarding search engine providers' ability to monitor

32) Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture.* New York: Portfolio, p. 193.

33) Hafner, K., & Richtel, M. *Google resists U.S. subpoena of search data*. The New York Times, pp. A1, C4; Mintz, H. (16 January 2006). *Feds after Google data: Records sought in U.S. quest to revive porn law*. San Jose Mercury News. Retrieved 19 January 2006, from http://www.siliconvalley.com/mld/siliconvalley/13657386.htm.

34) Hafner, K., & Richtel, M. (20 January 2006). *Google resists u.s. Subpoena of search data*. The New York Times, pp. A1, C4.

35) Hansell, S. (8 August 2006). *AOL removes search data on vast group of Web users*. The New York Times, p. C4.

36) McCullagh, D. (7 August 2006). *Aol's disturbing glimpse into users' lives.* Retrieved 3 December 2006, from http://news.com.com/AOLs+disturbing+glimpse+into+users+lives/2100-1030_3-6103098.html?tag=st.num

37) Barbaro, M., & Zeller Jr, T. (9 August 2006). *A face is exposed for AOL searcher no. 4417749*. The New York Times, p. A1; Hansell, S. (8 August 2006). *op.cit.*; McCullagh, D. (7 August 2006). *Aol's disturbing glimpse into users' lives*. Retrieved 3 December 2006, from http://news.com.com/AOLs+disturbing+glimpse+into+users+lives/2100-1030_3-6103098.html?tag=st.num

38) Glasner, J. (5 April 2005). *What search sites know about you*. Retrieved 2 August 2006, from http://www.wired.com/news/privacy/0,1848,67062,00.html; Ackerman, E. (19 August 2006) *What do Google, Yahoo, AOL and Microsoft's MSN know about you?* San Jose Mercury News.

39) See, for example, Electronic Frontier Foundation. (2007). *Privacy and search engines*. Retrieved 28 July 2007, from http://www.eff.org/Privacy/search/; Privacy International. (9 June 2007). *A race to the bottom: Privacy ranking of internet service companies*. Retrieved 10 July 2007 from http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-553961

40) Associated Press. (25 May 2007).*EU data privacy officers launch investigation into Google's internet search engine*. Retrieved 28 July 2007, from http://www.iht.com/articles/ap/2007/05/25/business/EU-FIN-EU-Google-Privacy-Probe.php; Lohr, S. (29 May 2007). *Google Deal said to bring U.S. Scrutiny*. Retrieved 27 July 2007, from http://www.nytimes.com/2007/05/29/technology/29antitrust.html

and aggregate activity across their array of products and services.[41] The encouragement of the creation of user accounts, the cross-promotion of services to keep users from leaving the search engine's website and the use of persistent Web cookies across services, arms search engines with the ability to collect and aggregate a wide array of personal and intellectual information about their users, not just extending beyond the keywords they search for, but also including the news they read, the interests they have, the blogs they follow, the books they enjoy and other intimate details about a person's identity, political interests, health status, sex life, religion, financial status and buying preferences.[42] The result is what John Battelle calls a "database of intentions":

> "This information represents, in aggregate form, a place holder for the intentions of humankind - a massive database of desires, needs, wants, and likes that can be discovered, subpoenaed, archived, tracked, and exploited to all sorts of ends. Such a beast has never before existed in the history of culture, but is almost guaranteed to grow exponentially from this day forward. This artifact can tell us extraordinary things about who we are and what we want as a culture."[43]

While many of our day-to-day habits – such as using credit cards, ATMs, cell phones or automated toll collection systems – leave countless "virtual footprints" of our activities, the drive for the perfect recall necessary to build the perfect search engine results in the construction of a robust infrastructure of dataveillance with the power to monitor and record users' search histories, e-mails, blog posts and general browsing habits, providing "an excellent source of insight into what someone is *thinking*, not just what that person is doing".[44]

3. Making the Perfect Search Audiovisual

With the proliferation of digital images and video content on the Internet and World Wide Web, search engine providers have long sought to include audiovisual content in their searchable indexes. In 1995, AltaVista was the first search engine to offer the ability to search for image, audio and video content, a feature soon offered by the other major search engines of the time, including Lycos, HotBot and Infoseek. These early audiovisual search services incorporated access to multimedia files primarily on the basis of file name extensions (.gif, .avi, .au) and textual references extracted from contextual materials, such as the file name, metadata, hyperlink text or surrounding words.[45] Today, while some specialty search engines survive (Picsearch, for example, offers the ability to search over two billion images), audiovisual search services are dominated by the large general-purpose search engines. Google, Yahoo, Microsoft and Ask each offer the ability to search images and video files, using much the same retrieval methods developed by AltaVista over a decade earlier.

While still heavily dependent on metadata to identify the content of images for retrieval, innovations have emerged in audiovisual searching, including both social and technological enhancements in audiovisual search services. The rise of photo-sharing websites has resulted in an explosion in the amount of personal images available online, dominated by Flickr. While providing a robust and easy-to-use platform for uploading and sharing photos online, Flickr's key innovation was in providing the ability for users to "tag" photos with their own descriptive labels, allowing searchers to find images relating to a certain topic such as place name or subject matter. Upon acquisition by the search engine Yahoo, Flickr's database of over 100 million photos was soon integrated into Yahoo's own image search results, greatly expanding the reach of their audiovisual search offerings.[46]

41) Glasner, J. (5 April 2005). *op. cit.*; Ackerman, E. (19 August 2006). *op.cit.*; Waters, D. (31 May 2007). *What search engines know about us*. Retrieved 2 April 2008 from http://news.bbc.co.uk/2/hi/technology/6700997.stm
42) Zimmer, M. (2008) *The gaze of the perfect search engine*. *Op.cit*.
43) Battelle, J. (13 November 2003). *The database of intentions*. Searchblog. Retrieved 16 May 2006, from http://battellemedia.com/archives/000063.php
44) Hinman, L. (2005). *Esse est indicato in Google: Ethical and political issues in search engines*. International Review of Information Ethics, *3*, 19-25, p. 23.
45) Schwartz, C. (1998). *Web search engines*. Journal of the American Society for Information Science, 49(11), 973-982.
46) Yahoo! (26 June 2007). *Flickr-izing image search*. Yahoo! Search Blog. Retrieved 10 September 2007, from http://www.ysearchblog.com/archives/000466.html

Relying on the photo tags created by the image's creator within Flickr helped Yahoo increase the accuracy of its image retrieval search engine: rather than using surrounding text or merely the name of the image in order to determine its relevancy to a search query, Yahoo can now take advantage of the multiple tags purposely placed by a human being to describe the image. Google soon developed its own way of improving image search results: the Google Image Labeler.[47] Users are randomly paired together and presented with an image from Google's database. Each user provides labels to describe the image, without seeing the other's submissions. If a match is found within 120 seconds, points are earned. Though the feature is enjoyable for the users, it is also a clever way for Google to ensure that its images are matched with the most appropriate keyword.

Along with the growing use of human and socially-generated tags to aid the identification and retrieval of images, advances in automated content-based image retrieval technology have greatly enhanced audiovisual searching.[48] Rather than relying on human-generated metadata, content-based systems attempt to recognize and catalogue different shapes, colors and textures within images, automatically distinguishing between an image of a brown dog and a red apple, for example. The ability to automatically identify and distinguish between *faces* within image retrieval systems has recently received significant attention, both for law enforcement and commercial ends.[49] Increasingly, these facial recognition systems are being deployed in the consumer realm and integrated into Web search engines. For example, after acquiring Neven Vision, a company that develops technology designed to detect and recognize objects and persons in images,[50] Google enhanced their Image Search service to restrict the results to return only images containing faces. For example, a basic search for "Zimmer" in Image Search provides varied results, ranging from the exotic car of that name, random pictures of rooms ("Zimmer" is the German word for "room") and some photos of people. But selecting the content-type "faces" in the advanced image search settings delivers only results containing people's faces.

Recently, various audiovisual search services have been released focusing solely on facial recognition within images. Launched in early 2006, the photo search service Riya provides face recognition and text recognition technology to identify faces within a user's digital photo collection and tag them accordingly.[51] Once the user has manually trained the system by tagging a person's face in numerous photos, the software takes over, and can automatically scan, identify and tag that person in any future photo uploaded onto the user's library.[52] Riya had a million photos uploaded in the first two days after launching, and seven million photos uploaded in the first seven weeks.[53] Building on this success, Riya announced plans to expand its face recognition technologies beyond people's personal photo collection to include any image on the Web. Through its planned "visual search engine," users can submit an image to Riya (a particular rug design, a purse seen in a photograph, or even a person's face), and it will return image results that are similar from across the Web.[54]

The Swedish-based Polar Rose plans similar innovations in using facial recognition for image retrieval.[55] Rather than relying on users to upload their personal image libraries to be analyzed (like Riya), Polar Rose plans to release a Web browser plugin that will identify faces in images on nearly any website, allowing users to tag faces they recognize as they surf the Web.[56] This data is then sent to

47) Lenssen, P. (1 September 2006). *Google image labeler*. Retrieved 11 July 2007, from http://blogoscoped.com/archive/2006-09-01-n44.html

48) See, for example, Lew, M., Sebe, N., Djeraba, C., & Jain, R. (2006). *Content-based multimedia information retrieval: State of the art and challenges*. ACM Transactions on Multimedia Computing, Communications, and Applications, 2(1), 1-19; Vasconcelos, N. (2007). *From pixels to semantic spaces: Advances in content-based image retrieval*. Computer, 40(7), 20-26.

49) Zhao, W., Chellappa, R., Phillips, P., & Rosenfeld, A. (2003). *Face recognition: A literature survey*. ACM Computing Surveys, 35(4), 399-458.

50) Google. (15 August 2006). *A better way to organize photos?* Retrieved 10 July 2007 from http://googleblog.blogspot.com/2006/08/better-way-to-organize-photos.html

51) See http://www.riya.com/

52) Ribeiro, J. (19 January 2006). *Riya photo search site readies for launch*. Retrieved 11 July 2007 from http://www.pcworld.com/article/id,124407-c,searchengines/article.html

53) Arrington, M. (15 June 2006). *Riya 2.0 on the way; major strategy shift*. Retrieved 11 July 2007, from http://www.techcrunch.com/2006/06/15/riya-20-on-the-way-major-strategy-shift/

54) Arrington, M. (15 June 2006). *Op.cit.*

55) See http://www.polarrose.com/. At the time of writing, Polar Rose is still a private beta test.

56) Simonite, T. (19 December 2006). *Face-hunting software will scour Web for targets*. Retrieved 11 July 2007, from http://technology.newscientist.com/article/dn10828-facehunting-software-will-scour-web-for-targets.html

Polar Rose's central server allowing anyone looking at an image containing that particular face to tell who it is, and confirm or correct its tag. Thus, if they see a photo that is clearly Bill Gates but tagged as Steve Jobs, they can make Polar Rose "smarter" by fixing the mistake. Users will also be able to search the web for more photos containing that face, based on the unique biometric profile created for that face by Polar Rose.

In summary, audiovisual search services have witnessed numerous technological advances, including the merging of user-generated images into searchable indexes, the widespread tagging of image files and the deployment of user-based face recognition systems. Through these innovations, Web search providers have been able to extend their *reach* to include the increasing amount of user-generated images available on the Internet. In fact, some search engines focus solely on providing access to audiovisual content, such as the European Quaero multimedia search research project, which includes among its goals the facilitating of access to cultural heritage audiovisual archives and digital libraries.[57] At the same time, as these new audiovisual search tools provide users with the opportunity to search for faces, sounds, videos and places, search providers can gain new insights into users' wants and desires, enhancing their ability to *recall* users' intentions. For example, knowing if a user typically searches for videos of celebrities rather than images of European landmarks can help a search engine provide the most relevant results for the keyword search "Paris Hilton." In total, the perfect search engine is increasingly becoming audiovisual.

4. Privacy and the Perfect Audiovisual Search Engine

Part Two of this article detailed the particular privacy threats that emerge from the increased reach and recall attained in the drive for the perfect search engine. As the perfect search engine increasingly becomes audiovisual, new, and perhaps more potent, privacy threats emerge. This section will detail how the attempts to add audiovisual components to the perfect search engine detailed above intensify the privacy concerns outlined above with regard to the perfect reach and perfect recall of the perfect audiovisual search engine.

4.1. Privacy and the Perfect Reach of Audiovisual Search

As described above, the perfect reach of search engines has reduced the "privacy via obscurity" of individuals by making visible websites, documents and other online files that were previously difficult to locate or access. The rise of advanced audiovisual components accelerates this disintegration of privacy via obscurity as users increasingly lose control over the accessibility and identifiably of their likenesses online. Image sharing websites like Flickr have simplified the ability to upload, tag and share images – both of oneself and of others. Combined with the ubiquity of cellphone cameras, almost any moment of one's life could be captured, tagged and placed online for all to see. With the incorporation of Flickr images into general Yahoo! image search results, users no longer need to be members of the Flickr community (or even aware that it exists) to access the images stored on the service. As a result, an individual might have their photo taken, posted online, labeled with their name and discoverable within the Yahoo search engine, without ever knowing it.[58] While individuals are frequently subject to having their photo taken (especially in public settings), this new ability to have these images tagged and searchable by millions threatens existing privacy norms regarding the availability of one's likeness to mass audiences.

The prevalence of detailed metadata exacerbates this concern. Digital photos typically make use of the exchangeable image file format (Exif), which captures various metadata elements within each image, including date and time information, camera settings, a thumbnail for previewing the picture on the camera's display screen or in software, and possibly copyright information. Many images are also geocoded with tags for location information, either manually by the user or automatically through a built-in GPS receiver on the camera itself. Such detailed metadata elements are viewable (and often

57) See http://www.quaero.fr/

58) For example, a search for "Michael Zimmer" in the Yahoo Image Search service reveals a photo of me delivering a talk, tagged with my name and posted to Flickr, all without my knowledge.

searchable) within online photo-sharing websites, allowing users to know precisely when and where particular photos were taken. As a result, not only are photos of individuals increasingly available online without their knowledge, but a wealth of detailed information about precisely where and when that person was when the photo was taken is also freely available, further lessening that person's presumed "privacy via obscurity."

The rise of robust face recognition tools to identify and search for images online reveals another instance where the powerful reach of the perfect audiovisual search engine threatens user privacy.[59] By no longer relying on users to recognize a person's face and tag it in a Flickr photo, faces are now easily discoverable based on their unique biometric features. With services such as Riya or Polar Rose, a user could upload dozens of personal photos to train the software to learn her own face, making the labelling of future images easier. However, Riya makes that facial profile available to other users to search and identify images as well. Thus, a different user could submit a random photo found online, and the Riya software will scan its database for matching faces and automatically label any recognizable faces within the new image. Users could also scan the entire Web for faces they have trained Riya to recognize. The second user previously had little way of identifying the person in the photo, but with face recognition tools such as Riya and Polar Rose, identification of unknown faces is much easier. Further, Polar Rose also makes it easy to search the Web for other images containing that person's face. This can result in finding images on social networking sites, personal homepages, employer websites, and so on. While previously all such images were publicly viewable, it was difficult for a random user to know they all existed, or where to find them. Through these new socially-oriented face recognition tools, users lose the ability to control the context and community in which certain images are viewable.

4.2. Privacy and the Perfect Recall of Audiovisual Search

Along with the threats to privacy engendered by the expanded reach of audiovisual search, similar threats exist when the *recall* of the perfect search engine is augmented by the presence of audiovisual content. Just as users' general search queries are tracked and logged by search engines in order to help predict their wants and needs, audiovisual search terms – names, places, pornography, and so on – will be similarly captured and potentially associated with users' accounts. More than simply being embarrassing, some search queries could shed light on an individual's social or political views, and perhaps even be of questionable legality. Just as law enforcement officials have shown interest in tracking or gathering general search activities,[60] attention might also be paid to audiovisual searches for controversial or illegal items such as child pornography, cartoons of the Islamic prophet Mohammed, Nazi memorabilia, or copyright protected material. In short, the unique type of query made possible by the addition of audiovisual content in web search services threatens to make the "database of intentions" more detailed and potentially more damaging.

5. Conclusion

In striving to make audiovisual content, including moving images, as "universally accessible and useful" as its textual cousins, powerful new tools have been added to Web search offerings, including the inclusion of user-generated images, robust photo tagging and metadata, and the emergence of sophisticated face recognition tools placed in the hands of everyday Web users. These new audiovisual enhancements bring with them significant privacy concerns, as both the reach and recall of the perfect search engine now include people's faces, biometric identifiers, detailed metadata and the history of the kinds of multimedia content they search for. Thus a kind of Faustian bargain emerges: Audiovisual search technologies promise to enhance the perfect search engine by making available a wealth of

59) Granick, J. (5 July 2005). *Face it: Privacy is endangered*. Retrieved 8 April 2008 from http://www.wired.com/politics/law/commentary/circuitcourt/2005/12/69771 ; Inman, M. (5 January 2007). *Face recognition for online photo searches sparks privacy fears*. Retrieved 5 April 2008 from http://news.nationalgeographic.com/news/2007/01/070105-photo-search.html

60) McCullagh, D. (20 December 2006). *Police blotter: Google searches nab wireless hacker*. Retrieved 13 June 2007 from http://www.news.com/2100-1030_3-6144962.html

multimedia content – much of it user-generated – while also tapping into the value of metadata and the Web 2.0 ideology of having users identify and label images across the Web to help future retrieval. But new privacy concerns also emerge, including the continued loss of "privacy via obscurity," the fact that users lose the ability to control whether images of themselves are uploaded, whether their faces are labeled, and the fact that audiovisual search activities can be added to the ever-growing "database of intentions" collected by search engine providers.

What options exist for renegotiating our Faustian bargain with these emerging audiovisual search tools? First, we must expand our conceptualization of privacy as it relates to content available online, including images uploaded by individuals or taken of them in public places. It becomes vital to embrace a more contextual notion of privacy,[61] recognizing that just because a person was in a public venue, it does not mean that an image should be taken, uploaded to the Web, identified, tagged and perhaps have a biometric scan of the person's face created and shared with all. Some images are meant for a particular context, and performing these additional tasks – even with the positive intent of improving the perfect search engine – might still violate one's privacy.

Second, we must explore new legal and regulatory frameworks to address the increased flow of personal images online and the growth of sophisticated technologies designed to identify, tag and make them searchable on the Web. Some jurisdictions have taken steps towards an outright ban on the publishing of people's likeness online without their explicit consent. Others require the blurring of any identifiable face in online images where individuals have not given their consent. One could also foresee a "notice and takedown" process similar to that used for alleged copyright violations under the US Digital Millennium Copyright Act. Here, users who find their likeness online can issue a notice to the uploader of the image, or the website hosting it, to remove the item in question.

Finally, various technological options exist to help mitigate the privacy concerns about emerging audiovisual search tools. Even in the absence of legal or regulatory requirements, the providers of audiovisual search could provide technological tools for individuals to request the removal of their image when found online, or create a robust reporting system – similar to that in the credit industry – to inform individuals if an image of their face is tagged or uploaded. With regard to the fact that users' audiovisual searches might be tracked and included in a "database of intentions," search providers should give users access to any search activity collected, allowing them to remove any data they do not want recorded, and prevent future recording of their search activities.

In sum, the rise of powerful audiovisual search tools has the potential to greatly enhance the drive for the prefect search engine. But only if the multiple privacy concerns can be addressed – through broadening our conception of privacy, creating policy protections and engaging in privacy-protecting design – will we avoid following the path of Faust and make a bargain with the devil in exchange for access to universal knowledge.

61) Nissenbaum, H. (2004). *Privacy as contextual integrity*. Washington Law Review, 79(1), 119-157; Barth, A., Datta, A., Mitchell, J. C., & Nissenbaum, H. (2006). *Privacy and contextual integrity: Framework and applications*. Paper presented at the IEEE Symposium on Security and Privacy; Zimmer, M.(2008). *Privacy on planet Google. Op.cit.*

Freedom of Expression Implications for the Governance of Search

Joris van Hoboken
Institute for Information Law, University of Amsterdam

Introduction

Search tools and services have become central in our information environment. The most straightforward explanation for this development is the abundance of online information, in particular on the World Wide Web, while innovations in search and information technology have made this development possible. As a result of this abundance, the question is now not so much whether certain information is published, but whether an end-user will be able to find it in the midst of all the other material out there. While the Internet and text-based media stand out because of the overwhelming amount of sources and material they have to offer to their users, audiovisual content is following closely on their heels. According to predictions of Internet traffic statistics by Cisco, "Video on demand (VoD), IP television (IPTV), and Internet TV will account for nearly 90 percent of all consumer IP traffic in 2012."[1] Therefore, the same gate-keeping power is shifting towards new intermediaries, and questions about the governance of the tools and services that index the available content and help users select the content of their choice are gaining importance in the policy debate about the information society.

This contribution will provide insight into the debate about the regulation of search from the perspective of freedom of expression and information. It will do so by looking at some of the governance issues relating to search engines, from the right to freedom of expression to the interests of the primary stakeholders in the free flow of information facilitated by the search medium. Where appropriate, particular attention will be paid to aspects that are specific to the regulatory debate about audiovisual search.

In the first section search engines are characterized in terms of their functionality and a few examples are given of their growing importance as an element in the Internet content value chain (1.1). Following this, the focus is placed on the primary stakeholders with regard to the governance of search engines (1.2) and the interests of the primary stakeholders with respect to the functioning of search engines (1.3).

The second section will address to what extent these interests are protected under the right to freedom of expression and information, as enshrined in Article 10 of the Convention for the Protection of Human Rights and Fundamental Freedoms. After a short examination of this general question (2.1),

1) Cisco Visual Networking Index - Forecast and Methodology, 2007–2012, white paper, 16 June 2008. Available at http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf (Last viewed 1 July 2008).

the analysis will proceed by comparing search engines to Internet filters (2.2) and the freedom of expression issues they have raised. In the context of this comparison, a recent Recommendation of the Council of Europe's Committee of Ministers on measures to promote respect for freedom of expression and information with regard to Internet filters will be discussed (2.3).[2] As will be shown, the Council of Europe's Recommendation on Internet filtering and the underlying expert Report[3] provide an interesting starting point for the analysis of the implications of freedom of expression for the governance of the inclusion and exclusion of information providers by search engines, in other words the issues surrounding the collection of information providers in the search engine's index that can be accessed through a particular search service (2.4).[4]

Section three offers a discussion on the question of the liability of search engines for third party content, an important issue for the governance of the inclusion and exclusion of information providers in a search platform. The section will give a short overview of the present legal regime for search engine liability at the EU level and characterize the freedom of expression perspective in this context. Section four concludes the contribution.

1. Search Engines

1.1. What Does a Search Engine Do?

Generally speaking, search engines help the end-users of a certain medium to find content, including audiovisual content, which can be enjoyed through that medium. Internet search engines like Google and Yahoo! provide such services for end-users of the Internet. The primary stakeholders with regard to a search service are the service provider itself, the end-users and the information providers. Search engines are in the business of organizing information. Commercial search engine providers usually make profits through the sales of targeted advertising. The advertisements can be linked to specific search actions of end-users. Most search engines are commercial and follow such a business model. For end-users, search engines are a tool that helps them select information from a list available through a particular medium. On the other side of the search service, we find the information providers that have an information output that could qualify for inclusion in a search engine. Thus, for information providers, search engines are a means to reach an audience and the search platform functions like a forum in which they compete for prominence and the end-users' attention. The importance of being included in a search engine's index is captured by the following phrase: "to exist is to be indexed by a search engine."[5] Although most information providers will aim to be included and ranked highly, in some cases information providers will in fact like to be excluded from a search engine.[6]

The functional steps that are needed to set up a search engine can be divided into two. First, the provider has to compose a list or index of information from the full set of information that can be made available to the end-user. Second, it can use this list to offer an interactive selection service, allowing the end-user to search the index. It follows that a Web search engine allows end-users to search their index, not the entire World Wide Web.

For example, in the case of the search engine for the Web as provided by Google the available information is all the information on the Web that its automatic spiders or crawlers can access. The

2) Recommendation CM/Rec(2008)6 of the Committee of Ministers to member states on measures to promote the respect for freedom of expression and information with regard to Internet filters. Available at https://wcd.coe.int/ViewDoc.jsp?Ref=CM/Rec(2008)6 (Last viewed 1 July 2008).

3) Report of the Group of Specialists on human rights in the information society (MC-S-IS) on the use and impact of technical filtering measures for various types of content in the online environment, as it appears in document CM(2008)37 add, Available at https://wcd.coe.int/ViewDoc.jsp?Ref=CM(2008)37&Ver=add (Last viewed 1 July 2008).

4) Schulz, W., T. Held, et al. (2005). *Suchmaschinen als Gatekeeper in der öffentliche Kommunikation: Rechtliche Anforderungen an Zugangsoffenheit und Transparanz bei Suchmaschinen im WWW.*

5) Introna, L. D. and H. Nissenbaum (2000). *Shaping the Web: Why the Politics of Search Engines Matters.* The Information Society 16(3): 169-186.

6) See infra section 2.4.

search engine sends these crawlers to look for information that is available online and systematically makes an index based on their findings. New technologies make it possible to detect audiovisual content and more and more search engines have developed ways to index audiovisual content. A lot of content on the Web is not included in this index. Many information providers use the *robots.txt* instruction file and other functionalities to instruct search engines' crawlers not to index parts of their websites and major search engines normally respect that. Other sources might not be included for other reasons, for instance because the crawlers do not have the technical capabilities to extract the information from a particular website. Yet another reason why Google might not include a website in its index is that it would be unlawful to show the reference, for instance because it has received a notice under the Digital Millennium Copyright Act (DMCA). As a result, Google's index contains a selection of the information available on the Web, a selection that is determined by information providers, third parties, Google itself, the technical properties of the Web and the law.

After the index is made, the search engine lets end-users interact with it by typing in keywords. The search results are organized and ranked to help the end-user find satisfactory results, and targeted advertisements are placed next to the results to make a profit. The methods that are used by major search engines like Google to organize and rank content are a complex combination of statistical methods. One of these methods is the frequency analysis of keywords on the websites. Many search engines also perform an analysis of the popularity and importance of websites, for instance by making a hyperlink analysis. Google's original PageRank algorithm is one of the most widely known and successful algorithms that use the hyperlinks on the Web to make an estimate of the relevance of a specific website. The original PageRank algorithm measures the relevance of a website, e.g., website X, by counting the links from other websites that link to X. A link carries more weight if it comes from a website that is more linked to itself. Therefore, under PageRank, a website is most relevant if it has many links coming from very relevant pages. Another factor that is used by search engines to determine the ranking of a website is the past viewing and searching behaviour of end-user(s). This search and viewing history can be used to personalize search results and make recommendations to end-users. The details of the more complex ranking algorithms are usually kept secret.

1.2. Examples of Regulatory Attention for Search Engines in Europe

Many have noted the importance of search engines and their impact on the information flows in the information environment. Internet search market leader Google receives particular attention, but search engines in general have also started to gain in importance in the eyes of European regulators and policy makers.

In the context of Internet content regulation, Ofcom (the UK regulator with responsibilities covering television, radio and electronic communications), has started to allocate search engines a separate segment in the Internet value chain.[7] It divides this value chain into two broad categories, namely entities that control the availability of content, such as content producers, content aggregators and Web hosts, on the one hand, and entities that control the accessibility to content, such as Internet service providers, navigation and search engines and consumer devices, on the other. Ofcom takes the position that traditional regulatory intervention or strategies with regard to the availability of content have become less effective, due to the global Internet content market and complications resulting from legal jurisdiction.[8] Whereas content regulation used to address entities in the content value chain that controlled the availability of content, Ofcom stresses the need to focus also on the entities that control the accessibility of content. Ofcom concludes that self-regulation, in particular by Internet Service Providers – which have strong ties with the jurisdictions of end-users – is a crucial element in a policy that aims to manage content risks for children. In the case of search engines, however, such a lack of strong ties with jurisdictions is notable. Ofcom concludes that filtering should be promoted as a solution, but that it is not feasible to legally require ISPs to filter at the network layer, a type of content control that is practiced for instance in China and Saudi-Arabia.[9]

7) See Ofcom - Office of Communications, Ofcom's Response to the Byron Review, 27 March 2008, pp. 50-51. Available at http://www.ofcom.org.uk/research/telecoms/reports/byron/byron_review.pdf (Last viewed 1 July 2008).
8) Ibid, pp. 50-51.
9) Ibid., p. 51.

Another European regulatory debate on the subject of search engines relates to their effects as concerns privacy and the protection of personal data. The user data retention policies of large Internet search engines were investigated by the European Union's Article 29 Data Protection Working Party.[10] Search engines tend to retain detailed data on the use of their services by end-users over a lengthy period of time. The logs of search engines normally include the search terms entered, the ip-address(es) from which a user is connected to the Internet at specific dates and times, a unique cookie identifying the user and the consequent user navigation on the search engine platform after the search results have been presented to the user. The Article 29 Data Protection Working Party recently concluded an investigation into the data processing practices of major search engines with a detailed official Opinion, which opens by stating that "Search engine providers on the World Wide Web fulfil a crucial role in the information society as intermediaries. The Working Party recognises the necessity and utility of search engines and acknowledges their contribution to the development of the information society."[11]

Finally, another reference to search engines can be found in the underlying Report of the Council of Europe in the context of the implications of freedom of expression for Internet content filters.[12] The Report clarifies that: "Most Internet users today rely on search engines to rapidly find the most relevant information for their needs. In this way, search engines play an important role as gatekeepers in determining and shaping which information is received through the Internet. This issue is accentuated when, as is the case today, there is one dominant search engine provider on the market. The way such a search engine lists and prioritizes search results may in practice have an important impact on what information a user receives."[13] In the second section of this contribution, the freedom of expression implications for search engines and Internet content filters will be further addressed.

1.3. The Interests of Information Providers, End-users, Search Engines and Third Parties

Before addressing some of the implications of freedom of expression for the governance of content accessibility through search engines, it is interesting to look at the interests of the primary stakeholders. In the following section, a broad characterization is given of what is at stake for the search engine provider, end-users and information providers with regard to the accessibility of content through a search platform. As will become apparent, these interests are not always in alignment. The question to be addressed later on in this contribution is to what extent they are protected under freedom of expression. Of course, third parties may also have interests in the governance of search engines, such as the protection of their privacy and reputation or intellectual property, however a discussion of these matters goes beyond the scope of this contribution.[14]

To start with the search process itself, one can distinguish different types of issues that arise in the governance of the accessibility of content. As explained above, the composition of the index is decisive for what can be found through a particular search engine and there are many reasons why a particular source could be excluded from the index of a search engine. However, not only the composition of the index, but also the consequent ranking of search results has enormous impact on what is being found in practice. Empirical research shows that end-users usually do not go beyond the first page of search

10) The Article 29 Working Party consists of the national data protection authorities of the EU's Member States and has the task of facilitating harmonized interpretation of the Data Protection Directive (95/48/EC), Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 23 November 1995, pp. 31–50.

11) Article 29 Working Party, Opinion on data protection issues related to search engines, WP 148, 4 April 2008. At several points in the Opinion, the Working Party addresses the importance of the fundamental right to freedom of expression and information in the context of search engines. ("[…] a balance needs to be struck by Community data protection law and the laws of the various Member States between the protection of the right to private life and the protection of personal data on the one hand and the free flow of information and the fundamental right to freedom of expression on the other hand."). Available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf (Last viewed 1 July 2008).

12) Supra, notes 2 and 3 and accompanying text.

13) Supra, note 3.

14) For a comprehensive overview of the interests involved in the governance of the various aspects of search engines and the information flows they facilitate, see James Grimmelmann, *The Structure of Search Engine Law,* Iowa Law Review, Vol. 93, No. 1, 2007. Available at http://ssrn.com/abstract=979568

results and tend to click more on the first couple of results listed: references to a film on the first page determine whether it will be found or not.[15] Therefore there is more needed to exist, than just to be indexed. As a consequence, a whole industry of search engine marketing and optimisation exists that assists information providers in achieving a prominent ranking in search results and managing effective advertising campaigns in search engines. Finally, the search service might be personalized for specific audiences and/or end-users. We could end up in a situation in which all end-users have their own personalized search service based on their own interests and contexts of use.

All primary stakeholders have specific interests with regard to these governance issues. The search provider's interest is to be in control over its own service, including the composition of its index, its functioning and the relevance and integrity of its search results. The interest of many information providers is to feature effectively in a search engine so as to reach an audience, although it is worth mentioning that, for various reasons, some may in fact not want to be included in a search engine. Since the business models of search engines and many online publishers are based on revenues from advertising, many publishers also compete with search engines for online advertising revenue.[16] Finally, the end-user's general interest is to search freely and find the information providers that provide the most value. Thus, he/she has an interest in being able to inform him/herself fully through using the medium. In the next section the right to freedom of expression and information will be discussed in relation to these interests of the primary stakeholders.

2. Freedom of Expression and Information and Search Engines

2.1. The Implications of Article 10 European Convention on Human Rights

According to Article 10 of the European Convention on Human Rights, everyone has the right to freedom of expression, which includes the freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.[17] Article 10 para. 2 provides that these freedoms are subject to possible restrictions which are prescribed by law, necessary in a democratic society and serve one of a listed set of goals, such as the protection of health or morals and the protection of the reputation or rights of others. The protection of the rights of others is understood to include the protection of intellectual property rights.[18]

The European Court of Human Rights (hereinafter referred to as "ECHR") has not yet addressed the question of search engines in any of its rulings, so it is necessary to examine other case law instead in order to address the implications of Article 10 for the governance of search engines. One judgment that contains a number of relevant considerations is the *Autronic* judgement, in which the European Court of Human Rights reaffirmed that Article 10 applies not only to natural persons, but also to profit making corporations. It concluded that:

> "neither Autronic AG's legal status as a limited company nor the fact that its activities were commercial nor the intrinsic nature of freedom of expression can deprive Autronic AG of the protection of Article 10 (art. 10). The Article (art. 10) applies to "everyone", whether natural or legal persons."[19]

15) For this and other empirical findings about search engine use see the special volume of the Journal of Computer-Mediated Communications, Volume 12, Issue 3, April 2007, Special Theme I, The Social, Political, Economic, and Cultural Dimensions of Search Engines, edited by Eszter Hargittai. Available at http://jcmc.indiana.edu/vol12/issue3/. Last viewed 1 July 2008.

16) The European Commission discusses this aspect in its recent Communication on Creative Content Online, See Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Creative Content Online in the Single Market. COM(2007) 836 final, Brussels, 3 January 2008, pp. 34-35.

17) Council of Europe, Convention for the Protection of Human Rights and Fundamental Freedoms, 4 November 1950, Europ.T.S. No. 5.

18) For a discussion see P. Bernt Hugenholtz, *Copyright and Freedom of Expression in Europe,* in: R. Dreyfuss et al. (ed.), Expanding the Boundaries of Intellectual Property. Innovation Policy for the Knowledge Society, New York: Oxford University Press, 2001, pp. 343-363.

19) ECHR, 22 May 1990, A 178, § 47 (*Autronic AG v. Switzerland*).

Thus the commercial and corporate nature of major search engines does not deprive them of possible protection by Article 10. In the same judgment the Court clarified that:

> "Article 10 (art. 10) applies not only to the content of information but also to the means of transmission or reception since any restriction imposed on the means necessarily interferes with the right to receive and impart information."[20]

It is unclear whether search tools in general should be understood to be a means of transmission and reception of information. Article 10 does not explicitly refer to a right to seek or gather information, but these communicative processes are generally not construed as being excluded from its scope. Interestingly, both the United Nations Universal Declaration of Human Rights and the United Nations International Covenant on Civil and Political Rights do contain an explicit reference to the right to seek information and ideas in Article 19.[21]

Van Eijk already noted the difficulties of qualifying search engines in the light of freedom of expression due to their complex role.[22] Because it combines both being an intermediary between information providers and end-users and being a communicator itself, the search engine is a complex medium. The Internet access provider of an end-user is an essential part in any online communication between that end-user and online information providers. If the access provider does not facilitate the communication between the end-user and the information provider, it cannot take place. Search engines, however, are not a necessary part in a linear chain of communications between information providers and end-users. An online information provider can be easily accessed without the use of a search engine, either directly by typing in the URL in a browser, or indirectly by following a link to the blog elsewhere, for instance on another website or through the use of another search or recommendation service.

For these reasons Van Eijk asserts that the freedoms to impart or receive information, as explicitly mentioned in Article 10, are not the main aim of search engines, since the information is already present and directly accessible.[23] A search engine does not and cannot change that. Van Eijk concludes that search engines *facilitate access* to information and do not offer access to information by themselves. He asserts that this activity of *making accessible* should have a similar status under Article 10 as the activity of disclosing or disseminating information and ideas.

This perspective focuses on the ultimate communication between information providers and end-users that is (partly) facilitated by search engines. It should receive significant weight because of the increasing importance of search engines as mediators in our information environment; as pathfinders and as fora to reach audiences. The question, however, is whether Article 10 does not also protect the communications between the information provider and the search engine and the communications between the search engine and the end-user. By zooming in on the communications that are taking place between the primary stakeholders, we may be able to draw a more detailed analysis of the freedom of expression implications for the governance of search engines.

The search engine's crawlers are a means of receiving information, in some ways a modern, automated version of the researcher or journalist delving for sources of information. The *Autronic* judgement of the ECHR clarified that Article 10 also protects the means of transmission and reception. Since crawling is a means of reception, crawling would be protected under Article 10 and interference

---

20) Ibid.

21) The United Nations Universal Declaration of Human Rights, Adopted and proclaimed by General Assembly resolution 217 A (III) of 10 December 1948 (Article 19: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."); The United Nations International Covenant on Civil and Political Rights, adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of 16 December 1966 (Article 19 para. 2: "Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.").

22) N. A. N. M. van Eijk, "Search Engines: Seek and Ye Shall Find? The Position of Search Engines in Law." IRIS *plus* 2006-2, p. 5. Available at http://www.obs.coe.int/oea_publ/iris/iris_plus/iplus2_2006.pdf.en (Last viewed 1 July 2008).

23) Ibid.

by public authorities would need to satisfy the test of Article 10 para. 2. Therefore, a law which would restrict crawling by search engines would need to satisfy this test. On top of that, the importance of search engines for the dissemination of information and ideas and the ability of the public to inform itself could be one of the determining elements in deciding the extent to which crawling and indexing are protected and the question of whether such interferences are in accordance with Article 10 § 2. Thus, the protection against interferences with crawling by search engines would be enhanced because of the importance of search engines in our information society.

Similarly, the publication of search results by a publicly accessible search engine and the querying and reception of search results by the end-user would be protected under Article 10. The publication of search results itself constitutes a communication between the search engine provider and the end-users. From this perspective, the search engine not only *facilitates access to information* of third parties, but also *offers access to information* about third parties. An interesting question in this context is to what extent the communications between a search engine and its users are in fact public in nature. Arguably, the querying by search engine users is not a public but a private communication. To the extent that the search service is made publicly available, the communication of search results to users is of a public nature.[24]

Based on the case law of the ECHR with regard to Article 10 and the interpretation of the "duties and responsibilities" in Article 10 para. 2 there are several things worth mentioning in the context of search engines. First, the potential impact of the medium and the nature of the content that can be found through a search engine could be of importance here. Several times the court has clarified that the particular impact of audiovisual media has to be taken into account when considering the permissibility of interference by public authorities. For instance in its *Jersild* judgment, the Court concluded as follows:

> "In considering the "duties and responsibilities" of a journalist, the potential impact of the medium concerned is an important factor and it is commonly acknowledged that the audiovisual media have often a much more immediate and powerful effect than the print media [...]. The audiovisual media have means of conveying through images meanings which the print media are not able to impart."[25]

Therefore, it is possible that audiovisual search engines would receive less protection against interference than text-based search engines.

Secondly, a search engine might carry specific duties and responsibilities connected to its role in the information environment and the (technical) nature of its service. According to the ECHR the duties and responsibilities, under Article 10 European Convention of Human Rights, also depend on the technical means used for expression and dissemination. The Court has stated that:

> "[...] whoever exercises his freedom of expression undertakes "duties and responsibilities" "the scope of which depends on his situation and the technical means he uses. The potential impact of those means must be taken into account when considering the proportionality of the interference. The safeguard afforded by Article 10 to journalists is subject, because of those very "duties and responsibilities", to the proviso that they provide reliable information in accordance with the ethics of journalism"[26]

To which the Court adds in a more recent judgment:

> "These considerations play a particularly important role nowadays, given the influence wielded by the media in contemporary society: not only do they inform, they can also suggest by the way in which they present the information how it [the information] is to be assessed. In a world in which the individual is confronted with vast quantities of information circulated

24) See Schulz & Held 2005.
25) ECHR 23 September 1994, § 31 (*Jersild v. Denmark*).
26) ECHR 14 June 2007, § 42 (*Hachette Filipacchi Associes v. France*).

> via traditional and electronic media and involving an ever-growing number of players, monitoring compliance with journalistic ethics takes on added importance."[27]

These last considerations do seem to resonate in the context of search engines but it remains an open question whether and how these considerations translate into the context of search engines and whether we will see a certain development towards a body of search engine ethics and self-regulation. Undoubtedly, the impact of search engines on our information environment is considerable. It would, however, be premature to give the ultimate judgment on these issues. The next section will continue the discussion by focusing on one specific aspect of search engine governance from the perspective of freedom of expression, namely the governance of the composition of the index of search engines. Before addressing this question, however, it is useful to consider a similar problem, that is to say the governance of Internet content filters, which has recently been dealt with in a recommendation by the Council of Europe's Committee of Ministers.

2.2. Internet Filters and Search Engines: a Possible Analogy?

There are many types of Internet content filters ("Internet filters") and they are deployed in a variety of circumstances. Of interest here is the type of Internet filter whose purpose it is to limit the accessibility of Internet content. The technicalities of Internet filters will be left aside, but there are filters that are – amongst other things – able to detect audiovisual content (for example, Google is working on filtering copyright protected content on Youtube based on a "fingerprinting" technique).[28] Internet filter software aims to block certain categories of content that would be accessible for end-users with an open, unfiltered Internet connection. Internet content filters can be used and installed on Internet terminals by end-users. They can be deployed by private actors on their networks, for instance by employers and the owners of Internet cafés. They can be voluntarily deployed by Internet Service Providers (hereinafter referred to as "ISPs") on the network level as a service to their customer, giving customers control over filtering. They can be installed by public libraries that want to restrict access to harmful and illegal content on their internet access points.[29] In a growing number of countries in and outside of Europe, ISPs have agreed with public authorities to filter child pornography at the network level voluntarily. In some jurisdictions, public authorities require ISPs by law to use filtering products at the network level.[30]

The function of Internet content filters is to block a part of the available information on the World Wide Web that would otherwise be accessible. As such, Internet filters fall into the broader category of selection intermediaries, a class of intermediaries that select and filter information travelling online from information providers to end-users.[31] Examples of selection intermediaries include Internet content filters, search engines, recommendation services and ISPs that block or filter material on the basis of content. Selection intermediaries fulfil an important function in our information environment. In this environment, which is characterized by abundance, selection intermediaries help end-users to find and select the information they consider relevant or useful and exclude information that they are not willing or allowed to access, for instance because it is harmful or illegal.

Since search engines and Internet filters are similar in the sense that they both make decisions about the collection of information that is accessible through their services, it is interesting to see what can be said about Internet filters from the perspective of freedom of expression. Taking into account the differences between Internet filters and search engines, this will provide a useful starting point for the analysis of the freedom of expression implications for search engines. One difference is that search engines are more complicated, in the sense that they not only make a selection of

27) ECHR 10 December 2007, § 104 (*Stoll v. Switzerland*).
28) For a discussion see: Report of the Group of Specialists on human rights in the information society, and references, supra, note 3.
29) In the United States some of the government funding to public libraries has been made conditional on the installation of such filtering software. See United States Supreme Court, *United States v. American Library Association*, 539 U.S. 194 (2003).
30) For a comprehensive overview and discussion of global Internet filtering see Ronald Deibert, John Palfrey et al. (ed.), *Access Denied: The Practice and Policy of Global Internet Filtering,* Cambridge MA: The MIT Press, 2007.
31) See Chandler, Jennifer A., "A Right to Reach an Audience: An Approach to Intermediary Bias on the Internet". Hofstra Law Review, 2008. Available at http://ssrn.com/abstract=1021344

information providers, i.e., sources on the Internet that can be found through their service, but in addition organize and order this selection of information providers with their ranking algorithms. Another difference is that Internet content filters are normally placed between the end-user and the Internet at the end-user's terminal or at some level of the network and that switching between Internet filters is different from switching between search engines. Although there is one dominant search engine on the Web today in terms of use, namely Google, there are other similar search engines and normally end-users can easily switch between them. Finally, the purpose of Internet content filters is to prevent content from being accessible, while the purpose of search engines is to help users to access content. On a functional level, these purposes are quite opposite.

Apart from the differences between search engines and Internet filters, however, there remains an important similarity between the governance of Internet filters and search engines, in particular when we restrict the focus to the governance of the inclusion and exclusion of information providers from a search engine's index. As noted before, both types of services manage the accessibility of information. For information providers the decision whether or not to be indexed by search engines or included on a blocking list of filters has similar effects on their availability to the public. For end-users, the composition of the index of search engines presents similar issues of access to information, transparency and control as in the case of Internet content filters.

Notably, search engines already use filters within their services. The image search functions of the major search engines all have an optional filter that an end-user can turn on or off. Google, for instance, offers the option of SafeSearch in its search preferences, allowing end-users to opt for a filtering of explicit images from image search results and a filtering of all search results, including text based search.[32] The preferences for filtering are saved through the placement of a cookie on the end-user's browser. Google has also started to offer filtering for corporate clients.[33] As a result of the problem of search engine spam (spamdexing) search engines might also use filters to filter pages that have no other purpose than to manipulate search results and receive an unfairly high ranking. In the context of unlawful search results, a recurring question is whether search engines should use filters to prevent unlawful results from showing up, or be obliged to use such filters by law or regulation.

These findings are in line with the findings of the Council of Europe's Group of Specialists on Human Rights in the Information Society, which prepared a report on Internet filters for the Committee of Ministers.[34] The Report notes that the "filtering effects of search engines can [...] be said to be more indirect than for filtering products", "[...] filtering products aim to protect users against illegal, harmful or otherwise unwanted content, search engines aim to provide the user with rapid access to information that is most likely to fulfil the user's information needs as expressed in the search terms employed." The Report also notes the use of content filters by search engines and asserts that "the findings in the report apply to search engine providers to the extent that they are users of filtering software." The Report states that "it is likely that at least some of the best practice guidelines identified in this report could form the basis of similar guidelines for search engines."[35]

From the above it is clear that although there are some important differences, an enquiry into Internet filters and the implications of freedom of expression for Internet filters is justified. This enquiry will be made along the lines of the recent Recommendation on Internet filters and freedom of expression of the Council of Europe's Committee of Ministers.

2.3. Recommendation on Internet Filters and Freedom of Expression

On 26 March 2008, the Council of Europe's Committee of Ministers adopted a Recommendation (CM/Rec(2008)6) on measures to promote the respect for freedom of expression and information with regard to Internet filters.[36] The Recommendation and the underlying Report acknowledge both the

32) Google SafeSearch filtering preferences, viewed 1 July 2008, available at http://www.google.com/support/bin/static.py?page=searchguides.html&ctx=preferences
33) 'Web Security for Enterprise', Cnet News.com, 8 May 2008. Available at http://www.news.com/8301-10784_3-9938812-7.html?tag=nefd.top (Last viewed 1 July 2008).
34) Report of the Group of Specialists on human rights in the information society, and references, supra, note 3.
35) Ibid.
36) Supra, note 2.

legitimate function of Internet filters and the ways in which Internet filters can impact on freedom of expression and information. It explicitly addresses some of the requirements of Article 10 of the European Convention on Human Rights in this context, some of which relate to the interests of information providers and end-users. The Recommendation calls upon the Member States of the European Convention to take measures with regard to Internet filters in line with a set of guidelines promoting user notification, user awareness and user control of Internet filters, and the accountability of the private and public parties involved. The guidelines are categorized into three sections. Section I gives guidelines about "using and controlling Internet filters in order to fully exercise and enjoy the right to freedom of expression and information". Section II deals with "appropriate filtering for children and young people" and section III is about the "use and application of Internet filters by the public and private sector".

The Recommendation differentiates between mandatory filtering and the use of Internet filters by public entities, such as public libraries and schools, on the one hand, and their use by private entities, such as enterprises in the context of internet access in the workplace, on the other. The Recommendation does not limit itself to the implications of freedom of expression on Internet filters in the public sector or legally prescribed filtering, but it also contains specific guidelines on the use of filters in the private sector. With regard to the use and application of Internet filters by public entities or mandatory filtering, the Recommendation qualifies the use of Internet filters as an interference with the right to freedom of expression and makes the test of Article 10 § 2 more explicit. With regard to the use of Internet filters by private entities, the Recommendation and the underlying Report refer to positive obligations upon Member States as guarantors of pluralism and in maintaining and enhancing legal and practical measures to prevent private censorship.

In section III of the guidelines, the Recommendation specifically addresses the demands of Article 10 ECHR with regard to the legal requirements relating to Internet filters. It calls upon the Member States to refrain from filtering Internet content in electronic communications networks operated by public actors for reasons other than those laid down in Article 10 § 2. Similarly, it states that Member States are only allowed to introduce nationwide general blocking or filtering measures if the conditions of the Convention, in particular Article 10 § 2, are fulfilled. According to the guidelines in the Recommendation, this means that the filtering has to concern "specific and clearly identifiable content", "a competent national authority should have taken a decision on its illegality" and "there should be an opportunity to have this decision reviewed by an independent and impartial tribunal or regulatory body, in accordance with the requirements of Article 6 of the European Convention on Human Rights".[37] Furthermore, the guidelines stipulate that Member States have to ensure that there is an evaluation of the proportionality of filters before and during their implementation in terms of their possible effects on the unreasonable blocking of content.

Several of the guidelines – in sections I and III – explicitly relate to the governance of the inclusion and exclusion of information providers by Internet filters and the rights of end-users and information providers in this respect. First of all, the guidelines provide that end-users, where appropriate, must be able to control the level of filtering. The guidelines further stipulate that end-users should have the possibility of challenging the blocking or filtering of content and seeking clarifications and remedies. With respect to the end-user's ability, where appropriate, to activate and deactivate filters and to be assisted in varying the level of filtering in operation, the guidelines call upon the Member States to ensure, in cooperation with the private sector and civil society, the existence of a number of more detailed guarantees. It is stated that end-users should receive guidance regarding the manual overriding of an activated filter, including, more specifically, whom to contact when it appears that content has been unreasonably blocked, as well as the reasons that may allow a filter to be overridden for a specific type of content or Uniform Resource Locator (URL). Furthermore, the Recommendation states that content that is filtered by mistake or because of an error has to be accessible without undue

37) Article 6 European Convention of Human Rights, first paragraph: "In the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law. Judgement shall be pronounced publicly by the press and public may be excluded from all or part of the trial in the interest of morals, public order or national security in a democratic society, where the interests of juveniles or the protection of the private life of the parties so require, or the extent strictly necessary in the opinion of the court in special circumstances where publicity would prejudice the interests of justice."

difficulty and within a reasonable time. With regard to the use and application of Internet filters by the public sector, Member States have to avoid the universal and general blocking of offensive or harmful content for users who are not part of the group for which a filter has been activated to protect, and of illegal content for users who justifiably demonstrate a legitimate interest or need to access such content in exceptional circumstances, particularly for research purposes. Member States should also provide for effective and readily accessible means of recourse and remedy, including suspension of filters in cases where users and/or authors of content claim that content has been blocked unreasonably.

2.4. Relevance for the Governance of Search

The Council of Europe Recommendation on Internet filtering and the underlying expert Report provide an interesting starting point for the analysis of the implications of freedom of expression for the governance of the inclusion and exclusion of information providers by search engines in the form of the issues surrounding the collection of information providers that can be accessed through a particular search service.

Following the Recommendation, there would be a difference between the freedom of expression implications with regard to legal obligations on search engines to limit access to certain information providers and the provision of search engines by the public sector, on the one hand, and the voluntary limitations on such access by search engines in the private sector, on the other hand. A legal obligation on search engines to filter out certain information providers from the search index would have to respect the test of Article 10 para. 2. Following the guidelines, this would mean that the removal of information providers from the search engine's index has to concern specific and clearly identifiable content. Imposing general preventive filtering obligations on search engines with regard to illegal or infringing content could be problematic from this point of view.[38] Another interesting point is the requirement that, in cases of mandatory filtering, a competent national authority should decide on the illegality of the information and an opportunity to have this decision reviewed by an independent and impartial tribunal or regulatory body should be offered.

The guidelines on Internet filters aim to guarantee a certain degree of user control over and transparency of the application of Internet filters and it is interesting to see how this plays out in the context of search engines. A first problem that arises here is that a search engine can not be turned on or off like an Internet filter. In the context of search engines, the Internet end-user has other options of redress. The end-user can, for instance, decide to use another search engine, something which is only different if the choice between search engines is restricted, which is normally not the case. The end-user can also access the information directly, which requires that he or she is aware of the location of the information on the Internet. In many cases, the use of a search engine implies that the user is not aware of the information's location. The additional filters of search results offered by search engines, such as the image filters discussed above, already give users control over filtering in the advanced settings. These possibilities could further increase due to technological developments.

With regard to the effective and readily accessible means of recourse and remedy in the case of an information provider not included in a search engine's index, there are several points worth mentioning. First, search engines partly rely on the suggestions of third parties, especially information providers, to include sources in their index. All major Web search engines have online forms and tools to submit websites for crawling. Notably, the control that information providers have over being indexed as a result of exclusion protocols like *robots.txt*, implies less control for end-users over the composition of the index. By respecting instructions like *robots.txt*, the major search engines favour the interests of information providers in having control over the accessibility of their content through a search service, over the end-users' interests in having such control. There are several reasons why general search engines have come to respect these protocols. A systematic discussion of these reasons goes beyond the scope of this contribution. One of the main reasons is that search engines on the Web

38) This seems to fit with Article 15 of the Directive on Electronic Commerce (2000/31/EC). However, one should note that Article 15 of the Directive on electronic commerce does not apply to search engines, as can be seen in Article 21 of this Directive. Infra, section 3.

do not restrict themselves to meta-data, but store large parts or even the complete contents of websites on their servers. By respecting the exclusion protocols, search engines evade some – but surely not all – of the potential conflicts with online publishers.[39] In its Opinion on search engines, the Article 29 Working Party recommends that search engines respect these protocols because they could be used to prevent personal data being indexed.[40]

The situation might be different when a search engine has purposely removed a specific website from its index. One of the reasons for such a removal could be that a website is using illegitimate techniques to influence its ranking. Another reason could be that the search engine provider considers the removal of certain references from its index as its particular responsibility.[41] Yet another reason could be that s/he is required by law or ordered by a judge to do so, because the search results are unlawful, for instance because they contain or link to illegal information.[42] This leads to the interesting question of the circumstances in which search engine providers find themselves under such a legal obligation to remove references from their index. This legal question about the liability of search engine providers for references to third party content is closely related to the question of the inclusion and exclusion of information providers in a search platform.

3. Liability of Search Engine Providers for Unlawful References

One of the reasons why search engine providers remove references to information providers from their index is that it would be unlawful for them to show these references to their users (for instance because they contain or provide access to defamatory content or link to unauthorised copies of copyright protected material). Clearly, the liability of search engine providers for unlawful references can have a significant impact on the composition of the index of search engines and thereby on the free flow of information facilitated by search engines.

The question about the liability of search engine providers for third party content is a question about online intermediary liability.[43] In the European Union, the liability of online intermediaries for third party content has been specifically dealt with by the Directive on Electronic Commerce.[44] This Directive contains special provisions for the liability of information society services acting as online intermediaries in Articles 12-15. The Directive defines three categories of intermediaries, namely "mere conduit" (Article 12), "caching" (Article 13), and "hosting" (Article 14).[45] For each of these

39) One of the most notable conflicts between publishers and search engines is the Belgium case Copiepresse et al. v. Google et al. - Court of First Instance 13 February 2007, Computer Law Review International, 2007-2, pp.48-50, Summary & Comment by Croux/Sasserath. The case is being appealed. See "Google loses copyright case launched by Belgian newspapers"', International Herald Tribune, 13 February 2007. Available at http://www.iht.com/articles/ap/2007/02/13/business/EU-FIN-Belgium-Google-vs-Newspapers.php (Last viewed 1 July 2008).

40) Supra, note 10, p.14 ("It is essential that search engine providers respect opt-outs expressed by website e*ditors.*").

41) In Germany, the major search engines have entered into a self-regulatory agreement to remove certain references from their index "to improve consumer protection and to protect children and young people from inappropriate content when using search engines". See Subcode of Conduct for Search Engine Providers of the Association of Voluntary Self-Regulating Multimedia Service Providers (*Freiwillige Selbstkontrolle Multimedia-Diensteanbieter* – FSM) (VK-S), 21 April 2004. Available at http://www.fsm.de/en/Subcode_of_Conduct_for_Search_Engine_Providers (Last viewed 1 July 2008). For a discussion of the agreement see Wolfgang Schulz & Thorsten Held, *Der Index auf dem Index? Selbstzensur und Zensur bei Suchmaschinen*, in Marcel Machill & Markus Beiler (ed.), *Die Macht der Suchmaschinen / The Power of Search Engines,* Köln: Halem, 2007, pp. 71-87.

42) For a collection of German and international case law on search engine liability for unlawful references see the collection by Stephan Ott at Links & Law. Available at http://www.linksandlaw.com/legalresources-publicationsenglish.htm. Last viewed 1 July 2008.

43) For a discussion of online intermediary liability see K.J. Koelman, *Liability for on-line intermediaries*, Imprimatur, IViR, Amsterdam, 1997; Rosa Julia-Barceló: *On-Line Intermediary Liability Issues: Comparing EU and US Legal Frameworks*, European Intellectual Property Review, 2000, Vol. 22 (3), 2000, P. 106-119; K.J. Koelman, *Online Intermediary Liability*, in P.B. Hugenholtz (ed.), Copyright and Electronic Commerce, Information Law Series-8, Kluwer, 2000, pp. 7-58.

44) Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on Electronic Commerce), OJ L 178, 17 July 2000, p.1.

45) Generally speaking, "Mere conduit" applies to transmission of third party content, "caching" applies to the temporary storage of third party content for technical reasons and "hosting" applies to the storage of third party content on servers in general.

categories it contains a horizontal[46] conditional liability exemption, often called a "safe harbour". For instance, in the case of the providers of hosting services, the condition tied to the safe harbour is that "(a) the provider does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent; or (b) the provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information". The liability exemptions explicitly leave open the possibility for a court or administrative authority to require the provider to terminate or prevent an infringement. The exemptions do not affect the lawfulness of the processing of information by providers of any of these types of intermediary services. The lawfulness has to be determined by applying the relevant laws of the Member States. Finally, the Directive proscribes general obligations on the providers of these services to monitor the information that they transmit or store or to seek facts or circumstances indicating illegal activity (Article 15).

The Directive focuses on the legal obligations of information society services, which include search engines.[47] But Article 21 on the evaluation by the European Commission of the Directive shows that search engines (location tool services) and hyperlinks are not covered by the intermediary liability regime in Articles 12-15 of the Directive. Article 21 para. 2 provides that "[…] in examining the need for an adaptation of this Directive, the report shall in particular analyse the need for proposals concerning the liability of providers of hyperlinks and location tool services, […]". The exclusion of search engines from online intermediary liability regulation at the EU level has led to a divergent treatment of search engines by legislatures and the judiciary in the various Member States. Because the Directive on Electronic Commerce does not provide for full harmonization of intermediary liability exemptions, the Member States could still regulate the liability of search engine providers and other intermediaries not covered by the Directive. According to the European Commission's first report on the application of the Directive on Electronic Commerce, this option has been used by a few Member States, including Spain, Portugal, Austria and EEA-member Liechtenstein.[48] The new EU Member States Hungary and Poland also extend the limitation on liability for providers of hosting services to information location tools. The safe harbours for search engine providers in Spain, Portugal, Hungary and Poland resemble the safe harbour for hosting providers (Article 14 of the Directive). The safe harbours for search engine providers in Austria and Liechtenstein resemble the safe harbour for mere conduit (Article 12 of the Directive).

The result of the absence of a European safe harbour for search engine providers and the lack of harmonization is a patchwork of laws in Europe governing the liability of search providers for unlawful references. The question is whether this situation, which implies a certain degree of legal uncertainty, is favourable from the perspective of the facilitating role of search engine services with regard to the free flow of information. It goes without saying that extensive duties of care for search engines can result in a chilling effect on legitimate information flows and perhaps the absence of effective search engines altogether.[49] Ideally, the law would set the proper incentives for search engine providers to remove illegal results on the one hand, without chilling the legitimate information flows facilitated by search engines on the other.

There are some examples of national courts taking account of the importance of search engines for the realization of freedom of expression and information. The Paperboy judgment of the German Bundesgerichtshof affirmed the permissibility of the use of hyperlink technologies by crawler-based search engines in a decision referring both to the social utility of search engines and the fundamental right of freedom of expression and information.[50] In a more recent ruling, the Oberlandesgericht

46) This means that the liability limitations apply both to civil and criminal liability, and cover liability for copyright, trademark, defamatory statements, hate speech, etc.

47) As can be seen in recital 18 of the Directive: "[…] information society services are not solely restricted to services giving rise to on-line contracting but also, in so far as they represent an economic activity, extend to services which are not remunerated by those who receive them, such as those offering on-line information or commercial communications, or those providing tools allowing for search, access and retrieval of data; […].", ibid, p. 6.

48) See "First report on the application of Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, 21 November 2003, COM(03)0702", p. 13.

49) For a discussion see Koelman in P.B. Hugenholtz (ed.), *op. cit.*, pp. 40-44

50) Bundesgerichtshof 17 July 2003, I ZR 259/00.

Hamburg builds on this judgment, concluding that the normal liability standard for publications is too strict in the case of the publication of references by search engines and requiring an exception for the case of search engines.[51] The court concludes that the liability of search engine providers for the possibly unlawful content of their references, which is the result of the automatic reproduction of unlawful information from the billions of websites on the Web, should be lowered, because of the significance of search engines for freedom of expression and information.

At this point in time it is unclear whether some kind of safe harbour for the providers of search engines will be adopted at the European level. Article 21 of the Directive instructs the European Commission to conduct a biannual report on the application of the Directive which has to contain an analysis of the need for proposals concerning the liability of providers of hyperlinks and location tool services. In the first report the Commission concluded there was no reason to amend the existing intermediary liability rules with regard to search engines. The Commission did note the diverging legislative choices and wrote the following about legal developments with regard to search engine liability:

"It is encouraging that recent case-law in the Member States recognizes the importance of linking and search engines to the functioning of the internet. In general, this case-law appears to be in line with the Internal Market objective to ensure the provision of basic intermediary services, which promotes the development of the internet and e-commerce. Consequently, this case-law does not appear to give rise to any Internal Market concerns."[52]

The Commission subsequently placed the issue on the agenda of the next review, which, according to Article 21 of the Directive, should have been finalized in 2005. In the second report special attention will be paid to the question of intermediary liability, including the question of liability of search engine providers for third party content. In its review, the European Commission will also have to address the perspective of freedom of expression. Whereas the primary rationale of the Directive lies in the strengthening of the Internal Market for E-commerce and related services, the Directive acknowledges that the "free movement of information society services can in many cases be a specific reflection in Community law of a more general principle, namely freedom of expression as enshrined in Article 10(1) of the Convention for the Protection of Human Rights and Fundamental Freedoms [...]."[53]

4. Conclusion

Search engines have become central in the online information environment and play a crucial facilitating role with regard to online public communications. They mediate between the interests of information providers and end-users that relate to the free flow of information online and have a significant impact on access to information, including audiovisual content. Because of their facilitating role in the free flow of information, more specifically, the ability of information providers to reach an audience and the interest of end-users in searching for the content of their choice, this contribution has focused on the implications of the fundamental right to freedom of expression and information as enshrined in Article 10 of the European Convention on Human Rights for the governance of search engines. Some of the general questions about these implications have been examined, in particular the implications on the governance of the inclusion and exclusion of information providers from a search engine's index. The recent Council of Europe Recommendation on Internet filters has been shown to be an interesting starting point for studying these issues. As it is a related topic, the liability of search engines for third party content at the EU level has also been discussed. This liability regime and the lack of legal certainty could be considered sub-optimal from the point of view of freedom of expression and information. All in all, the legal and regulatory challenges with regard to search engines can be said to be significant. Search engines are multifaceted and complex and some of the questions raised in this contribution will only be answered over time.

51) OLG Hamburg, 20 February 200 – AZ.: 7 U 126/06.
52) First report on the application of Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, 21 November 2003, COM(03)0702, p. 13. Available at http://ec.europa.eu/internal_market/e-commerce/directive_en.htm
53) Directive 2000/31/EC, recital 9.

In Search of the Audiovisual Search Tools in the EU Regulatory Frameworks

Peggy Valcke
Interdisciplinary Centre for Law & ICT of K.U. Leuven
K.U. Brussel

Introduction

In his contribution to the IRIS *plus* 2006-02, Nico van Eijk noted that "the search engine does not have a place in law".[1] He attributes the limited attention search engines have received so far from lawmakers[2] to the fact that they find themselves in the grey zone between electronic communications (transmission) and broadcasting (content). Meanwhile, both the "Television without Frontiers" Directive[3] (see infra 1.3.) – the cornerstone of the EU's broadcasting regulation – and the Electronic Communications Directives[4] (see infra 1.2.) have been (or are currently) subject to a thorough revision. Has this changed (or is it changing) something in the legal treatment of search engines, or in more general terms, content navigation tools? It is the purpose of this contribution to assess the EU regulatory environment and check to what extent the respective frameworks (could) apply to audiovisual search or content navigation tools.[5] In a second part, some preliminary ideas on "how to fill the gaps" will be formulated.

Warning Notice

I would like to start my contribution with the following warning notice: what makes this analysis extremely complicated is the existence of a wide variety of audiovisual search tools. Some search engines order audiovisual material on the Web, as part of a more general or horizontal search site (including also text, images, locally stored files, etc., such as Exalead[6]) or a social bookmarking service

*) Peggy Valcke lectures in media and communications law at K.U.Leuven and K.U.Brussel and has been visiting professor at Central European University in Budapest in 2006-2007; she works as a postdoctoral researcher at the Interdisciplinary Centre for Law & ICT of K.U.Leuven (www.icri.be), which is a founding member of the Flemish Institute for Broadband Technology (www.ibbt.be).

1) Van Eijk, Nico. 2006. Search engines: Seek and ye shall find? The position of search engines in law. IRIS *plus* 2006-2; Strasbourg: European Audiovisual Observatory, available at http://www.obs.coe.int/oea_publ/iris/iris_plus/iplus2_2006.pdf

2) Though not from courts, as is demonstrated by numerous court rulings relating to search engines in a number of law domains (intellectual property, unfair trade practices, data protection…).

3) Council Directive 89/552/EEC of 3 October 1989 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the pursuit of television broadcasting activities, [1989] OJ L298/23, as amended by Directive 97/36/EC, [1997] OJ L202/60.

4) In short: Directive 2002/21/EC (Framework Directive) [2002] OJ L108/33; Directive 2002/19/EC (Access Directive) [2002] OJ L108/7; Directive 2002/20/EC (Authorisation Directive) [2002] OJ L108/21; Directive 2002/22/EC (Universal Service Directive) [2002] OJ L108/51; Directive 2002/58/EC (ePrivacy Directive) [2002] OJ L201/37. These directives are accompanied by Commission Directive 2002/77/EC ("Competition Directive" [2002] OJ L249/21) and a number of regulations, decisions and recommendations (for more information, see: http://ec.europa.eu/information_society/policy/ecomm/current/index_en.htm).

5) Audiovisual search tools and content navigation tools are used as synonyms in this article.

6) http://www.exalead.com

7) http://del.icio.us

(e.g., del.icio.us[7]). Others are dedicated to serving up videos, sometimes combining the search function with (free) hosting capacity for users who want to upload their own videos (e.g., GoogleVideo[8]). In some cases, the focus is no longer on the search function, but on the hosting and distribution functions, as is the case with the mushrooming online video portals. Of these, some focus on user-generated content (like YouTube, Dailymotion, GarageTV or BlipTV)[9] and others on professional channels, shows, movies and music videos (like Joost).[10] Other portals combine both types of content (like Gotuit Media[11]) and/or offer users communication and rating tools to share their experience with friends while they are watching (e.g., instant messaging software or a blog space).[12]

At the other end of the spectrum, we have the traditional (edited) electronic programme guides (EPGs), used in the context of a walled-garden digital television environment. Now that established broadcasters are expanding their territory from the traditional television platforms to the Internet, we find more and more of their programme listings and guides online as well, thereby allowing users to check what is on a channel at that moment, that night, the next day, etc. (one of the many examples is the one offered by the BBC[13]). Similar online listings are also offered by TV guide magazines that reproduce their print product on the Internet – often with additional features and usually for free (i.e., advertising based, without subscription from the user; e.g., TV Guide[14]) – or by new entertainment portals (like the Irish entertainment.ie website[15]). It is standard practice for these to present a grid of television programmes that are shown on the various channels at a certain moment, thereby offering users the option of adjusting grid preferences (e.g., to change the order of the channels), searching for a specific programme and checking additional programme information. Some even offer an online scheduling service, i.e., a tool to schedule recordings of programmes on your digital video recorder by simply clicking the programme in the online grid (e.g., TVGuide which has partnered with TiVo® DVR for that purpose). Contrary to the electronic programme guides on the traditional television platforms, they do not offer[16] the possibility of accessing the channel or programme directly via the online grid. At least that is, not yet; in the light of the growing convergence between the Internet and television, it is only a matter of time before online television listings will enable such accessing and zapping as well, while at the same time, "traditional" electronic programme guides will include similar features enabling access and search of the user-generated content available on the Internet.

In other words, there is an almost endless continuum of search services and products to assist the user in finding his/her way through the "audiovisual abundance". But that common feature – the search function – does not automatically make these services susceptible to equal legal treatment (it is not because a vehicle has wheelsthat you need a driver's license to steer it). It makes a huge difference whether the search function is part of a broader content offering (although in that case the provider may still be largely left "untouched" when it looks more like the electronic equivalent of print or text products), part of a broader audiovisual content offering (in which case the search service may be subject to the same rules applicable to the audiovisual media service which it accompanies – unless the content service does not fulfil all criteria of the legal definition of audiovisual media service; *infra*) or part of transmission services (like the EPGs on traditional television platforms, which are offered by conditional access providers; *infra*).

It is not the aim of this article to offer a complete legal analysis of all possible forms of audiovisual search tools. The intention is to pinpoint relevant provisions in the existing EU regulatory frameworks (i.e., provisions containing a reference to search engines or electronic programme guides), analyse the concepts that they use, as well as the scope of the rules they impose.

8) http://video.google.com/
9) http://youtube.com/ http://www.dailymotion.com/ http://www.garagetv.be/ http://blip.tv
10) http://www.joost.com
11) http://www.gotuit.com
12) For an interesting analysis of these new portals from the perspective of content regulation, see: Ariño, Monica. 2007. Content Regulation and New Media: A Case Study of Online Video Portals. *Communications and Strategies* (66):115 – 135.
13) http://www.bbc.co.uk/cgi-perl/whatson/search/grid.cgi?MEDIUM=tv
14) http://www.tvguide.com/Listings/default.aspx (a similar example taken from the Flemish context is http://www.tvgids.be/)
15) http://entertainment.ie/tv/whats-on-now.asp
16) At least those which we have checked.

1. Content Navigation Tools under EC Regulation for the Info-communications Sector

1.1. In Search of Search Tools...

Audiovisual search tools have to comply with a series of general EU legal provisions, including the competition rules in Articles 81 and 82 EC-Treaty, the rules relating to unfair commercial practices in Directive 2005/29/EC,[17] the advertising rules contained in, for instance, Directives 2006/114/EC (misleading and comparative advertising),[18] 2003/33/EC (tobacco advertising)[19] and 2001/83/EC (medicinal products).[20] These instruments will, however, not be analysed further here.

This contribution aims to focus on the rules specifically adopted for the sectors of broadcasting and audiovisual services, communications, and information – in short the "info-communications sector". EC sector-specific regulation for the info-communications sector is based on a demarcation between transmission and content. The electronic communications directives, adopted in 2002, cover all transmission networks and services (as a result of convergence), but not the content of services delivered over such networks (such as broadcasting content, financial services and certain information society services[21]). Services at the content layer, consisting of the editing and delivery of audiovisual content or information, are regulated by the "Television without Frontiers" Directive[22] (recently amended and renamed the Audiovisual Media Services Directive; *infra*) and by the E-Commerce Directive.[23] Inextricable links between transmission and content, however, render it impossible to regulate one completely separately from the other or to exclude hybrid or "intermediary" measures. The electronic communications framework, for instance, contains rules that apply to transmission networks or associated facilities for content-related reasons (like cultural diversity) – think of the must carry-provisions in Article 31 Universal Service Directive or the rules relating to bottleneck facilities in digital television (*infra*).

1.2. ...in the Electronic Communications Directives

2002 Regulatory Framework

As mentioned earlier, the 2002 Electronic Communications Directives apply to transmission networks and services, as well as the so-called "associated facilities". Since it is obvious that search tools do not constitute a network of their own, we can confine our analysis to the notions of electronic communications services and associated facilities.

The former are defined by Article 2 (c) Framework Directive as "services normally provided for remuneration which consist wholly or mainly in the conveyance of signals on electronic communications networks, including telecommunications services and transmission services in networks used for broadcasting". The definition explicitly excludes "services providing, or exercising editorial control over, content transmitted using electronic communications networks and services" as well as "Information Society services, as defined in Article 1 of Directive 98/34/EC, which do not consist wholly or mainly in the conveyance of signals on electronic communications networks." The exceptions seem rather to be written with broadcasting and on-demand audiovisual services in mind, excluding those which edit that type of service, as well as those which provide (or distribute) such

17) Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No. 2006/2004 of the European Parliament and of the Council ("Unfair Commercial Practices Directive"), [2005] OJ L 149/22.

18) Directive 2006/114/EC of the European Parliament and of the Council of 12 December 2006 concerning misleading and comparative advertising, [2006] OJ L 376/21.

19) Directive 2003/33/EC of the European Parliament and of the Council of 26 May 2003 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the advertising and sponsorship of tobacco products, [2003] OJ L 152/16.

20) Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use, [2001] OJ L 311/67, as amended by Directive 2004/27/EC of 31 March 2004, [2004] OJ L 136/34.

21) Cf. Recital 5 Framework Directive.

22) *Supra*, footnote 3 and *infra*, footnote 39.

23) *Infra*, footnote 47.

content to the public. Neither the editing, nor the provision or distribution of content via the Internet or more traditional platforms, like cable television, is considered an electronic communications service.[24] Search tools act as directory services, facilitating access to third party information, but they do not edit that information or content themselves, and it also seems difficult to argue that they provide the content themselves. At the same time, it cannot be denied that the service they offer does not consist wholly or mainly of the conveyance of signals. The core of their service is not merely to bring signals from point A to point B, but to assist the user in finding the information he/she is looking for by providing references to third party information, ranked, according to relevance, in the light of search terms and/or past user preferences. So, even though search tools may show some aspects of routing services, their added value lies in their functionalities as marketing instruments and content-related services and will therefore fall outside the scope of electronic communications services in the sense of Article 2 (c) Framework Directive.[25]

The notion of associated facilities is defined in Article 2 (e) Framework Directive *as "those facilities associated with an electronic communications network and/or an electronic communications service which enable and/or support the provision of services via that network and/or service. It includes conditional access systems and electronic programme guides."* Hence, EPGs – a form of audiovisual search tool as explained in our introduction – are explicitly covered by the directives. But how far does this notion extend? Does it, for instance, also include search engines on the Internet? Contrary to the concepts of conditional access systems (CAS) and application program interfaces (APIs),[26] EPGs are not further defined. Only by looking at the relevant rules in the Access Directive does it become clear[27] that the notion of an EPG is also confined to the context of digital radio and television broadcasting services and only regulated from the angle of its technical (transmission) aspects:

- Article 5, para. 1 (b) and Article 6 Access Directive (in conjunction with Annex I) contain a specific access regime for associated facilities, more specifically conditional access systems (Article 6), application program interfaces and electronic programme guides (Article 5), *in the area of digital radio and television broadcasting services*.[28] This makes it unlikely that Internet search engines would be considered electronic programme guides in the sense of the electronic communications directives.

- The purpose of these provisions is to provide tools to ensure access on fair, reasonable and non-discriminatory ("FRND") terms to these facilities. Access is to be interpreted in a technical sense here, covering transmission aspects only (access to the underlying technology or routing facilities only). We can infer this not only from the definition of "access" itself in Article 2 (a) Access Directive,[29] but also from Article 6, al. 4, which states that "conditions applied in accordance with

24) See also Recital 10 Framework Directive ("The same undertaking, for example an Internet service provider, can offer both an electronic communications service, such as access to the Internet, and services *not* covered under this Directive, *such as the provision of web-based content*.") and Recital 20 Authorisation Directive ("The same undertaking, for example a cable operator, can offer both an electronic communications service, such as the conveyance of television signals, and services *not* covered under this Directive, *such as the commercialisation of an offer of sound or television broadcasting content services*, and therefore additional obligations can be imposed on this undertaking in relation to its activity as a *content provider or distributor*, according to provisions other than those of this Directive, without prejudice to the list of conditions laid in the Annex to this Directive").

25) Cf. van Eijk (2006) with regard to search engines. See also the recent Article 29 Working Party Opinion on data protection issues related to search engines, mentioned *infra*, footnote 57, "Search engines therefore fall outside of the scope of the definition of electronic communication services" (which does however not prevent the search engine provider from being subject to the Electronic Communications Directives if he also offers additional services that meet the definition of electronic communications service, such as a publicly accessible email service).

26) Respectively Articles 2 (f) and 2 (p) Framework Directive.

27) For conditional access systems and application program interfaces, this is already clear from the definitions themselves.

28) Article 5(1) and Annex 1 Part II of the Access Directive state that in the absence of significant market power (SMP), national regulatory authorities are able to impose obligations on operators to provide access to APIs and EPGs on fair, reasonable, and non-discriminatory terms to ensure accessibility for end-users to digital radio and television broadcasting services specified by the Member State.

29) "Access means the making available of facilities and/or services, to another undertaking, under defined conditions, on either an exclusive or non-exclusive basis, for the purpose of providing electronic communications services. It covers inter alia: access to network elements and associated facilities, which may involve the connection of equipment, by fixed or non-fixed means (in particular this includes access to the local loop and to facilities and services necessary to provide services over the local loop), access to physical infrastructure including buildings, ducts and masts; access to relevant software systems including operational support systems, access to number translation or systems offering equivalent functionality, access to fixed and mobile networks, in particular for roaming, access to conditional access systems for digital television services; access to virtual network services."

this Article are without prejudice to the ability of Member States to impose obligations in relation to the presentational aspect of electronic programme guides and similar listing and navigation facilities." Hence, Member States remain free to regulate content aspects of electronic programme guides and "similar listing and navigation facilities" (like Internet search engines), as these are not covered by the Electronic Communications Directives.

Could search tools, other than electronic programme guides, nevertheless fall within the broader category of "associated facilities"? Possibly, as this notion is formulated broadly (not confined to the digital broadcasting sphere) and it could be argued that search tools "enable and/or support the provision of services via that network and/or service" (just like electronic programme guides do in the area of digital broadcasting). In that regard, it is worth noting that the notion of associated facilities was deliberately extended during the legislative process leading to the adoption of the Framework Directive in 2002 from facilities necessary to provide "electronic communications services" to facilities enabling or supporting the provision of "services" (including broadcast or information society services).[30] The usefulness of such qualification would, however, remain limited. Even though the general regime of significant market power could, in principle, be applied to associated facilities, triggering the application of *ex ante* obligations for providers of those facilities with a position on the market equal to dominance is highly unlikely to happen in the short to medium term. Firstly, such a market is not included in the Commission's Recommendation on Relevant Markets,[31] meaning that national regulatory authorities who would want to analyse it face a substantial burden of proof, in that they would have to demonstrate that the conditions for the so-called three criteria-test are fulfilled.[32] Secondly, the possible remedies that can be imposed always relate to the notion of "access" in a technical sense (as we have already explained), covering transmission aspects only, and are hence not (well) suited to address problems related to content aspects.

The Reform Proposals

The Electronic Communications Directives are currently being revised. The European Commission published its legislative proposals on 13 November 2007, and the European Parliament held its first reading on 24 September 2008.[33] No fundamental changes have been introduced with regard to "associated facilities" or distribution functions, notwithstanding several suggestions by consultants in the context of preparatory studies. We could think of the attention that the Cullen Study of 2006 has drawn to the need for more clarification of the distribution function in the context of the electronic communications framework (as it is now largely in the grey zone).[34]

30) See, for instance, Report Paasilinna (ITRE) on the Proposal for a Directive of the European Parliament and of the Council on a common regulatory framework for electronic communications networks and services (COM(2000) 393 – C5-0428/2000 – 2000/0184(COD)), A5-0053/2001, PE 297.096, 7 February 2001, amendment 19, p.17-18.

31) European Commission Recommendation of 17 December 2007 on relevant product and service markets within the electronic communications sector susceptible to ex ante regulation in accordance with Directive 2002/21/EC of the European Parliament and of the Council on a common regulatory framework for electronic communications networks and services, [2007] OJ L344/65.

32) In order to identify markets that are susceptible to *ex ante* regulation, three cumulative criteria need to be fulfilled: (1) the presence of high and non-transitory barriers to entry, (2) absence of a dynamic market structure that tends towards effective competition within the relevant time horizon, (3) application of competition law alone would not adequately address the market failure(s) concerned. If a market is included in the Commission's list, there is a presumption that the market in question meets these criteria; if not, the national regulatory authority has to provide the relevant evidence.

33) European Parliament legislative resolutions of 24 September 2008 on the proposal for a directive of the European Parliament and of the Council amending Directive 2002/21/EC on a common regulatory framework for electronic communications networks and services, Directive 2002/19/EC on access to, and interconnection of, electronic communications networks and associated facilities, and Directive 2002/20/EC on the authorisation of electronic communications networks and services, P6_TA-(2008)0449 (COM(2007)0697 – C6-0427/2007 – 2007/0247(COD)) – A6-0321/2008 and on the proposal for a directive of the European Parliament and of the Council amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on consumer protection cooperation, P6_TA-(2008)0452 (COM(2007)0698 – C6-0420/2007 – 2007/0248(COD)) – A6-0318/2008. Available at: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+20080924+SIT+DOC+WORD+V0//EN&language=EN

34) Cullen International. 2006. *Study on the Regulation of Broadcasting Issues under the New Regulatory Framework,* 11-14.

The notion of "associated facilities" will remain unchanged when it comes to electronic programme guides. They are not affected by the changes the Commission proposed to Article 2 (e) Framework Directive: "associated facilities means those facilities associated with an electronic communications network and/or an electronic communications service which enable and/or support the provision of services via that network and/or service **or have the potential to do so**, and include **number or address translation systems**, conditional access systems and electronic programme guides, **as well as physical infrastructure such as ducts, masts, street cabinets, and buildings**)".[35] The European Parliament adopted an amendment adding the following facilities, which again have no link with electronic programme guides: "**entries to buildings, building wiring, towers and other supporting constructions**, ducts, **conduits**, masts, **antennae, manholes and** cabinets **and all other network elements which are not active**". Also the suggested (minor) changes to Articles 5 and 6 Access Directive do not alter the legal status of electronic programme guides.

The Commission, however, proposed to extend the notion of "access" in Article 2 (a) Access Directive as follows: "access means the making available of facilities and/or services, to another undertaking, under defined conditions, on either an exclusive or non-exclusive basis, for the purpose of providing electronic communications services **or delivering information society services or broadcast content services.** It covers *inter alia:* access to network elements and associated facilities, […] access to physical infrastructure including buildings, ducts and masts; access to relevant software systems including operational support systems; access to number translation or systems offering equivalent functionality; access to fixed and mobile networks, in particular for roaming; access to conditional access systems for digital television services; access to virtual network services".[36] This would imply that content providers and/or distributors would be entitled to access and interconnection under the same conditions as providers of electronic communications networks and services today and would resolve ongoing discussions about whether content providers and/or distributors can be direct beneficiaries of access obligations imposed on network operators. The proposed modification of Article 20 Framework Directive – which clarifies that disputes between "service providers" would fall within the powers of dispute resolution mechanisms of national regulatory authorities from the moment "one of the parties is an undertaking providing electronic communications networks or services" (meaning that the other party could be a broadcaster or information society service provider) – goes in the same direction.

It remains to be seen whether these Commission proposals will be retained in the final text of the directives. The draft Trautmann Report suggested that the reference to information society services or broadcast content services in the definition of "access" be removed, in order "to prevent the regulation from covering content issues, which would open a much bigger field of litigation (already covered in AVMS and eCommerce Directives)".[37] However, after the vote in the ITRE Committee on 7 July 2008, the final Trautmann Report contained the peculiar amendment 91 which suggests changing the text into "making available of facilities and/or services, to another undertaking, under defined conditions, on either an exclusive or non-exclusive basis for the purpose of providing electronic communications services, **including the delivery of information society services or broadcast**

35) European Commission, Proposal for a directive of the European Parliament and of the Council amending Directives 2002/21/EC on a common regulatory framework for electronic communications networks and services, 2002/19/EC on access to, and interconnection of, electronic communications networks and services, and 2002/20/EC on the authorization of electronic communications networks and services, COM(2007)697 rev1, 13 November 2007, available at:
http://ec.europa.eu/information_society/policy/ecomm/library/proposals/index_en.htm
The text in bold marks the suggested modifications.

36) COM(2007)697 rev1, 13 November 2007. The text in bold marks the suggested modifications.

37) European Parliament (ITRE Committee – Rapporteur: Trautmann), Draft Report on the proposal for a directive of the European Parliament and of the Council amending Directive 2002/21/EC on a common regulatory framework for electronic communications networks and services, Directive 2002/19/EC on access to, and interconnection of, electronic communications networks and services, and Directive 2002/20/EC on the authorization of electronic communications networks and services (COM(2007)0697 – C6-0427/2007 – 2007/0247(COD)), PE398.542v02-00, 23 April 2008, Amendment 60, p. 46 (see also Amendment 615 of Pleguezuelos; Amendments to the Draft Trautmann Report, vol. 5 (amendments 603-716), PE407.731v01-00, 10 June 2008, p. 15). Available at
http://www.europarl.europa.eu/meetdocs/2004_2009/organes/itre/itre_20080707_1900.htm or
http://www.europarl.europa.eu/activities/committees/publicationsCom.do?language=EN&body=ITRE

content services."[38] This amendment contains in our view a contradiction with the definition of electronic communications services (explicitly excluding services providing, or exercising editorial control over, content, *supra*), but was nevertheless adopted by the European Parliament on 24 September 2008.

Proposals in the sphere of consumer protection (e.g., transparency for users; Article 20 Universal Service Directive) remain limited to contractual situations (i.e., relations between subscribers and providers) and to "undertakings providing electronic communications services and/or networks" (and not to providers of associated facilities).

Unless some drastic reforms are proposed by the Council, the new telecoms package will not fundamentally change the legal position of search tools in the context of EU transmission regulation.

1.3. ...in the Audiovisual Media Services Directive

On 11 December 2007, the European legislator adopted its long-awaited Audiovisual Media Services Directive (AVMS Directive), modernising and renaming the "Television without Frontiers" Directive (TVwF Directive).[39] The aim of this Directive has always been (and still is) the creation of an internal market for television services and, in this light, the provision of minimum harmonisation in certain areas of broadcasting policy (including protection of minors and of human dignity, cultural diversity and media pluralism, access to information and commercial communications).

The new Directive covers more than just traditional broadcasting services. Its scope of application includes both linear broadcasting services, irrespective of the underlying technology or platform, and on-demand audiovisual media services (such as video-on-demand). These are grouped under the common heading of an "audiovisual media service", which is defined as "a service as defined by Articles 49 and 50 of the Treaty which is under the editorial responsibility of a media service provider and the principal purpose of which is the provision of programmes in order to inform, entertain or educate, to the general public by electronic communications networks within the meaning of Article 2(a) of Directive 2002/21/EC".

The various elements of this definition – which are cumulative criteria[40] – are further clarified in various recitals. It would fall outside the scope of this contribution to engage in an in-depth analysis of these criteria,[41] so we will look at only those aspects that are relevant to search tools. Recital 18 stresses that the definition of audiovisual media service should only cover mass media in their function to inform, entertain and educate the general public and that it excludes all services whose principal purpose is not the provision of programmes, i.e., where any audiovisual content is merely incidental to the service and not its principal purpose. "Examples include websites that contain audiovisual elements only in an ancillary manner, such as animated graphical elements, short advertising spots or information related to a product or non-audiovisual service."[42] Recital 18 continues that, "for these reasons", games of chance (involving a stake representing a sum of money), lotteries, betting and other

38) European Parliament (ITRE Committee – Rapporteur: Trautmann), Final Report on the proposal for a directive of the European Parliament and of the Council amending Directive 2002/21/EC on a common regulatory framework for electronic communications networks and services, Directive 2002/19/EC on access to, and interconnection of, electronic communications networks and services, and Directive 2002/20/EC on the authorization of electronic communications networks and services (COM(2007)0697 – C6-0427/2007 – 2007/0247(COD)), A6-0321/2008, PE398.542v03-00, 22 July 2008.

39) Directive 2007/65/EC of the European Parliament and of the Council of 11 December 2007 amending Council Directive 89/552/EEC on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the pursuit of television broadcasting activities [2007] OJ L332/27, available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:332:0027:0045:EN:PDF

40) Cf. recital 25: "All the characteristics of an audiovisual media service set out in its definition and explained in Recitals 16 to 23 should be present at the same time."

41) See, for instance, Chavannes, Remy and Castendyk, Oliver, in *European Media Law*, edited by Castendyk, Oliver, Dommering, Egbert, and Scheuer, Alexander, Alphen a/d Rijnn, 2008. Chapter I. Definitions, 813-819. Valcke, Peggy and Lievens, Eva. 2008. Rethinking European broadcasting regulation - Unravelling Europe's policy for the digital media landscape: critical analysis of the Audiovisual Media Services Directive. In *Rethinking European Media and Communications Policies*, edited by Pauwels, Caroline, and Kalimo, Harri. Brussel: VUB Press (forthcoming).

42) Think of the website of a travel agent or car manufacturer where clips of holiday resorts, respectively the latest car model are shown.

forms of gambling services, as well as on-line games and *search engines* should be excluded from the scope of the Directive.

Since in most cases search engines on the Internet will not fulfil the criterion of providing programmes as their principal purpose, they will not qualify as audiovisual media services, even if they were to occasionally show audiovisual material.

If the search engine were to be a means of accessing video clips – like GoogleVideo– some commentators argue that it might potentially qualify as an on-demand audiovisual media service, amounting as it does to a catalogue of programmes.[43] In their view, this will depend on whether such an offer would be sufficiently "television-like" (cf. recital 17) to be "comparable to the form and content of television broadcasting", as Article 1 (b) of the Directive (containing the definition of "programme") puts it, and consequently on how the search engine is presented to the public. They refer to a version of YouTube (which clearly has search functions) made available through an AppleTV as an example that might very well qualify.

In my view, however, the absence of editorial responsibility in the sense of the Directive will prevent such an audiovisual search tool from being considered an audiovisual media service. This notion of "editorial responsibility" has – at the instigation of Parliament and Council – become a key element of the definition of audiovisual media service.[44] It means "the exercise of effective control both over the selection of the programmes and over their organisation either in a chronological schedule, in the case of television broadcasts, or in a catalogue, in the case of on-demand audiovisual media services" (but does not necessarily imply any legal liability under national law for the content or the services provided; Article 1 (c) AVMS Directive).

It is difficult to argue that the machine-powered generation of search results based on algorithms is similar to the effective control over the selection and organisation of programmes in a chronological schedule or catalogue by a broadcaster or video-on-demand provider. Even if search engines dedicated to video or video sites like YouTube adopt a policy of removing videos that have been flagged by the users of the community themselves as inappropriate, this intervention is *ex post,* acting upon complaints from users.[45] Moreover, the editorial control is what Van Hoboken calls "distributed", i.e., shared with the information providers and Web users themselves tagging and ranking the content they post and consume online.[46]

What if the selection and presentation of audiovisual material would be performed by human-powered organisations with editorial centres? In that case, we come close to the notion of electronic programme guides, which are mentioned in recital 22: "While the principal purpose of an audiovisual media service is the provision of programmes, the definition of such a service should also cover text-based content which accompanies programmes, such as subtitling services and electronic programme guides. Stand-alone text-based services should not fall within the scope of this Directive, which should not affect Member States' freedom to regulate such services at national level in accordance with the Treaty."

In other words, electronic programme guides are not considered an audiovisual media service in themselves (since their principal purpose is not the provision of programmes, but the listing of those programmes with additional information and functionalities like zapping, recording, etc.), but if they accompany an audiovisual media service, they nevertheless fall under the scope of the Directive. The electronic programme guide is then considered part of the service package of the audiovisual media service provider and, as such, should follow the rules of the Directive as well (for instance, those relating to the protection of minors and of human dignity, or *mutatis mutandis* the rules on audiovisual commercial communications).

43) Chavannes, Remy and Castendyk, Oliver. 2008. *op.cit.*, 814.
44) See Schulz, Wolfgang and Heilmann, Stefan. 2008. In IRIS Special: Editorial Responsibility, edited by Nikoltchev, Susanne, Strasbourg, European Audiovisual Observatory.
45) See, for instance, YouTube Community Guidelines at http://www.youtube.com/t/community_guidelines
46) Van Hoboken, Joris. 2008. Legal Space for Innovative Ordering – The need for more clarity regarding the liability of selection intermediaries for third party information. Paper presented at the EuroCPR2008 conference, Seville, March 2008, at p. 4.

A contrario, if the electronic programme guide is offered as a stand-alone service (i.e., not part of an audiovisual media service), for instance by the cable or satellite operator, the rules of the Directive no longer apply.

1.4. ...in the E-Commerce Directive

The E-Commerce Directive,[47] adopted in 2000, intends to create a common fundamental legal framework for electronic commerce in the Internal Market and stimulate the development of cross-border online services. It applies to the so-called "information society services", defined as "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services".[48] This notion covers a wide variety of services provided online, ranging from online newspapers, specialised news services (such as business or financial information) and online selling of various products and services (like books, DVDs, computer hardware and software, travel services, pharmaceuticals, etc.) to the online provision of financial services (online banking and online investment).

Recital 18 of the Directive reads: "…information society services are not solely restricted to services giving rise to on-line contracting but also, in so far as they represent an economic activity, extend to services which are not remunerated by those who receive them, such as those offering on-line information or commercial communications, or those providing tools allowing for search, access and retrieval of data". In other words, online search engines are, in principle, included in the scope of application of the Directive.

The relevant rules, however, remain limited. Firstly, because the coordinated fields in the Directive and the harmonisation achieved by it remain limited in scope, and secondly, because the rules that seem most relevant for search engines – notably the liability rules for intermediary service providers – do not apply to them (only to mere conduit, hosting and caching activities). We will not further address this issue, as it is dealt with in more detail in the contribution of Joris van Hoboken.[49]

1.5. Interim Conclusion

The previous paragraphs have shown that the current EU legal framework deals only in a very fragmented manner with audiovisual search tools. Limited provisions exist – both in the Electronic Communications Directives and in the Audiovisual Media Services Directive – with regard to electronic programme guides (which we interpret here as "edited audiovisual search tools", linked to digital broadcasting in a walled-garden environment). In contrast, online search tools – even though they are included in the general scope of application of the E-Commerce Directive – are not even covered by the safe harbour provisions in the E-Commerce Directive. In other words, not much has changed with regard to the legal position of search tools since van Eijk's analysis in IRIS *plus* 2006-2 ("Seek and Ye Shall Find?").

2. Preliminary Thoughts on Regulating the Audiovisual Search Function

2.1. Do we Need to Regulate?

A fundamental question is whether we need to regulate at all. We need to have good arguments in order to regulate functions for which we have never enacted sector specific regulations in the print and analogue environment – think of TV listing guides, bookshops, newspaper kiosks, libraries; they

47) Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), [2000] OJ L 178/1, available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:178:0001:0016:EN:PDF

48) Article 1 (a) Directive 98/34/EC of the European Parliament and of the Council of 20 June 1998 laying down a procedure for the provision of information in the field of technical standards and regulation, [1998] OJ L204/37, as amended by Directive 98/48/EC of 20 July 1998, [1998] OJ L217/17.

49) See also van Hoboken, Joris. 2008. Legal Space for Innovative Ordering – The need for more clarity regarding the liability of selection intermediaries for third party information. Paper presented at the EuroCPR2008 conference, Seville, March 2008. See also the article by Joris van Hoboken in this publication.

have all been selecting and facilitating access to content for many years, without anyone defending the need to subject them to special neutrality or other rules. What is so different about the online, digital environment – apart from the fact that we are, at this stage, not as familiar with it and that the providers of the search services can be located virtually anywhere on the planet?[50] In the context of online text-based search engines, the (hidden) manipulation of search results and the dominance of one provider (Google) have been mentioned as possible reasons motivating the introduction of regulatory intervention.[51] But such issues have not yet been brought to the forefront with regard to audiovisual search tools.

2.2. Whom/what Do we Want to Regulate?

The second question: whom/what do we want to regulate? Only audiovisual search tools that involve some form of editorial control (like the prior selection of the channels or programmes that will be listed and be presented in a searchable form) or also "automated" search services (with removal of certain references *ex post*, upon receipt of complaints from users, at the maximum)? There is something to be said for the first option, since these imply a greater control by the provider. But, ironically, this may lead to perverse effects: the more you control, the more you will be regulated, so let us control less…

2.3. Why do We Want to Regulate (and how)?

Thirdly, what are the policy goals that would need regulatory intervention (and how could they best be achieved)? Concerns over market power and market foreclosure might trigger intervention, but the first tools for realising free and fair competition are the general competition rules (in the first place, in our case, the prohibition of abuse of a dominant position in Article 82 EC Treaty).[52] Only in sectors that were formerly characterised by monopolies, have sector specific instruments been introduced allowing the imposition of *ex ante* remedies on undertakings with a certain position on the market (like the significant market power regime in telecommunications). The digital television sector has been a peculiar case, with access obligations for conditional access providers irrespective of their market position. The current Access Directive inherited this regime and created comparable (though not similar) rules for "new" digital bottlenecks, like application programme interfaces and electronic programme guides. As discussed already, these rules relate to transmission aspects of those facilities only and say nothing about presentational or content issues.

When it comes to cultural objectives (such as media pluralism), the need for intervention could arise when certain content that is considered of public importance would no longer appear at a visible place in the search list. Therefore, the suggestion made by Cullen in the 2006 Study to specify that must-carry guarantees access to the transmission network and to all the associated facilities – in their view to be interpreted as all distribution functions, hence covering, besides CAS, API and EPG, also DRMs, set-top-boxes, the multiplex, and search engines[53] – seems worth considering, but may be only a first step. Indeed, in an era of abundance, the important thing is not merely to be listed, but to be listed in a *prominent* place. The German provisions on "Navigatoren" in the *Satzung digitaler Zugang*, containing the obligation for a "balanced" listing of the programme offer of both public and commercial channels, can be cited as an example of provisions pursuing that aim.[54]

---

50) Although if they really want to target a specific geographically defined audience, they are likely to have at the medium to long term at least a local branch or representative to stay in touch with that audience's culture and preferences, and establish close contacts with local advertisers.

51) Van Eijk 2006, *op. cit.* at p. 6 and van Eijk, Nico. 2007. Search Engines, the new bottleneck for content access. Paper presented at the International Telecommunications Society 19th European Regional Conference, Istanbul, 2-5 September 2007, at p. 20 (available from http://www.ivir.nl/publications/vaneijk/Paper_SearchEngines_ITS_2007.pdf)

52) Merger control rules could also come into the picture if powerful search tool providers started a process of horizontal or vertical integration through alliances and take-overs, or the prohibition of cartel agreements and concerted practices may apply to (exclusive) vertical agreements between content providers and search tool providers.

53) Cullen International, *op. cit.*, footnote 344, pp.120 and 125.

54) Satzung über die Zugangsfreiheit zu digitalen Diensten gemäß § 53 Abs. 6 Rundfunkstaatsvertrag, § 13 (3): *„Auf das öffentlich-rechtliche und private Programmangebot muss gleichgewichtig hingewiesen werden."* (adopted by the Direktorenkonferenz der Landesmedienanstalten (DLM) on 13 December 2005, available at http://www.alm.de/fileadmin/Download/Gesetze/Satzung_gem.____53_-End-.pdf).

Consumer protection might be another relevant policy goal in the area of audiovisual search tools. Under this heading, rules seeking to enhance transparency of ranking methods (and "sponsored" links) could be envisaged, as well as rules dealing with the removal of (and/or the warning for) search results referring to illegal (and/or harmful) content (like setting the conditions and procedures for such removal or warning notices). As an example of this, we can refer to the code of conduct for search engines in Germany, which is described in the contribution of Wolfgang Schulz.[55]

Privacy protection could be a rising star in the firmament of policy goals justifying (more) regulatory intervention. Concerns about people's privacy are growing in the light of emerging technologies designed to trace and recall past user activity online or on digital TV platforms in order to personalise services. As was demonstrated by Michael Zimmer,[56] dozens of products and services are being linked via cookies and user accounts and people are leaving behind an increasing amount of "digital fingerprints" on the Internet (like contacts, blogs, books, stocks, appointments, e-mail, friends, computer files, discussion groups, URLs). Moreover, they lose control of what Zimmer calls "their likeness online", since they might never know what personal data relating to them (including audiovisual material, like photos or videos) other people take and place online (let alone consent to it). In this respect, we would like to draw the reader's attention to the recently adopted Opinion 1/2008 on data protection issues related to search engines, in which the Article 29 Working Party – urged on by growing concerns among European users of search engines about potential breaches of their privacy rights – scrutinises search engine activity and data storage under the current EU directives relating to the protection of personal data and data retention.[57] This opinion clarifies that search engine providers, depending on their exact activities, will in many instances be considered as controllers of personal data and have to comply with the applicable rules and principles. These include the duty to delete or make anonymous personal data once they are no longer necessary for the purpose for which they were collected, as well as the need for search engine providers that specialise in the creation of value added operations, such as profiles of natural persons – the so called "people search engines" – and facial recognition software on images not only to have a legitimate ground for processing (such as consent), but also to meet all other requirements of Data Protection Directive 95/46/EC (such as the obligation to guarantee the quality of data and fairness of processing).

3 Final Comments

At present, both electronic programme guides and search engines fall largely outside the scope of the sector-specific media and (tele)communications regulatory frameworks. Nevertheless, we are of the opinion that most policy goals about which concerns are growing in the context of search tools can be fairly addressed by applying existing and/or general principles or rules, such as competition rules (to address market power concerns), data protection legislation (to cater for privacy concerns) or rules on unfair commercial trade practices and advertising (to enhance transparency on the "manipulation" of results). It may be more a matter of ensuring consistency with these rules, rather than having to enact new rules.

We agree, though, with worried commentators[58] that developments should be monitored closely, as the search function indeed becomes crucial in an era of abundance (perhaps even more important than the provision of content itself, now that we are witnessing an unprecedented lowering of the threshold for becoming a content provider yourself).

55) See the article by Wolfgang Schulz in this publication.
56) See the article by Michael Zimmer in this publication.
57) Article 29 Working Party, Opinion 1/2008 on data protection issues related to search engines, 4 April 2008, WP148. The Article 29 Working Party is a European Union advisory body comprised of representatives of the Member States' data protection agencies. Opinions of the Article 29 Working Party are not legally binding, but often become the "EU standard", and therefore are highly influential in shaping EU policy as well as Member State implementation in the privacy field.
58) See, for instance, the critical comment with regard to the lack of transparency in internet search engines and the hidden manipulation of search results : Machill, Marcel, Neuberger, Christoph and Schindler, Friedemann. 2003. *Transparency on the Net: functions and deficiencies of Internet search engines*. 5 *Info* 1, pp.52-74.

In the audiovisual context, the search function is often bundled with other distribution functions (like the provision of a platform for hosting and sharing audiovisual content), which makes us wonder whether potential regulatory attention should not be broader than the search function itself. In some EU Member States[59] and also in Japan[60], legislators have introduced or are introducing a three-layered regulatory approach for the info-communications sector, distinguishing between content, distribution and transmission.[61] Hence, the providers of "intermediate functions" (i.e., intermediate between the editing and provision of content, on the one hand, and the provision of transmission networks and services over and via which that content travels, on the other hand) would fall under a common regulatory framework. If this were to be the case at EU level as well, this would allow the European legislator to set down the basic rules and principles that would apply (like openness and fairness), while at the same time leaving scope for further refinement of these basic rules through – in first instance – co- and/or self-regulatory instruments. After all, finding the right balance between conflicting policy goals and/or rights and freedoms may be a lengthy process. For instance, where does the freedom of expression – and hence the freedom to adopt an active selection policy – of the search provider end and where does his obligation of non-discrimination start? How should we correlate the search provider's liability with the extent of active selection? Such questions will undoubtedly occupy us for many more years to come.

There is a risk in regulating too much and too soon (which may provide the wrong incentives and stifle innovation); the promises of new technologies should only be tempered when the threats become apparent and serious (although we should not strike the Faustian bargain that has been referred to by Michael Zimmer).[62] Perhaps there will come a time when we will curse search tools (once they start to know us better than we know ourselves), but for the time being I believe they can still be considered extremely useful tools for the detection of relevant information. Admittedly, there is a lot of information that we do not find with the popular search tools because they rank other items first for commercial reasons or because they filter certain things out, but wouldn't there be *a lot more* information that we would not find *without* them (and no, I'm not sponsored by one of them to write this)?

Certainly, providers of search services should be aware of their increasing role in the structuring of the public sphere and their growing impact on society. Therefore, they should be prepared to assume their social responsibility and, if they fail, public powers may have to intervene to "bind" the private powers. In this light, it is our role as academics to foster and continue the public discourse on these issues, since it cannot be left to the providers themselves alone to set the appropriate standards; they are not expected to think and act in the general interest. Initiatives for further analysis and research are extremely valuable in this context. So, search tools, be warned: you are not only watching us, we are also watching you!

59) For instance Belgium (the French Community introduced a three-layered approach in its 2003 Broadcasting Act – as we described in our previous contribution "The Future of Must-carry". 2005. In IRIS Special: To Have or not to Have Must-carry Rules, edited by Nikoltchev, S., Strasbourg, European Audiovisual Observatory – and the Flemish Community will adopt a similar three-layered approach in its new Media Act, implementing the AVMS Directive); see also the French Communications Act (Loi n°86-1067 du 30 septembre 1986 relative à la liberté de communication, "Loi Léotard") that uses the concept of "distributor" (in particular in the context of digital terrestrial broadcasting).

60) In its presentation at the 17th Biennial Conference of the International Telecommunications Society in Montréal (24-27 June 2008), Prof. Minoru Sugaya (Institute for Media and Communications, Keio University) presented the current policy reforms in Japan, explaining how Japan intends to integrate its existing regulatory frameworks for (tele)communications and broadcasting into a single comprehensive "information and communications law", structured along the lines of "content", "platforms" (as intermediate layer) and "transmission infrastructure" (http://www.canavents.com/its2008/plenary3/13.pdf ; for more information about the recommendations of the Study Group on a Comprehensive Legal System for Communications and Broadcasting to which the presentation of Sugaya refers, see: http://www.soumu.go.jp/joho_tsusin/eng/Releases/NewsLetter/Vol18/Vol18_21/Vol18_21.html).

61) Or actually four-layered, if we consider terminal equipment as the fourth layer to be regulated.

62) See the article by Michael Zimmer in this publication.

Regulating Search Engines? On the Use of Self- and Co-regulation in the Field of Internet Search

*Wolfgang Schulz**
Hans-Bredow-Institut

Introduction

When media managers from different sectors congregate nowadays, it is likely – at least in Germany – that they will begin squabbling among themselves about the economic challenges they face and the perceived cut-throat character of the competition. Yet, there is one common topic which brings them all back on speaking terms again: the looming fear of Google eating away at the advertising market.

Hence, the regulation of search engines, their possible influence on public opinion making, their ability to discriminate, and other regulatory issues have climbed up the policy agenda. But there is no clear idea of how to deal with the phenomenon. This contribution outlines the concept of information intermediaries, such as search engines, it identifies which public policy goals could be endangered and, finally, it asks whether concepts of co- and self-regulation might be adequate to meet the regulatory challenges.

1. Concept of Information Intermediaries

Concepts of media regulation implicitly or explicitly make assumptions concerning the functioning of individual and public communication. The specific protection of mass media under Article 10 ECHR or Article 5 Section 1 *Grundgesetz* (the German Constitution – GG) is based on the presumption that mass media fulfil a specific function when it comes to public and individual communication in a democratic society.[1] The German constitutional court has explicitly emphasized the role of mass media and distinguishes between the protection of speech, i.e., of the actual content of communication regardless of whether it is uttered as individual communication or is published by mass media services, and the protection of mass media services (broadcasting, press and film are mentioned in the Constitution, but it can encompass new mass media services as well).[2] From this assumption, the federal constitutional court deduces specific elements of constitutional protection against State interference, as well as the constitutional obligations for the lawmaker to ensure that the power of public opinion making – that in principle is exercised if one controls a mass medium – are controlled effectively.

---

*) Director of the Hans-Bredow-Institut, lecturer at the Law Faculty of the University of Hamburg

1) Hoffmann-Riem/Schulz in: *Politische Kommunikation in der demokratischen Gesellschaft*, Wiesbaden 1998, 154 (159 f.) (in German only); Hoffmann-Riem, *Regulating Media*, New York 1996, 119.

2) Cf. Decisions of the German Constitutional Court (in German only): BVerfGE 73, 118 (158); 74, 297 (324); 83, 238 (298); 90, 60; see also the latest decision: BVerfG, 1 BvR 2270/05, 11 September 2007 (in German only) available at: http://www.bverfg.de/entscheidungen/rs20070911_1bvr227005.html

The specific role of mass media communication is an assumption which is still in line with actual findings in communications science today. Yet, there has been a structural change in public communications. While mass media are characterised by controlling publication as such and, furthermore, by editorial pre-selection, more and more forms of communication have emerged where other types of service providers are of structural importance. The latter means that State interference with those services not only affects a single act of communication, but the overall structure of public communication. I would like to call these other types of service providers "service intermediaries". Communication science is still in the process of characterizing types of online services which might form the basis for future legal debate. Certain types of platforms such as Internet forums might be seen as "information intermediaries".

The important role of search engines has been proven in several studies.[3] Whether it is appropriate to say that, as a precondition, to exist on the Internet it is necessary to be indexed by a search engine[4] or that this might rather be an exaggeration, is anybody's guess. However, it is evident that search engines fulfil a basic function for making information on the Internet accessible. Even if users are becoming less dependent on search engines as they become increasingly adept at using the Internet, this function still remains important.

At first glance one could say that search engines, as special intermediaries in communication, are completely different to classical media service providers. There is no journalistic-editorial pre-selection. The search engine only shows what the user has specified in his or her query. Schönbach has emphasized this difference by calling traditional media "display media". It is a pre-selection which shall surprise the user (in the case of classical media services you want someone else to choose the content according to a given professional standard and, therefore, to surprise you by the content). On the other hand we find the so called "research-media", of which search engines are one very prominent, as well as paradigmatic, example.[5] However, on closer examination, this classification of search engines does not completely match all relevant factors, especially when taking into consideration the fact that some of the algorithms used by search engines try to target the most up-to-date pages or that search engines use indicators of prominence which are relevant for journalistic-editorial choices as well. Still, the main function for which search engines are being used remains the detection of a specific piece of information about which the user already has some clue, but the exact location of which he does not yet know.

The specific role of information intermediaries in individual and public communication may lead to specific risks triggered by these intermediaries. Nevertheless, the specific function fulfilled by search engines for individual or public communication might also require specific protection. This might call for specific safeguards against regulation, rather than suggest legislative action. The choice depends on the specific construction of the regulatory framework, which cannot be elaborated upon here. In this context, the fundamental rights and, in particular the guarantee of freedom of speech, has to be kept in mind.[6]

3) Cf. Van Eijk, Nico. *Search engines: Seek and ye shall find? The position of search engines in law*. IRIS *plus* 2006-2; Strasbourg, European Audiovisual Observatory; available at: http://www.obs.coe.int/oea_publ/iris/iris_plus/iplus2_2006.pdf ; Philippi, Internet Suchmaschinen: Suchmethodik – Erlösmodelle – Preisstrategien, (2006); Nitsche, Suchmaschinen im Internet - Bedeutung, Trends, Chancen und Entwicklungen (2005); Machill/ Neuberger/ Schweiger/ Wirth, Wegweiser im Netz - Qualität und Nutzung von Suchmaschinen (2003); Introna/Nissbaum, Shaping the Web: Why the Politics of Search Engines Matters, *The Information Society* 171 (2000); Maurer et al., Report on dangers and opportunities posed by large search engines, particularly Google (2008), available at: http://www.google-watch.org/gpower.pdf

4) Cf. Introna/Nissbaum, Shaping the Web: Why the Politics of Search Engines Matters, *The Information Society* 171 (2000).

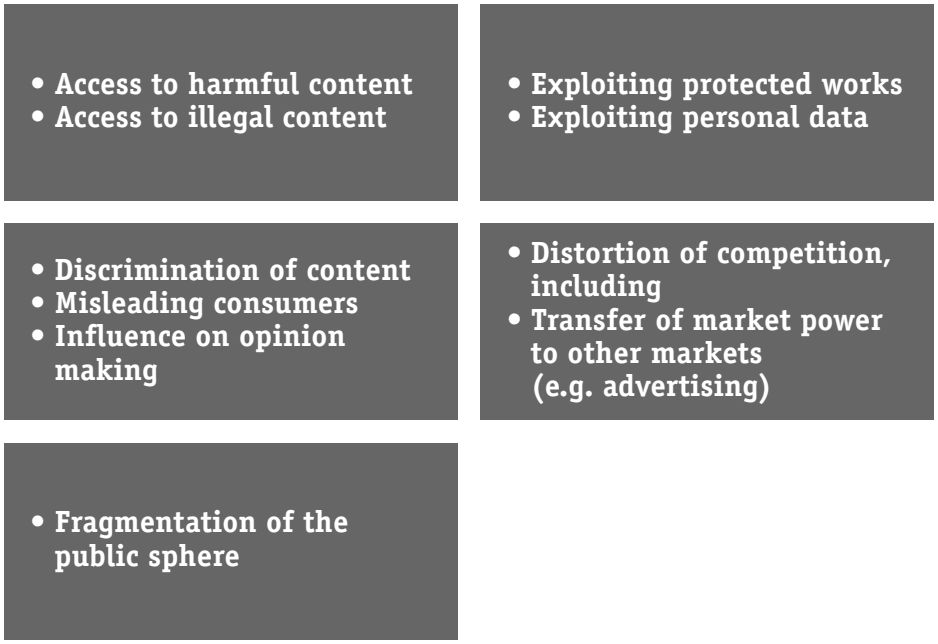
5) Cf. Schönbach, Das Eigene im Fremden. Zuverlässige Überraschung – eine wesentliche Medienfunktion? [The own in the foreign. Reliable surprise: an important function of the media], *Publizistik*, 2005 (50): 344-352.

6) For search engines relating to German constitutional law see Schulz, Von der Medienfreiheit zum Grundrechtsschutz für Intermediäre? CR 2008, S. 470 *et seq*.

2. Risks Triggered by Search Engines

This contribution shall focus on the risks presented by Internet search engines and the possible regulatory approaches for facing these risks.

The risks arising from Internet search engines are manifold:



As a downside of the important role that search engines play as gateways (if not gatekeepers) to the Internet, each and every legal issue on the Internet can sooner or later be found to affect them.

The first block of risks triggered by search engines is that they make harmful content or illegal content (more easily) accessible. What is considered to be harmful or illegal varies from country to country. In most countries at least, child pornography[7] is illegal without exception and in most countries there are also specific restrictions with regard to the access of young persons to content which might impair their development.[8]

Another set of risks might be triggered by the possibility for search engine providers to choose which pages they will include in their index, to remove pages from their index and to arrange their result lists. This might lead to discrimination of content and, in the case of commercial communication, mislead consumers as well as, on the whole, influence the process of public opinion making. During the war in Iraq rumours circulated that certain search engines had removed pictures of US soldiers torturing prisoners from their indexes.[9] Given the fact that Google provides over 80 % of the Internet search results requested in Germany, it is no exaggeration to state that the first listings on Google's search results on a specific topic are highly relevant for public opinion making.[10] Furthermore, this has to be seen against the background of a shift from traditional media, such as broadcasting and press,

7) Cf. for example, §4 JMStV (Germany); Federal Law 18 U.S.C. §2252 (USA); for Europe also see http://www.saferinternet.org/ww/en/pub/insafe/safety_issues/faqs/pornography.htm ; harmonisation by the Council of Europe by way of the Convention on Cybercrime (ETS No. 185).

8) See for example: Awareness and Harmful Media Content in the Digital Age, Ulla Carlsson (Ed.) (2006), available at: http://unesdoc.unesco.org/images/0014/001469/146955E.pdf

9) In regard to Internet Filtering cf. Deibert/ Palfrey/Rohozinski/Zittrain, Access Denied: The Practice and Policy of Global Internet Filtering (2008).

10) Cf. Schulz/Held/Laudien, Search Engines as Gatekeepers of Public Communication Analysis of the German Framework Applicable to Internet Search Engines Including Media Law and Anti-trust Law, in: *German Law Journal,* 2005 (vol 6, no 10), pp. 1419-1433, available at*:* http://www.germanlawjournal.com/pdf/Vol06No10/PDF_Vol_06_No_10_1419-1432_Developments_SchulzHeldLaudien.pdf

to online media, due to the informational needs of specific user groups, while the traditional mass media services still remain important for public communication.[11]

Returning to the differences between display-media and research-media mentioned above, some say that the fragmentation of the public sphere has already led to more and more special interest services (and vice versa) and has been pushed for by research media. To begin with, empirical studies show that there is a significant change in the overlapping of relevant issues on an individual's agenda in comparison with those on the agenda of other citizens, if individuals inform themselves mainly through the use of research media.

Furthermore, it cannot be denied that search engines provide (easy) access to protected works and to specific services, like searches for images or audiovisual content. In the case of images, the role of the search engine providers is more active than just making the links to works available to the public, seeing as the search engines create thumbnails of these pictures. However, it is arguable that copyright is not infringed.[12] The same is largely applicable to audiovisual content.

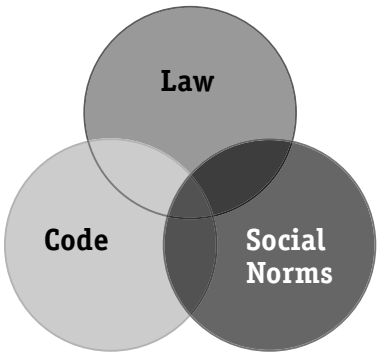
Search engines are among the most frequently visited sites on the Internet and, furthermore, the queries submitted to them by users make an excellent basis for personal profiling. Therefore, data protection issues arise from the use of search engines as well.[13] As far as portrait search is concerned, other aspects of personal rights might also be threatened.

Finally, since in several countries one and the same search engine provider has gained an extremely high market share for both search services and targeted online advertising, the problem of competition distortion has to be addressed as well. Given that companies like Google extend their activities into other fields, like advertising services,[14] the transfer of market power to other markets, like the advertising market, also increases in relevance.

3. Regulatory Approaches

3.1. General Remarks

Based on the fact that one of the above-mentioned risks is considered as being regulated and assuming that the current regulation is not sufficient to address the existing problems, the right regulatory choice has to be made. On a very abstract level there is a regulatory triangle describing which regulatory concepts have to be taken into consideration:



11) Cf. Ridder/Engel, Massenkommunikation 2005: Images und Funktionen der Massenmedien im Vergleich. In: *Media Perspektiven*, Nr. 9/2005, S. 422 *et. seq*.

12) Cf. Wimmers/Schulz, *Wer nutzt? - Zur Abgrenzung zwischen Werknutzer und technischem Vermittler im Urheberrecht*, CR 2008, 170 ff.; LG Hamburg Case No. 308 O 449/03, MMR 2006, 697-699; LG Erfurt Case No.: 3 O 1108/05, MMR 2007, 393-394; Commented on by Berberich in CR 2007, 393-395 and Roggenkamp in K&R 2007, 328-330; OLG Jena, Judgment of 27 February 2008 - 2 U 319/07. For further information about legal aspects of search engines, linking and framing and German Copyright Law see http://www.linksandlaw.com by Stephan Ott.

13) See for the debate on European level the Article 29 Group, Data Protection Working Party, available at: http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf

14) In this context especially the doubleclick-deal on Google is of importance, Cf. http://www.businessweek.com/technology/content/apr2007/tc20070414_675511.htm

Social norms, like professional rules of journalism, govern certain kinds of social behaviour; there are settings of norms which specifically apply to the Internet (e.g., netiquette). There might be explicit rules for certain groups, like codes of conduct or, on a lower level of formality, but even more common, internal rules, which nevertheless can be powerful. Legal regulation can tie in with these social norms, but legal regulation can also be eroded by them or weakened in its enforcement, especially when interfering with social norms. That may occur in particular if social norms are not taken into account by the legislator when drafting laws.

On the Internet especially, the software and hardware architecture also imposes constraints, which are known as code.[15] It is less common for legal regulation, though, to take this type of code into consideration when designing a regulatory framework, which might produce a weakness of regulatory concepts for Internet regulation in the future.

3.2. Co- and Self-regulation

The specific regulatory approaches on which this article focuses are self- and co-regulation. The use of co-regulation is encouraged by the EU; explicitly in Art. 3 of the Audiovisual Media Services Directive (which does not apply to Internet search engines), and also in general terms as part of the concept of better lawmaking.[16] Talking about co-regulation and self-regulation means talking about the interaction between State regulation and non-State regulation and the leeway each side has in determining the regulatory concept. On the one hand, there is the ideal of "command-and-control-regulation", entirely designed and enforced by the State, while, on the other hand, there is pure self-regulation (which in practice is usually surrounded by law, but is, for the issue in question, entirely in the hands of non-State control).

One general advantage of self-regulation lies in the expertise of the actors, which is most effective if they can take their own decisions. This is in addition to the efficiency, the problem-orientation, the flexibility and the high acceptance for this type of regulation. But self-regulation is generally criticized as lacking democratic and legal legitimacy, circumventing mechanisms to protect basic freedoms (enforced private censorship), lacking accountability and transparency, being unfair and causing legal uncertainty, while, at the same time, arousing fears about fragmentation of the law.

Co-regulation is a mixture of both the above-mentioned regulatory approaches. From the Hans-Bredow-Institute's point of view, co-regulation can be defined as follows: "Co-regulation means combining non-State regulation and State regulation in such a way that a non-State regulatory system links in with State regulation".[17]

The latter definition specifies the following criteria which must be met by a regulatory concept so as to classify it as being co-regulation:

The non-State element of the regulatory system includes:

- Creation of specific organisations, rules or processes
- in order to influence decisions
- performed by the addressees themselves

15) Cf. Lessig, CodeV2, available at: http://codev2.cc/download+remix/
16) Cf. Interinstitutional Agreement on Better Lawmaking 2003/C 321/01.
17) Cf. Baldwin/Cave, Understanding regulation: Theory, Strategy, and Practice. (1999); Ayres/ Braithwaite, *Responsive Regulation*, (1992); Hans-Bredow-Institut, Study on co-regulation measures in the media sector, available at: http://ec.europa.eu/avpolicy/info_centre/library/studies/index_en.htm ; Held/Schulz, Regulated self-regulation, available at: http://www.hans-bredow-institut.de/publikationen/apapiere/8selfreg.pdf

With regard to the link between a non-State regulatory system and State regulation, one can speak of co-regulation if the following criteria are met:

- Achievement of public policy goals
- Legal connection
- Discretionary power of the non-State regulatory system
- State uses regulatory resources[18]

Co-Regulation requires special conditions. First of all, a certain regulatory culture is needed. This is especially important for models of co-regulation that depend on industry associations to draft codes or enforce rules, for example. Moreover, incentives for the co-operation and enforcement of rules are of essential importance. The motivation for the industry players to co-operate among themselves has to be adequately high. In most cases the incentive lies in the avoidance of State regulation: The State must therefore reduce its regulatory power significantly. However, it must still appear likely to the industry that the State will implement its own regulation, if a failure of non-State regulation becomes apparent.

Beside the incentives, a functioning co-regulatory system still has to enforce the regulation by implementing adequate and proportional sanctions. Another condition for a co-regulatory system is that the State should retain resources that are used to influence the outcome of non-State regulatory processes. These resources are necessary, because the State remains responsible for guaranteeing a sufficient level of protection for the respective public goal to be achieved. In this regard, a certification procedure for codes and/or non-State bodies has proven efficient. Other resources could be the delegation of members to non-State bodies, financial incentives or publicity. Finally, a clear legal basis and task-sharing are essential for co-regulation. A clear division of tasks is a strong incentive for the industry to participate, whereas an unclear division of competences might cause lack of transparency, increase the risk of "capturing" the regulator and lead to smokescreens.

3.4. Examples for Co-regulation of the Internet in Germany

When it comes to the protection of minors in the media, in Germany non-State bodies have traditionally played an important role: i.e., they have been, and still are, responsible for age-classification. The federal *Jugendschutzgesetz* (Federal Act for the Protection of Minors – JuSchG) distinguishes between different levels of content: content that is harmful to children ("*jugendgefährdend"*) is classified by the *Bundesprüfstelle für jugendgefährdende Medien* (Federal Department for Media Harmful to Young Persons – BPjM). The task of the BPjM is to protect children and adolescents in Germany from any media which might contain harmful or dangerous content. The BPjM monitors, among others, the following media: videos, DVDs, computer games, audio records and CDs, print media and Internet sites. Objects are considered harmful or dangerous to minors if they tend to endanger the process of developing a socially responsible and self-reliant personality. In general, this applies to objects that contain indecent, extremely violent, crime-inducing, anti-Semitic or other racist material. If an object is dangerous to young persons ("*jugendgefährdend*"), the title of this object is placed on the "list of youth-endangering media" ("*Liste jugendgefährdender Medien" - "index*"). Distributors of that medium are then forbidden from selling, renting out or even presenting this object in public or broadcasting it. The content that is harmful to minors must not be shown at places to which children have access and it must not be provided to children. In the field of movies and video games, content that is not harmful to children, but capable of impairing the development of children ("*entwicklungsbeeinträchtigend"*), is rated by the *Oberste Landesjugendbehörde* (State Authorities Responsible for the Protection of Children).

This age classification is carried out in co-operation with non-State bodies: the *Freiwillige Selbstkontrolle Filmwirtschaft* (Voluntary Self-regulation of the Film Industry – FSK) and the

18) Cf. Schulz/Held, Together they are Strong? In: *Regulation, Awareness and Harmful Media Content in the Digital Age, Ulla Carlsson* (Ed.) (2006), pages 49 *et seq.* (pages 51+52), available at: http://unesdoc.unesco.org/images/0014/001469/146955E.pdf

Unterhaltungssoftware Selbstkontrolle (Self-regulation of Entertainment Software – USK). There is a close link between the State and the non-State bodies.[19]

Following this tradition, search engines have established a completely voluntary self-monitoring system. The increasing popularity of search engines and their continuous development leads to an increase in their significance for the protection of youth in the media as well. With this fact in mind, many well-known search engine providers met under the umbrella of the above-mentioned self-regulatory body, FSM, and founded the Self-Monitoring of Search Engine Providers in February 2005.[20] This union is based on a joint Code of Conduct. The aim of this Code is to improve consumer protection, as well as the protection of children and young persons using search engines in Germany. Under this Code, search engine providers agree to explain the search engine's mode of operation to the user. They also agree to structure their search result pages transparently. Furthermore, the signatories to the Code shall endeavour, within their means, to enable technical precautions, which are suitable for promoting the protection of children and young persons from content harmful to them. Moreover, search engine providers integrate the so-called "BPjM Module" provided by the BPjM into their search. The spreading of pornographic and other harmful media content via the Internet is a criminal offence under German penal law. Pornographic content on the Internet is legal only if technical measures prohibit minors from gaining access to the object (Age Verification System or Adult-Check-System – AVS). Content which is harmful to minors ("*jugendgefährdend*") is placed on the above mentioned "index". The "BPjM Module" is a list of foreign Internet services that contain harmful content kept by the BPjM and shared with the FSM. The FSM can use this list for technical precautionary measures (filtering) and the providers, as mentioned above, have agreed to remove and/or not to show respectively any URLs which are placed on the list by the BPjM, insofar as the providers have access to the URL and the commercial expense is reasonable. In case of any violation of this Code of Conduct, sanctions may be imposed on the violator. Search engines have agreed not to show respective results on their .de services.

4. Conclusion: Regulation in the Field of Search Engines

Returning to the list of risks triggered by search engines and applying the criteria that determine in which cases co- and self-regulation might be useful, there are some areas remaining where a closer examination might be crucial.

Hence, the German example shows that, in the field of protection of minors, co- and self-regulation could function. The same goes for the problems of discrimination of content and misleading consumers. Again, the voluntary self-regulation of search engine providers in Germany already addresses some of the issues that concern the transparency of the selection process (not of the algorithm as such). Regarding the risk that search engines might play a role in exploiting protected (audiovisual) works or personal data, there also seems to be at least some leeway for co-regulation.

However, when it comes to public policy goals, like controlling the influence of public opinion making, and the fragmentation of the public sphere, which might be aggravated by search engines, there is no incentive for search engine providers to co-operate. Moreover, the distortion of competition and the transfer of market powers is obviously not a field in which it could be expected that service providers would offer their cooperation voluntarily. In these fields, if any regulation is called for, it would be traditional State regulation that would seem to be necessary.

However, when contemplating different regulatory approaches, one should keep in mind that search engine providers are not completely aware of their role in public communications. The providers of Internet search engines have been focusing on elaborating the search technology. They operate on the

19) Cf. Schulz/Held, Together they are Strong? In: *Regulation, Awareness and Harmful Media Content in the Digital Age*, Ulla Carlsson (Ed.) (2006), pages 49 *et seq*. (57 *et seq.*), available at: http://unesdoc.unesco.org/images/0014/001469/146955E.pdf ; on the issue of self-censorship Held/Schulz, Der Index auf dem Index? Selbstzensur und Zensur bei Suchmaschinen, in Machill/Beiler, *Die Macht der Suchmaschinen – The Power of Search Engines* (2007).

20) Cf. http://www.fsm.de/en/Search_Engines

assumption that they are just creating tools for the users' hands and that they have a limited role in structuring public communication. The concept of information intermediaries with special functions for public communication is relatively new even to the academic debate. Thus, it should be assumed that the same applies to persons working in this field. Anyway, the new developments have to be seen as an advantage, because corresponding professional ethics can still be established and legal regulation might also play a vital role in establishing them.

Furthermore, all service providers are interested in externalising responsibility when it comes to making choices regarding the content. Accordingly – to use the same example again – the list of sites containing harmful content which is provided by the German BPjM has been warmly welcomed by the industry, for the simple reason that it relives the industry players from having to decide which sites they have to block.

Finally, for providers operating internationally there is limited scope for responding to national regulatory initiatives. Self- and co-regulation are able to provide flexible tools for international cooperation, while, at the same time, allowing for global solutions and tying the process in with national regulatory development paths. Therefore, it remains vital that the code allows for localisation (search engines providing for variations of algorithm).

Searching for Audiovisual Content

Participants to the Workshop held on 12 April 2008 in Amsterdam

Christina	**ANGELOPOULOS**	Institute for Information Law (IViR), NL (rapporteur)
Mónica	**ARIÑO**	Ofcom, UK
Pierre-Jean	**BENGHOZI**	CRC, Ecole polytechnique, FR
Machiel	**BOLHUIS**	European Policy Council, Google, EU
Francisco	**CABRERA**	European Audiovisual Observatory, CoE
Ramón	**COMPAÑÓ**	IPTS, ES
Anja	**DOBBELSTEEN**	Institute for Information Law (IViR), NL (logistics)
Nico van	**EIJK**	Institute for Information Law (IViR), NL (moderator)
Urs	**GASSER**	University of St Gallen, CH
Janet	**GRECO**	Broadcastprojects, UK
Natali	**HELBERGER**	Institute for Information Law (IViR), NL
Joris van	**HOBOKEN**	Institute for Information Law (IViR), NL (rapporteur)
Erik	**JOSEFSSON**	Electronic Frontier Foundation, US
Eddan	**KATZ**	Electronic Frontier Foundation, US
Miha	**KRISELJ**	APEK, SI
Tarlach	**MCGONAGLE**	Institute for Information Law (IViR), NL
Milton	**MUELLER**	Syracus University, USA, Technical University, NL
Susanne	**NIKOLTCHEV**	European Audiovisual Observatory, CoE
Michael	**ROTERT**	EuroISPA, BE
Thomas	**ROUKENS**	Telenet, BE
Alexander	**SCHEUER**	Institut for European Media Law (EMR), DE
Wolfgang	**SCHULZ**	Hans-Bredow-Institut, DE

David	**STEVENS**	ICRI, University Leuven, BE
Ralph	**TRAPHOENER**	Empolis, DE
Peggy	**VALCKE**	ICRI, University Leuven, BE
Erik	**VALGAEREN**	Stibbe, BE
Adam	**WATSON-BROWN**	DG Infosoc, European Commission, UE
Michael	**ZIMMER**	University of Wisconsin-Milwaukee, US

IRIS Special Serie

Available in English, French and German

What can you expect from IRIS *Special* in terms of content?

IRIS *Special* is a series of publications from the European Audiovisual Observatory that provides you comprehensive factual information coupled with in-depth analysis. The themes chosen for IRIS *Special* are all topical issues in media law, which we explore for you from a legal perspective. IRIS *Special*'s approach to its content is tri-dimensional, with overlap in some cases, depending on the theme. It offers:

1. a detailed survey of relevant national legislation to facilitate comparison of the legal position in different countries, for example IRIS *Special: Broadcasters' Obligations to Invest in Cinematographic Production* describes the rules applied by 34 European states;
2. identification and analysis of highly relevant issues, covering legal developments and trends as well as suggested solutions: for example IRIS *Special, Audiovisual Media Services without Frontiers – Implementing the Rules* offers a forward-looking analysis that will continue to be relevant long after the adoption of the EC Directive;
3. an outline of the European or international legal context influencing the national legislation, for example IRIS *Special: To Have or Not to Have – Must-carry Rules* explains the European model and compares it with the American approach.

What is the source of the IRIS *Special* expertise?

Every edition of IRIS *Special* is produced by the European Audiovisual Observatory's legal information department in cooperation with its partner organisations and an extensive network of experts in media law.

The themes are either discussed at invitation-only workshops or tackled by selected guest authors. Workshop participants and authors are chosen in order to represent a wide range of professional, academic, national and cultural backgrounds.

IRIS *Special* - unique added value

IRIS *Special* publications explore selected legal themes in a way that makes them accessible not just to lawyers. Every edition combines a high level of practical relevance with academic rigour.

While IRIS *Special* concentrates on issues and interactions within Europe, it takes a broader geographical scope when the theme so requires.

Who needs to read IRIS *Special*?

If you hold a position of responsibility in the media industry, particularly cinema, broadcasting or new media, if you work for a specialised law firm, for a regulatory authority or in media and cultural policy making, or if media studies is your academic discipline, then the IRIS *Special* series is a must for you.

So far published in the IRIS *Special* series:

Editorial Responsibility
ISBN 978-92-871-6476-6 - 50 pages - © 2008 – EUR 75
http://www.obs.coe.int/editresp

Legal Aspects of Video on Demand
ISBN 978-92-871- 6391-2 - 91 pages - © 2008 - EUR 89
http://www.obs.coe.int/vod-legal

The Public Service Broadcasting Culture
ISBN 978-92-871-6188-8 - 184 pages - © 2007 - EUR 62,50
http://www.obs.coe.int/psb

Audiovisual Media Services without Frontiers
ISBN 978-92-871-6115-4 - 76 pages - © 2006 - EUR 58,50
http://www.obs.coe.int/av-media-services

Broadcasters' Obligations to Invest in Cinematographic Production
ISBN 978-92-871-5971-7 - 127 pages - © 2006 - EUR 62,50
http://www.obs.coe.int/tv-invest-film

To Have or Not to Have - Must-carry Rules
ISBN 978-92-871-5897-0 - 52 pages - © 2005 - EUR 58,50
http://www.obs.coe.int/must-carry

Tomorrow's Delivery of Audiovisual Services
ISBN 978-92-871-5707-2 - 56 pages - © 2005 - EUR 35
http://www.obs.coe.int/tv-future

Political Debate and the Role of the Media
ISBN 978-92-871-5675-4 - 125 pages - © 2004 - EUR 44
http://www.obs.coe.int/political-debate

Regulating Access to Digital Television
ISBN 978-92-871-5401-9 - 127 Pages + 28 pages Glossary - © 2004 - EUR 44
http://www.obs.coe.int/digital-tv-access

Co-Regulation of the Media in Europe
ISBN 978-92-871-5141-4 - 140 pages - © 2003 - EUR 27
http://www.obs.coe.int/co-regulation

Jurisdiction over Broadcasters in Europe
ISBN 978-92-871-4854-4 - 64 pages - © 2002 - EUR 27
http://www.obs.coe.int/jurisdiction

Television and Media Concentration
ISBN 978-92-871-4595-6 - 89 pages - © 2001 - EUR 27
http://www.obs.coe.int/media-concentration

European Audiovisual Observatory: tel. +33 (0) 3 88 14 44 00 - fax +33 (0) 3 88 14 44 19 - E-mail obs@obs.coe.int