Techno-economic Challenges of Audiovisual Search Engines

Ramón Compañó (*)
European Commission - Joint Research Centre (JRC)
Institute for Prospective Technological Studies (IPTS)
C/ Inca Garcilaso, s/n
E - 41092 Sevilla – Spain
Email: ramon.compano@ec.europa.eu

1. INTRODUCTION.......................................................................................................2

2. TECHNOLOGICAL CHALLENGES........................................................................3
   2.1. Search Engine Basics ......................................................................................3
   2.2. Challenges .......................................................................................................5
       2.2.1. Gathering Information: Dealing with the Data Explosion ..................5
       2.2.2. Understanding User Queries: Bridging the Semantic Gap.............7
       2.2.3. Returning More Relevant Results: Getting the User into the Loop .................................................................9
       2.2.4. Towards Content-based Retrieval .......................................................11

3. ECONOMIC CONSIDERATIONS.............................................................................13
   3.1. Web Search ...................................................................................................14
   3.2. Business Search Solutions ...........................................................................17
   3.3. Mobile Search ...............................................................................................18

4. SUMMARY ..............................................................................................................19

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 of 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.
1. **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 excerpts. 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 transform rapidly in scale and type. There will be a growing need to investigate new audio-visual 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, for the development of a top-notch AV search portal, or the German Theseus research programme, provide further evidence of the important policy dimension 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 from the publicly-accessible Internet. 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 inside Wikipedia), or search engines that retrieve information which cannot be accessed by web search engines. Similarly, we also exclude from the definition those search engines that retrieve data from closed peer-to-peer networks or applications which

---

3  It is more adequate to refer to search results as "content" or "information", rather than web pages, because a number of search engines retrieve other information than web pages. Examples include search engines for music files, digital books, software code, and other information goods.
4  See for a similar definition, James Grimmelmann, *The Structure of Search Engine Law* (draft), October 13, 2006, p.3, at [http://works.bepress.com/james_grimmelmann/13/](http://works.bepress.com/james_grimmelmann/13/) (last visited: 6th 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 increases 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 to 30 billion pages (as opposed to sites), of which one fourth to one fifth is estimated to be accessible by search engines. See and compare [http://www.pandia.com/sew/383-web-size.html](http://www.pandia.com/sew/383-web-size.html) (last visited: 6th June 2008) and [http://technology.guardian.co.uk/online/story/0,,547140,00.html](http://technology.guardian.co.uk/online/story/0,,547140,00.html) (last visited: 6th June 2008).
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 specifically designed for children, for people search, focused on health-related topics, for retrieving software codes, for radio search, music search, and for science, etc. A regularly updated list of search engines by domain is available at Pandia.

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 aims to find 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.

2. TECHNOLOGICAL CHALLENGES

2.1. 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.

![Search Engine Processes Diagram](image)

**Figure 1 Search Engine Processes**

In essence, a search engine is composed of a number of technical components that fulfill 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. Once the

---

6 e.g. Fragfinn ([www.fragfinn.de](http://www.fragfinn.de))
7 e.g. Spock ([www.spock.com](http://www.spock.com))
8 e.g. Healia ([www.healia.com](http://www.healia.com))
9 e.g. Koders ([www.koders.com](http://www.koders.com)) or Krugle ([www.krugle.org](http://www.krugle.org))
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
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 frequent 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 was.

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 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, and SingingFish for audio content.

While text-based search is efficient for text-only files, this technology has important drawbacks when it has to retrieve information from contents 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 instance, although a video may contain a red mountain, the search engine will not retrieve this video when a user inserts the words "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.

There is thus 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


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: 6th June 2008).
AV content in text format. A solution could be found by, for instance, developing "intelligent" software that automatically tags audio-visual content.\textsuperscript{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.\textsuperscript{16} However, manual tagging is time consuming and thus costly.

\subsection*{2.2. 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 briefly discussed in the following section.

\subsection*{2.2.1. 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 billion gigabytes in 2008.\textsuperscript{17,18} For the sake of comparison, the last figure is equivalent to three million times the information in all books ever written. This figure is expected to reach 988 billion gigabytes by 2010.\textsuperscript{19} 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 users.\textsuperscript{20} This underlines the view that in coming years, users will gradually shift from being mere consumers to becoming

\begin{itemize}
\item \textsuperscript{15} See about this James Lee, Software Learns to Tag Photos, Technology Review, November 9, 2006, at http://www.technologyreview.com/Infotech/17772/.
\item \textsuperscript{17} 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 increases 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 to 30 billion pages (as opposed to sites), of which one fourth to one fifth is estimated to accessible by search engines.
\item \textsuperscript{18} 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).
\item \textsuperscript{19} See Andy McCue, Businesses face data 'explosion', ZDNet, 23 May 2007, at http://news.zdnet.co.uk/tmanagement/0,1000000308,39287196,00.htm (last visited: 6th June 2008), referring to IDC/EMC Study The expanding Digital Universe.
\item \textsuperscript{20} Fauto Rabitti, Sapir presentation at Concertation meeting Vilamoura (Faro, Portugal)16 April 2008
\end{itemize}
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 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 even a larger 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 large resources in terms of computer processing power and storage capacity. Second, they 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 3: 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. Source: Sapir

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 in 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\textsuperscript{22} and some companies, like Faroo,\textsuperscript{23} 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 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. If, and when, P2P architectures will be ready to replace distributed systems is unclear. 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. color AND shape) are not efficient yet.

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, ...) make the indexing task far more complex than text-based formats do. Until new architectures are in place and solutions for non-indexed content is available, it is possible that there will be a widening AV information retrieval 'gap'. For a considerable period, it is likely that search engines will rely 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 ease 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, like 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 ease the finding of relevant content.

2.2.2. 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 to handle for machines as contextual search is not possible. For instance, if we introduce the term 'Jaguar' the search engine is unable to discriminate whether we mean the automobile or the animal, as processing the user's query is linked to the machines 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

\textsuperscript{22} \url{www.sapir.eu} (last visited 6 June 2008)
\textsuperscript{23} \url{www.faroo.com} (last visited 6 June 2008)
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, because their individual features (eyes, mouth, 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.

Figure 2: 'Semi-artificial' pictures

Semi-artificial pictures are more complex from a retrieval point of view. These are images taken under well-defined conditions, like one single object per picture against a homogeneous background. The sets of fish and bracelets in Figure 2 are examples that fulfil these conditions. Searching for this kind of picture is useful for education, particularly for encyclopaedias, and also for commercial purposes (i.e. searching for particular products). 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.

Technological 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, detecting copyright infringement, 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

---

24 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 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 beyond. 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.
25 e.g. www.riya.com (last visited 6 June 2008)
28 Joly, A., Buisson, O., and Frelilot, C. Robust content-based copy detection in large reference database, Int. Conf. on Image and Video Retrieval, 2003
dominant rhythm, melody, or frequency spectrum in audio, are tightly correlated to the search goals of the particular application. For instance, in SongTapper, a user can search for a song by typing the rhythm of the song on the space bar of the keyboard or, in Midomi, by humming the song into a microphone.

High-level features are far more complex, as an 'understanding' of the contents is often required. For instance, humans can, on the spot, identify a two-dimensional projection of a three-dimensional object, i.e. we can also easily recognize a different picture as just another projection (from a different angle) of the very same object. Machines, however, are unable to recognize the 2D representation of 3D objects. They lack the understanding of the image content and therefore 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 was an early content-based retrieval system addressing semantic gap issues in the query interface, indexing, and results. This early search engine allowed a 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.

### 2.2.3. 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.

---

29 [www.songtapper.com](http://www.songtapper.com) (last visited 6 June 2008)

30 [www.midomi.com](http://www.midomi.com) (last visited 6th June 2008)

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 or the information, and the link to other pieces of information is visualized. Clusty\(^{32}\) is an example of a search engine that clusters results, and KartOO\(^{33}\) 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\(^{34}\), or getting user's feedback into search process like the engine ChaCha.\(^{35}\)

Search services would become more efficient and relevant if likely queries could be anticipated, at least to a certain 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'.\(^{36}\)

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. 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.

\(^{32}\) [www.clusty.com](http://www.clusty.com) (last visited 6\(^{th}\) June 2006)

\(^{33}\) [www.kartoo.com](http://www.kartoo.com) (last visited 5\(^{th}\) June 2008)

\(^{34}\) [www.mister-wong.com/](http://www.mister-wong.com/) (last visited 5\(^{th}\) June 2008)

\(^{35}\) [http://search.chacha.com](http://search.chacha.com) (last visited 9\(^{th}\) June 2008)

2.2.4. 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. 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 words '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.

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.

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. 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) shows. NIST carries out periodic benchmark evaluating the quality of spoken document retrieval evaluations. Comparing the performance of indexing for broadcast news during the subsequent text retrieval conferences (TREC from 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 are a challenge of 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. 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

---

37 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 Newswire, February 12, 2007.


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.

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.\(^{42}\) 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.\(^{43}\)

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\(^{44}\) and Riya\(^{45}\) 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\(^{46}\) 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\(^{47}\) and Audiocipping\(^{48}\) allow users to search radio, podcasts, and TV programmes by keyword.\(^{49}\) Blinkx\(^{50}\) and Podzinger\(^{51}\) 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.\(^{52}\) For instance, Pixlogic\(^{53}\) offers the ability to search not only the metadata 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

\(^{50}\) http://www.blinkx.com (last visited: 6th June 2008).
across different domains, such as the previously discussed semantics and the information overflow.

Table 1: Overview of the approaches, applications and challenges for content-based retrieval technologies by type of format. [Source: Adapted from54]

<table>
<thead>
<tr>
<th>Text</th>
<th>Speech</th>
<th>Music</th>
<th>Images</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technological Approach</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Named entity recognition, SVM, PLSIIB ayesian semantic reasoning</td>
<td>• Speech recognition: spoken document retrieval, subword indexing</td>
<td>• Music segmentation: spectral flatness, generic algorithms, music retrieval and recommendation</td>
<td>• Low-level image processing (histograms, shapes, textures, MPEG7-visual, SIFT), image similarity measurements</td>
<td>• Shot detection, keyframe generation</td>
</tr>
<tr>
<td>• Caption augmentation</td>
<td>• Speech segmentation: speaker clustering and recognition speech-to-video transcoding</td>
<td></td>
<td></td>
<td>• Object tracking based on motion based features, closed captions recognition, etc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Relevance feedback</td>
<td></td>
<td>• Object detection and recognition</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Video annotation and summarization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Video event detection</td>
</tr>
</tbody>
</table>

| Applications | | | | |
| • Classification of news and documents in companies | • Indexing of broadcast news/archives | • Indexing of music collections | • Content-based retrieval in image collections | • Indexing broadcast material, media observation, indexing of videocast material, recommendation engines, video fingerprinting, logo detection, security, etc. |
| • Email filtering | • Podcast/videocast search (Pötzing, Blinkx) audio archives (Parliament data, historical archives) | • Query by humming | • Object recognition, face recognition (security, photo collections), automatic annotation of image collections with keywords and textual descriptions | • 3D video |
| • Text-based search engines, semantic analysis of multimedia (automatic) annotations | | • Audio-music identification recommendation engines | | |

| Challenges | | | | |
| • Semantics, ontologies design (e.g. enrichment) | • Variability of content (e.g. background noise) domain and language dependency, scalability of subwords approaches | • Genre classification | • 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 |
| | | • Polyphonic instrument recognition, Affective analysis | | |

3. **Economic Considerations**

The data explosion in the digital realm has rendered search a critical functionality. Data abundance is at the origin of the 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, content securities, etc.

Although these retrieval tools do share similar technological challenges, the nature of their application 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.

3.1. 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 first 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 only in that month (August 2007). 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 a year ago, 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.

The search engine market consolidation becomes evident when observing the evolution of hits over longer time periods. Figure 3 and Figure 4 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 followers.

55 see www.nielsen-netratings.com
56 comScore Press releases, available at www.comscore.com
These data highlight the fact that the search engine market is highly concentrated and also that search engines have become a part of our lives. The average German, for instance, uses Google more than forty times a month and three quarters of Internet users get to Internet offers through search engines. Although the audience of the search engine providers may vary from one country to another, the user experience is similar for most western countries.

Consultancy firms metering the market share, such as Nielson/NetRatings, Compete, Hitwise or comScore, measure the search behaviour of internet users by installing real-time meters the computes to 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 partially

57  http://webhits.de
59  www.secrets2moteurs.fr
60  comScore German data June 2007
61  Internetverbreitung in Deutschland: Potenzial vorerst ausgeschöpft? Birgit van Eimeren, Heinz Gerhards and Beaeste Frees, Media Perspektiven, Vol 8, page 350 - 370
explain why comScore's traffic data for Germany and France differs from the hit counts 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. Why Google is more dominant in Europe than in the USA may have multiple reasons, 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 and Rambler. Yandex claims to have superior technology as it masters better the declinations and conjugations of the Russian language than other search engines. Other Slavic search engines, like the Czech Morfeo or the Polish NetSprint, 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 into 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 content will also further increase. Crawling is a resource consuming task, but not huge discriminating factor amongst search engine providers having an 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 the income and expenditures. 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-

---

62 www.yandex.ru
63 http://morfeo.centrum.cz/
64 www.netsprint.pl/serwis/
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) is amongst 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 requirements in terms of performance in retrieving relevant content for the user. The need to deliver increasingly good relevance partially 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. Therefore only a few general search engines are operating world-wide with success. In view of the fact that even more processing power for audiovisual content is needed than for text, the entry barrier is unlikely to diminish for the new incumbents of future audiovisual search engines. A consolidation around a few general web search engines is likely. An alternative is specialisation into a niche. This is the case of the thematic search engines. They provide relevant retrieval in specific domains and the barrier entry 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.

### 3.2. 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' data bases, 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 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 enterprises 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 $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 with regard to their core business. Contrary to 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 amongst the most prominent of these, including 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-visual 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.65

### 3.3. 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 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

---

65 [http://www.01net.com/editorial/368946/microsoft-s-achete-la-place-de-numero-un-de-la-recherche-en-entreprise/](http://www.01net.com/editorial/368946/microsoft-s-achete-la-place-de-numero-un-de-la-recherche-en-entreprise/)
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 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 of mobile broadband connection. 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. 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 worldwide, of whom roughly 40% have 2.5G 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 become potential customers for search engine providers. In fact, eMarketer considers that the mobile search advertisement market 221 m$ in 2008 and will increase to 2,361 m$ by 2011.

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.

4. **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,

---

66 [www.gsmworld.com](http://www.gsmworld.com)
67 [www.eMarketer.com](http://www.eMarketer.com)
retrieving relevant audio-visual material will benefit from new content-based search (audio, video, images) but it will also largely rely 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 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 which 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 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 blur. 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.