A vision[1]: New search technologies change the world of TV

Robert Ortgies

Institut für Rundfunktechnik, Munich, Germany

E-mail: ortgies@irt.de

Abstract: **Search engines are omnipresent on the Internet and are expected to be embedded into television receivers in the future. The embedded search engines together with metadata for audio-visual material will determine how hard or how easily content can be found. Having multiple providers for video/TV search and metadata [2], will promote competition and the freedom of information. This paper presents that strategic vision of separating the provision of audiovisual material, associated metadata and search functions.**

Keywords: TV, television, audiovisual search, search engines, embedded, vision, freedom of information, comfort, video, network, metadata, provider, video portal, personalisation, context, receiver, recorder, European Commission, funding

1 INTRODUCTION

Search engines are omnipresent on the Internet. Far away from today's mass market and the consumer's attention there are new fields of applications for search engines. One new thing is that search engines are beginning to be able to deliver search results from audio visual content and not only from text information. The second new thing is that search engines are not only focused on the Internet but are going to be embedded in local devices such as TV recorders. This is sometimes invisible for the user especially when combined with the third new thing: to trigger or initiate a search automatically. That means it is not necessary that the user sits in front of a keyboard to initiate a search. Search can be context-initiated. For example the search engine proposes a TV show because it knows what kind of TV entertainment the users like to consume when they sit on a sofa on Monday evening alone or in a crowd and expressing a certain mood detected by sensors. The fourth new thing is, how metadata can be generated and provided. Metadata is information about information: e.g. the date and location of the production, keywords (or non-textual information) of the story, genre, actor, intellectual property information, etc. that is needed or useful for video search. Currently, there is no technical solution for search in large video data bases by analysing the video on the fly. In other words, today search for video/TV content is only possible on the basis of associated metadata.

2 COMBINATION LEADS TO COMFORT

If applied to a TV receiver/recorder, the combination of these new things mentioned above has the potential to change the world of television: A video search engine embedded in a TV set, optionally initiated or triggered by context or personalisation.

An example for this is: Around the clock a search engine decides which of the TV signals are worth to be recorded for its owners based on the interests the engine learned during first switch on since buying. When the TV screen is switched on by the user, a list of the most interesting TV recordings for the user is offered. Optionally, the listed order of the TV recordings can be modified in relation to context of use. To give an example of context: who is sitting at what time in what mood in front of the TV could change the order of the list by broadcast time, genre, actor, most viewed etc.). And without any need for user to press any other than the switch-on button of the TV set, the context could automatically trigger a player to represent the video material. The context for this trigger could be that the user is not able or not in the mood to choose a TV recording from a list.

Like Internet browsers that are equipped with a link to a search engine provider, TV sets could come up as well with a search engine and could, at switch-on, present interesting hits found during last hours, days or other periods of time. In contrast to an Internet browser-based search engine, there will be no need for the TV user to enter an explicit query with a keyboard to obtain a recommendation. In more general words: It is desirable to offer best possible comfort while benefiting from the new audio-visual search technologies.

2.1 The Mass Market Today

Today, for TV signals, non-context aware implementations are already available in the mass market, for example by the company Tivo. That company applies proprietary search engines and provides proprietary metadata which both have a big influence on what TV material is recorded and consumed.

Tivo has not foreseen a possibility for the user to select an alternative or additional metadata provider other than the one to which the Tivo TV recorder is permanently linked. So the users cannot influence the search results by, for example,

[1] The term "vision" means here: The state that is desired to reach in the future. Parts of this vision and many other visions were drafted in the Intermediate Vision Document by the EU funded Coordination Action "CHORUS" in 2008. [1]

[2] i.e. search for video/TV content on the bases of associated metadata

Corresponding author: Robert Ortgies, Institut für Rundfunktechnik, Floriansmühlstr. 60, 80939 München, Germany, +49 89 32399310, ortgies@irt.de

selecting alternative metadata sources, so that the users may find the material they are really interested in.

In the US market, Tivo is dominant with their service. In some countries Tivo is regulated by local authorities or governments. And in many countries Tivo don't even offer their TV recorder and their service.

Today in the Internet market there are a number of video portals like Youtube. Especially Youtube is offering a huge (if not the biggest) collection of video material and a search engine to search for material. Similar to Tivo, Youtube determines with their search engine and their metadata how easy or hard it is to find desired material. Youtube is a dominant provider in numerous countries including many countries of the EU.

2.2 Advancing freedom of information

To ensure access to information for all citizens, it appears useful to have alternatives to such dominant "all-in-one-providers" like Youtube which are providing all decisive elements from one hand: the material, the search engine and the necessary metadata for search like illustrated on the right side in Figure 1. The search engine, together with metadata, determines how easy or difficult it is to find specific content. Therefore it is desirable that the users can add or exclude as many metadata providers they like for the search engines of their choice independent of the material provider. This is illustrated on the left side in Figure 1. Allowing access to multiple providers for metadata, search engines and video material in an interoperable way, will promote competition and the freedom of information. For this an openly standardized and commonly accepted technology is needed in order to provide the required interoperability between material providers, potential search providers and potential metadata providers. Today there are already a few metadata providers but they are acting within proprietary business solutions.

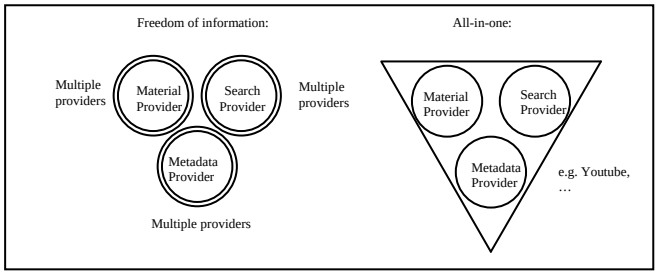

Figure 1: In the future multiple entities will provide material, search and metadata in an interoperable way to do business and ensuring freedom of information. Instead of leaving this market to some dominating „all-in-one" providers.

2.3 Dependencies and needed interoperability

The quality of metadata and the functionalities of the search engines will together determine how easy or hard it is to find desired audio-visual material. Providers of video material are very interested that their videos can be searched and found easily. So the material providers depend more and more on the performance of search engines. On the other hand, search engines are depending on the availability of high-quality metadata whilst the generators of metadata depend on best technical video quality and on ancillary data generated during production like date and location of the shot, for example.

This closes the loop of dependencies, each one of those three providers is depending on each other provider (illustrated in Figure 2).

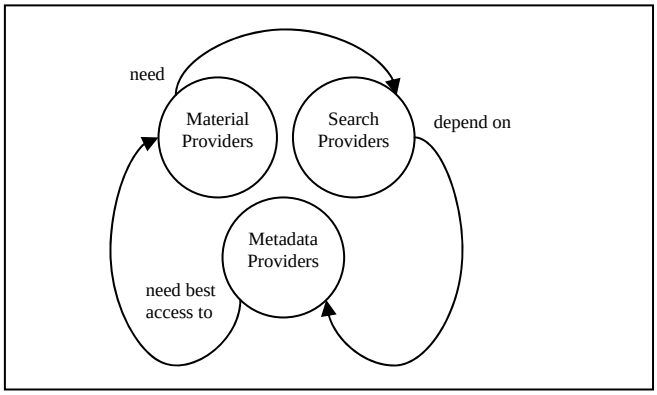

Figure 2: Material Providers need Search Providers. Search Providers depend on Metadata Providers. Metadata Providers need access to best material quality.

2.4 Further success factors

Access to the Internet should only be an option for "TV search" and not a precondition. The lower half of Figure 3 shows this IP based option. In Figure 3 the upper half illustrates TV search using only unidirectional networks like DVB-x. A TV set may be equipped with an Internet access such as xDSL; such "hybrid" TV sets (especially HDTV sets) are expected on the market by late 2009. Whilst an Internet connection were to provide a wider choice as all videos available in the world-wide web would potentially be included in the search, the minimum requirement for TV search should be considered fulfilled if set-top box or TV set is connected to the unidirectional broadcasting network only. That also helps to ensure privacy and may help to keep the Internet load low.

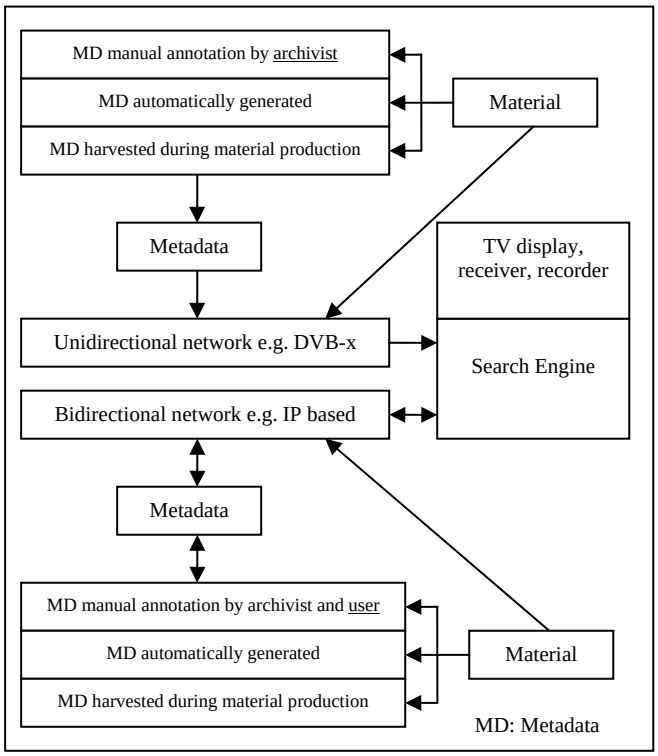

Figure 3: Metadata from different sources for TV search. Having bidirectional networks like Internet should only be an option (lower half of the Figure) but not a precondition for TV search.

But the option to use bidirectional networks like the Internet opens the door for collecting user annotation of material, provided the user agrees to. Bidirectional networks are illustrated by bidirectional arrows in the lower half of Figure 3, where metadata annotation involves the user in addition to the professional archivist (third rectangle from the bottom). User annotation is one source to generate metadata to advance TV search. Another very important source of metadata is the production chain. In most cases of content signal generation, important information like camera location, time and date of production and the script is today preserved manually, often however, the information gets lost during professional production, postproduction and storing. This is also true for user-generated video. A well-known example is a user who tries to produce a Video-DVD with material from a camcorder. During the necessary transcoding (e.g. from DV to MPEG-2 compression scheme) most metadata produced by the camcorder gets lost (date, time, location, face recognition etc.). And lost metadata cannot be used for search, of course. There are already ongoing activities within the industry to address parts of this technological key issue. One recent example in the photographic sector was documented in a press release entitled “Metadata Working Group Introduces First Specification for Interoperability and Preservation of Metadata in Digital Photography” [3]. Compared to the video sector, the photographic sector is in a better position for preserving metadata along production and postproduction due to the dominant market share of JPEG compression and the EXIF format for storing metadata contained in only one single dominant file format. In the video sector, new methods are needed to preserve the essence (i.e. the audio-visual material itself) and related metadata because there are at least twenty different compression, metadata and file formats having a reasonable market share [2]. Additionally, preserving metadata to advance search in video is more complex due to its nature: Time continuation, additional tracks for audio, time codes etc.

The third way to generate metadata needed for video search is the so called automatic generation of metadata. An example is object and face recognition. The state of the art regarding face recognition is that the available techniques can distinguish and recognise only a very limited number of faces. They are far away from being able to distinguish every face on earth. But the integration at the beginning of the production chain can help to lower the recognition error rates. Face recognition integrated into a camera can lower the recognition error rate if the user volunteers to help the engine by correcting mismatches, and because of the fact that cameras are often used to capture only a very limited number of different faces.

This example shows that even for automatically generated metadata preservation is of importance. This is true for the whole chain from production to consumption in order to advance video and TV search.

Object (e.g. face) recognition is a precondition for automatic semantic annotation, which, in turn, is one of the biggest challenges for the long term research in the area of audio-visual processing.

3 THE BUSINESS MODELL

Technology should guarantee competition and enable business for several metadata providers and search engine providers in the future. If not provided by the TV set manufacturer itself, the search engine could be a downloadable software via the TV signal network (not only via Internet or SD card) to be executed on the TV receiver/recorder. To enable the software, an individual code could be bought and entered by the consumer. The same mechanism could be provided for metadata. Metadata are essential for video search, and they represent an asset of intellectual properties by themselves. This allows doing business with metadata for several metadata providers, and those metadata can be used by future competing search engines. This independence from specific search engines preserves the competitive conditions for unlimited number of material providers.

4 CONCLUSIONS

The challenge is to establish receiver/recorder embedded TV search within the next years into the mass market by ensuring comfort, freedom of information, competition, business and privacy. If time to market constraints are not met with respect to specification/standardization then proprietary and vertical solutions are likely to become dominant — similar to the

situation we have today on the Internet with dominant search engines and video portals or with personal TV recorders in the US market.

TV search could bring back users from their PCs to their TV sets, and this might be interesting for the TV receiver/recorder manufactures as well.

The same vision is even true for Internet video portals and could change the situation of today from one dominant "all-in-one" provider into several material, metadata and search providers doing a shared and competitive business.

To achieve the vision for video/TV search, the first step would be an agreed specification within the concerned broadcast industries for interoperable “search ready” TV receivers/recorders. That specification should give each player room for new developments, but should ensure the needed interoperability to achieve “the combination” of the new “things” (ref. Section 1) that will lead to the best possible comfort for the end user.

In the opinion of the author “the bill of materials”, in comparision to conventional recorder designs, will only be slightly increased for such “search ready” TV recorders. The very first step (before such a specification agreement is approved) would be to reserve the necessary amount of flash memory space for a search engine in a TV recorder and to prepare the embedded software (OS etc.) for this. This very first step would keep open the door for new business and freedom of information of tomorrow.

5 ACKNOLEDGEMENT

This article contains only a fraction from the visions created by the CHORUS Think Tank. All visions and other deliverables are available on www.ist-chorus.org.

CHORUS ran from November 2006 to September 2009. Its purpose was to coordinate the EU funded R&D projects in the area of audio visual search and to provide visions for research in this domain.

Sincere thanks are given to the 53 stakeholders and scientists of the seven CHORUS Think Tank meetings that had taken place in several European countries. They succeeded in disrupting conventional thinking. The Think Tank encompassed 32 organisations including major companies of search engines, consumer and mobile electronics, broadcasters and a press agency, archives, telecom providers, capital goods industries, institutes and universities. The author likes to thank them for their effort and especially the European Commission for having initiated and funded the coordination action CHORUS.

References

[1] R. Ortgies, C. Dosch et. al., (2008) CHORUS Deliverable 3.3: Vision Document, intermediate version, http://www.ist-chorus.org/D3_3_CHORUS_Vision_Document.zip

[2] Rolf Bardeli, Nozha Boujemaa et. al., (2008) CHORUS Deliverable 2.2: Identification of multi-disciplinary key issues, http://www.ist-chorus.org/documents/D2_2_Chorus_Gap_Analysis_final.pdf

[3] Pressrelease of the Metadata Working Group: (COLOGNE, Germany — Sept. 24, 2008) „Metadata Working Group Introduces First Specification for Interoperability and Preservation of Metadata in Digital Photography. Adobe, Apple, Canon, Microsoft, Nokia and Sony form Metadata Working Group“, http://www.microsoft.com/presspass/press/2008/sep08/09-24PhotokinaPR.mspx