IFLA WLIC 2013
SINGAPORE

Submitted on: April 30, 2013

TIB´s Portal for audiovisual media: New ways of indexing and retrieval

Janna Neumann
DOI-Service
German National Library of Science and Technology
Hannover, Germany
Janna.Neumann@tib.uni-hannover.de

Margret Plank
Competence Center for nontextual Material
German National Library of Science and Technology
Hannover, Germany
Margret.Plank@tib.uni-hannover.de

Abstract:

The German National Library of Science and Technology (TIB) is developing a web-based platform for audiovisual media. The forthcoming audiovisual portal optimises access to scientific videos such as computer animations, lecture and conference recordings. TIB's AV- Portal offers new methods for searching within videos enabled by automated video analysis with scene, speech, text and image recognition. Search results are connected to new knowledge by linking the data semantically. This paper aims at describing the TIB´s portal for audiovisual media and the multimedia retrieval technologies as well as the added value for libraries and their users.

Keywords: AV-Portal, Metadata, Multimedia Retrieval, Automated Video Analysis, Search Tools

1 BACKGROUND

The German National Library of Science and Technology (TIB)[1] ranks as one of the largest specialized libraries worldwide. It is jointly financed by the federal government and the federal states ("Länder") of Germany. TIB is a member of the Leibniz-Association, an umbrella organisation for 86 institutions conducting research and providing scientific infrastructure. The TIB's task is to comprehensively acquire and archive literature from around the world pertaining to all areas of engineering as well as architecture, chemistry, information technology, mathematics and physics. The TIB´s portal GetInfo provides access to more than 150 million data sets from specialised databases, publishers and library catalogues. Further, the TIB actively participates in a large number of projects with a focus on visual analytics.

The Competence Centre for non-textual Materials at TIB is engaged in achieving fundamental improvements to conditions pertaining to the access and use of non-textual material such as audiovisual media, 3D objects and research data. This material is to be systematically collected at the Competence Centre, and preserved as cultural heritage. In this context the TIB together with the Hasso-Plattner Institut for software system technology GmbH (HPI)[2], is developing a web-based platform for audiovisual media. The future AV-Portal will optimise access to and the use of scientific videos from the fields of engineering and science. The TIB is converting known, multimedia analysis methods such as scene, speech, text and image recognition in order to enhance bibliographic metadata. The results are connected to new knowledge by linking the data semantically. The aim is to make it as easy for users to locate and use the growing stock of non-textual material as it is for them now to procure textual media. In 2011, a partially functioning prototype of the AV-Portal was developed; in 2012-2013, the further development and the beta operation of a system followed and, for 2014, the full operation of the portal is planned.

2 RELATED WORK

In the face of a rapidly increasing number of non-textual objects and the necessity of indexing the contents even of individual film sequences, an intellectual, "manual" indexing is unthinkable. This drives the demand for efficient automated metadata extraction. Additionally new tools and technologies are necessary in order to improve access to non-textual objects. During the last years there have been made different steps in this direction.

The Project ‘Mediaglobe’[3] was funded by the German Federal Ministry of Economics as part of the THESEUS research programme[4] from 2007 to 2012. The project’s objective was to develop solutions which allow media archives to not only optimally digitise, comprehensively index and efficiently administer their growing inventory of audiovisual documents on German history, but also to make them accessible online. The project partner, the Hasso-Plattner Institut for software system technology GmbH incorporated the development of automated and semantic media analysis and metadata generation, as well as semantic search technologies.

Yovisto[5] is a video portal for lecture recordings. Research and development foci are on automated video analysis and on the integration of so-called user-generated Web 2.0 services like tagging, evaluation and annotation. Currently, the project is being continued by the Semantic Web research group at the Hasso-Plattner Institut for software system technology GmbH (HPI).

ScienceCinema[6] is a video portal created by the Office of Scientific and Technical Information (OSTI)[7] one of the operation offices of the U.S. Department of Energy[8] as well as the European Organization for Nuclear Research (CERN).[9] Using innovative audio indexing and speech recognition technology from Microsoft Research, ScienceCinema allows users to search for specific words and phrases spoken within video files, whereas the search term is highlighted in the audio snippets.

Voxalead News[10] is using multimedia search technologies from Exalead S.A.[11] It also searches the spoken content of radio and television programmes, thus enabling innovative navigation within the video.

In this paper we focus on the generation of enhanced metadata by means of automated video analysis, the user benefits in the search process by named entity recognition as well as representing it in a web-based portal.

3 THE VIDEO ANALYSIS METHODS

Use case for the AV-Portal

A researcher produces a video and uploads it via the TIB´s web form into the media asset management system, where the video is transcoded. The bibliographic metadata like author, title, description etc. needs to be provided by the author. Then the video is processed for receiving enhanced, comparatively fine granular metadata which will be also indexed next to the bibliographic metadata. The enlarged index makes it possible to improve search results. The automated video analysis contains the following processes:

The process of scene recognition provides a visual table of content of the video. In this case, the video will be scanned and decomposed into time-based fragments. The cuts will be set automatically at scene edges. The scenes are furthermore divided into shots and subshots. Each time-based scene is represented by one keyframe in the visual table of content. For the automated scene recognition the used algorithm has to be trained with enough and diverse video material to obtain satisfying and correct results. Like implemented in the above mentioned portals Yovisto or Mediaglobe, the scenes will be visualized in form of a visual table of content. By clicking onto a keyframe the user can jump directly into the selected scene. The visualisation in the web-based portal allows the user to easily browse across the video.

The process of automated image recognition allows the detection of categories. The categories are classified as visual concept in six subject areas. These are architecture, chemistry, information technology, mathematics, physics, and engineering, the main fields TIB is focusing on. For every subject area a list of specific concepts was defined. Some cross-subject concepts have also been integrated for each subject area like e.g. lecture, conference, interview and screencast. The training of the concepts was realized by using manually annotated video material. For this annotation experts of TIB had to find enough images for each concept. For example in the subject area of engineering the specific concept "shipping" was defined. The annotated keyframes contain different shipping images as the main part of the scene content, which could easily be analyzed by object detection. However the principle problem that appeared in this part of the process was the definition of the specified subject concepts. The concept definition for the applied sciences (e.g. engineering and architecture) was found much easier than the definition of the other sciences (e.g. chemistry and physics). Most difficulties were given in the subject field of mathematics, because of very abstract video scenes, that could not be defined into a concept. Therefore in this case mainly the cross-subject concepts were used. As the video is analysed by automated image recognition the detected concepts are indexed and included as enhanced metadata. Within the faceted search of the portal the user can easily narrow his search with the given facets and find relevant videos by concept.

The process of speech recognition automatically extracts the spoken text within the video. The audio analysis is divided into two different processes. The structural analysis distinguishes between the spoken word and other sounds (e.g. music).[12] The second part of the automated speech recognition (ASR) is a speech to text analysis[13], where the auditive structure is matched to (the spoken) words. The quality of the results is, however, dependent upon the quality of the speaker; dialects, background noises and voice overlaps can be problematic. Further the challenge is to achieve good matching, which requires a preceded domain training of the spoken text. The training material was selected and provided by TIB experts, so that subject specific vocabulary could be added to each subject domain. Therefore up to 170 videos were required including the corresponding transcript in German as well as English. Successful audio analysis allows the user to navigate across the spoken text of the video.

The process of the text recognition extracts textual information within the video images. The so called intelligent character recognition (ICR) contains of text pre- and postprocessing as well as standard optical character recognion (OCR)[14]. The text preprocessing analysis recognizes and extracts the written text. The standard OCR transforms the extracted text block to textual information. The postprocessing analysis corrects the textual information by using lexical analysis. The extracted textual information e.g. from slides is also included as enhanced metadata within the faceted search.

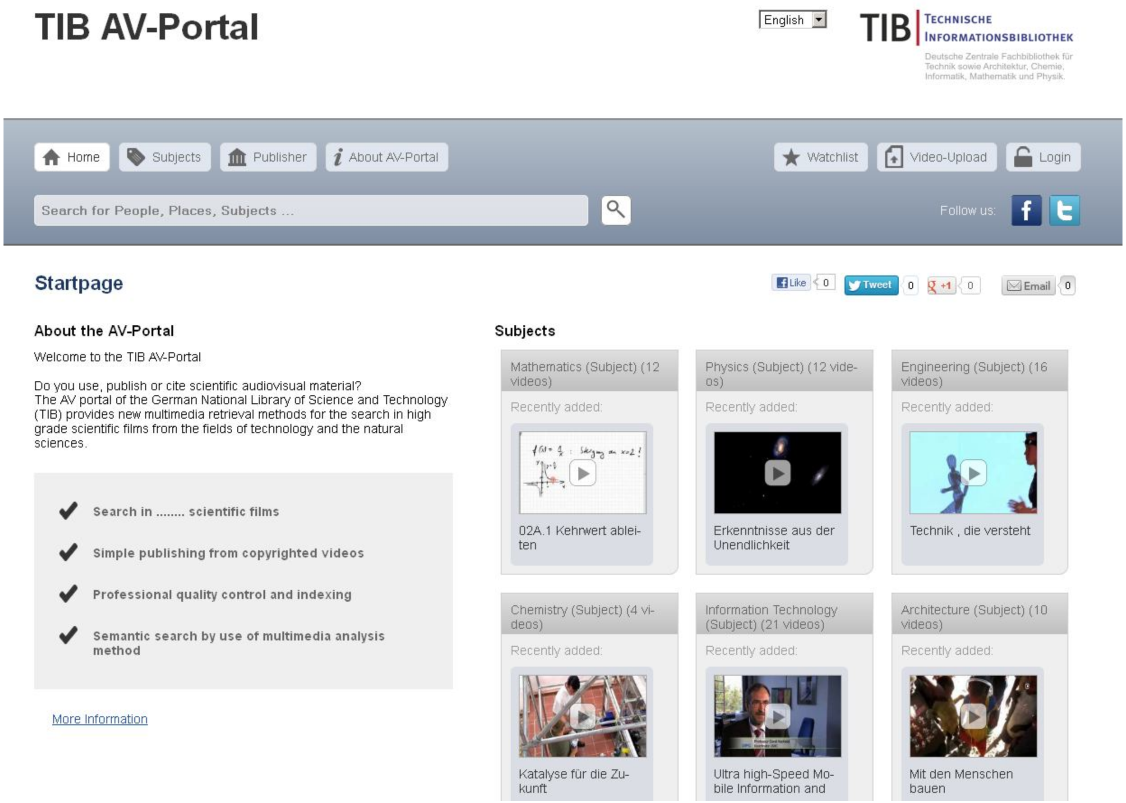
As the automated video analysis is finished the extracted textual information is linked to the underlying knowledge base of the AV-Portal by using the process of Named Entity Recognition (NER). Named Entity Recognition means ‘locating and classifying atomic elements […] into predefined categories such as names, persons, organizations, locations, expressions of time, quantities, monetary values, etc.[15] It allows exploring the content in much greater depth. The process needs the development of a knowledge base in German as well as English. The extracted German text will be linked to the German Authority File, Gemeinsame Normdatei (GND), which functions as the underlying German knowledge base within the portal. The GND is a standardized vocabulary used for cataloguing in German libraries. It provides connections to synonymous and super-/subordinated terms respectively. The English knowledge base for the AV-Portal has not yet been selected. The different options like e.g. Library of Congress Subject Headings (LCSH)[16] have all been dismissed because of missing linkage to the German knowledge base. The linkage however is essential for the enhanced search process.

On the basis of the video analysis technologies and the connection to the knowledge base the TIB is able to develop with HPI and provide a portal for audiovisual media, which offers the user new ways of searching and browsing within scientific films from the fields of science and technology.

4 THE AV-PORTAL

In order to ensure the future accessibility and usability of knowledge via the AV-Portal, the development has been accompanied by user-centred methods. As a process model, this offers several methods for the development of information systems which can be meaningfully used in a library context to develop user-friendly approaches. DIN EN ISO 9241-210 (DIN 2010)[17] serves as a basis for this user-centred approach. There, the process of designing utilisable systems based on the phase analysis of the usage context, definition of the requirements, conception and design/prototyping and evaluation is described. The following measures were used:

- Expert interviews with representatives from scientific institutes, film institutes, libraries and universities
- Context analysis: Research into publicly available AV-Portals, automated metadata extraction, content-based search methods and visualisation
- Development of a low fidelity prototype of the AV-Portal on the basis of the results
- Focus groups with users from the target groups (engineering and science)
- Development of a high fidelity prototype on the basis of the results
- Usability testing with 12 users
- Optimization
- Usability testing / eye tracking with 30 users


TIB AV-Portal
English
TIB Technische Informationsbibliothek
Deutsche Zentrale Fachbibliothek für Technik sowie Architektur, Chemie, Informatik, Mathematik und Physik.
Home
Subjects
Publisher
About AV-Portal
Watchlist
Video-Upload
Login
Search for People, Places, Subjects ...
Follow us:
Like 0
Tweet 0
+1 0
Email 0
More Information

Startpage

About the AV-Portal

Welcome to the TIB AV-Portal

Do you use, publish or cite scientific audiovisual material?
The AV portal of the German National Library of Science and Technology (TIB) provides new multimedia retrieval methods for the search in high grade scientific films from the fields of technology and the natural sciences.

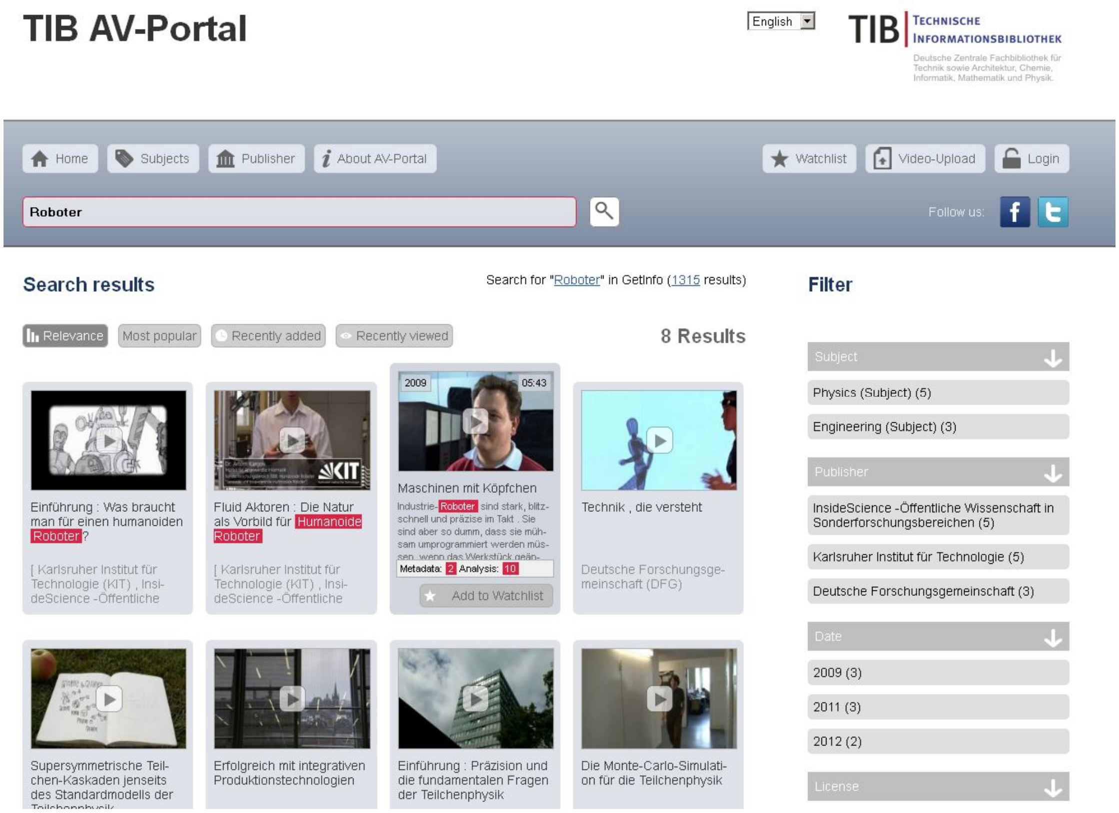
- ✓ Search in .... scientific films
- ✓ Simple publishing from copyrighted videos
- ✓ Professional quality control and indexing
- ✓ Semantic search by use of multimedia analysis method

Subjects

Mathematics (Subject) (12 videos) Recently added: 02A.1 Kehrwert ableiten	Physics (Subject) (12 videos) Recently added: Erkenntnisse aus der Unendlichkeit	Engineering (Subject) (16 videos) Recently added: Technik , die versteht
Chemistry (Subject) (4 videos) Recently added: Katalyse für die Zukunft	Information Technology (Subject) (21 videos) Recently added: Ultra high-Speed Mobile Information and	Architecture (Subject) (10 videos) Recently added: Mit den Menschen bauen

Figure 1: Start page

On the left side of the start page (see Fig. 1) a short description is displayed. Here the video retrieval technologies, which have been used are explained. At first a search scenario has been supported whereby the user knows precisely what he/she is searching for giving the opportunity to use the boolean AND to connect the searchterms. Additionally the option of browsing to see what content and functionalities the portal offers has been supported. Therefore on the right side of the start page the six TIB subjects have been displayed together with additional subjects at the bottom. However, users who want to access basic information about his/her subject, and do not know a great deal about the relevant themes, subject areas and authors can use filters like "subject "and "producers". Also a watchlist, the video upload section and the login section have been placed in the top section. The Login is important for users who want to watch or download licensing restricted videos. There will be a German and an English version of the portal, the English version has not yet been completed.



Roboter

Search results

Search for "Roboter" in GetInfo (1315 results)

8 Results

Einführung : Was braucht man für einen humanoiden Roboter?
[Karlsruher Institut für Technologie (KIT) , InsideScience -Öffentliche

Fluid Aktoren : Die Natur als Vorbild für Humanoide Roboter
[Karlsruher Institut für Technologie (KIT) , InsideScience -Öffentliche

2009 05:43
Maschinen mit Köpfchen
Industrie-Roboter sind stark, blitzschnell und präzise im Takt . Sie sind aber so dumm, dass sie mühsam umprogrammiert werden müssen, wenn das Werkstück geän-
Metadata: 2 Analysis: 10

Technik , die versteht
Deutsche Forschungsgemeinschaft (DFG)

Supersymmetrische Teilchen-Kaskaden jenseits des Standardmodells der Teilchenphysik

Erfolgreich mit integrativen Produktionstechnologien

Einführung : Präzision und die fundamentalen Fragen der Teilchenphysik

Die Monte-Carlo-Simulation für die Teilchenphysik

Subject
Physics (Subject) (5)
Engineering (Subject) (3)

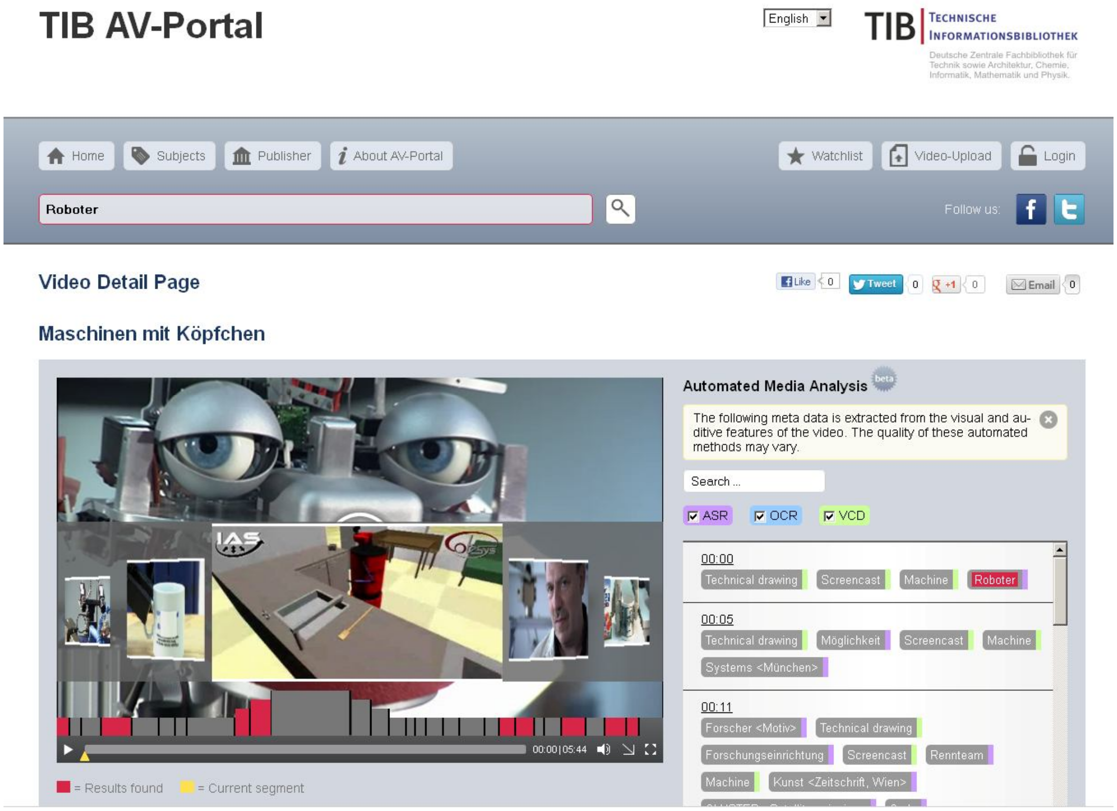
Publisher
InsideScience -Öffentliche Wissenschaft in Sonderforschungsbereichen (5)
Karlsruher Institut für Technologie (5)
Deutsche Forschungsgemeinschaft (3)

Date
2009 (3)
2011 (3)
2012 (2)

License

Figure 2: Search results

The relevant hits have been displayed in a list together with a thumbnail and a snippet (Fig. 2). It can be checked weather the hits derive from the metadata or from the media analysis like speech, text or image recognition. To narrow the list of search results a faceted search can be used. The facets *Subject*, *Publisher*, and *License* derive from the static metadata, which the author has delivered together with the video. The entries included in the other facets have been extracted by automated video analysis. The facet *Category* derives from the visual analysis and includes the genre of the video like e.g. interview, conference, inside or outside shot. The facets *Person*, *Organisations*, *Places*, and *Other* concepts derive from either speech recognition or text recognition.



Roboter

Video Detail Page

Maschinen mit Köpfchen

The following meta data is extracted from the visual and auditive features of the video. The quality of these automated methods may vary.

☑ ASR ☑ OCR ☑ VCD

00:00
Technical drawing
Screencast
Machine
Roboter

00:05
Technical drawing
Möglichkeit
Screencast
Machine
Systems <München>

00:11
Forscher <Motiv>
Technical drawing
Forschungseinrichtung
Screencast
Rennteam
Machine
Kunst <Zeitschrift, Wien>

Figure 3: Player and aggregated metadata from automated media analysis

Picking an interesting video from the list of search results, detailed information is displayed (Fig. 3). A HTML 5 player has been implemented as well as a flash player as a backup. By moving over the segments the visual index is displayed. This provides a quick overview of the video content and facilitates access to particular segments. The red segments contain the search term whereas the yellow segment is the current segment. The segments have been indexed based on time code. The time code is displayed over on the right side followed up by the extracted metadata from the automated media analysis. The data gained from speech, text and visual analysis has been aggregated and colour coded. Again the current segment is highlighted. By clicking on one of the keywords the user can jump to the desired segment of the video. The user can also search within the extracted metadata using the search entry.

5 CONCLUSION

The supply, use and significance of non-textual media is continually increasing but only a tiny proportion of these materials can be searched and explored right now. To face these new challenges libraries have to open up their library portals to non-textual information, develop new tools for indexing, searching, browsing and displaying the data as well as enrich the data with semantic information.

The TIB converts state of the art multimedia analysis methods for searching within videos enabled by automated analysis with scene, speech, text and image recognition in order to generate additionally metadata. The search results are connected to new knowledge by linking the data semantically.

- Scene recognition: a visual table of contents provides a quick overview of the video content, facilitating access to particular segments.
- Image recognition: based on visual features in the video (such as colour distribution), the system automatically recognises whether it is a lecture, an interview or an experiment.
- Speech and text recognition: both the spoken word and lettering in the video (for example, in logos or slides) are automatically recognised. The search term is highlighted, enabling navigation within the video.
- Semantic search: by adding semantic information to the data gained from video analysis, explorative navigation of the stock can be performed, enabling connections between audiovisual media

The objective is to expand the hitherto text-based search in bibliographical metadata to a media and cross-data search. In doing so, digital full texts with numerical data and facts, other research information, audiovisual media, visualisations, etc. will be integrated into a single user interface. The search space is widened, due to the connection between audiovisual media and TIB's portal GetInfo that contains information such as digital full texts, numerical data and facts as well as research data. Audiovisual media will be clearly referenced by the allocation of a Digital Object Identifier (DOI). The search tools provided by the AV-Portal offer innovative search scenarios and new ways of tapping into knowledge.

6 REFERENCE

[1] www.tib.uni-hannover.de

[2] http://www.hpi.uni-potsdam.de

[3] www.projekt-mediaglobe.de

[4] www.theseus-programm.de

[5] www.yovisto.com

[6] http://www.osti.gov/sciencecinema/

[7] http://www.osti.gov/home/

[8] http://energy.gov/

[9] http://home.web.cern.ch/

[10] http://voxaleadnews.labs.exalead.com/

[11] http://www.3ds.com/products/exalead/overview/

[12] Schneider D, Schon J, Eickeler S (2008) Towards large scale vocabulary independent spoken term detection: advances in the Fraunhofer IAIS audiomining system. In: Köhler J, Larson M, Jong de F, Kraaij W, Ordelman R (eds) Proc of the ACM SIGIR workshop "searching spontaneous conversational speech". Singapore

[13] Nandzik, J. et al., (2013) Multimed Tools Appl 63:287–329; DOI 10.1007/s11042-011-0971-2

[14] Liu, M. et al., (2012) EURASIP Journal on Advances in Signal Processing, 2012:109 http://asp.eurasipjournals.com/content/2012/1/109

[15] C.J.Rijsbergen, Information Retrieval (1979)

[16] http://id.loc.gov/authorities/subjects.html

[17] http://www.beuth.de/de/norm/din-en-iso-9241-210/135399380\\