

## **El portal de la TIB (Technische informationsbibliothek) para medios audiovisuales: nuevas formas de indización y recuperación**

*Spanish translation of the original paper: TIB's Portal for audiovisual media: New ways of indexing and retrieval*

*Translated by: Traducción: Lourdes Alonso Viana, Cayetano Hernández Muñiz, Biblioteca Nacional de España, Madrid, Spain*

### **Janna Neumann**

DOI-Service

German National Library of Science and Technology

Hannover, Alemania

Janna.Neumann@tib.uni-hannover.de

### **Margret Plank**

Competence Center for nontextual Material

German National Library of Science and Technology

Hannover, Alemania

[Margret.Plank@tib.uni-hannover.de](mailto:Margret.Plank@tib.uni-hannover.de)



This is a Spanish translation of "TIB's Portal for audiovisual media: New ways of indexing and retrieval" Copyright © 2013 by **Janna Neumann** and **Margret Plank**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/3.0/>

---

### **Resumen:**

*La Biblioteca Nacional Alemana de Ciencia y Tecnología (TIB) está desarrollando una plataforma Web para medios audiovisuales. El futuro portal audiovisual mejora el acceso a los vídeos científicos, tales como animaciones por ordenador y grabaciones de conferencias y congresos. El Portal Audiovisual de la TIB (AV-Portal) ofrece nuevos métodos para hacer búsquedas en los vídeos gracias a un análisis automático del vídeo con reconocimiento de escena, habla, texto e imagen. Los resultados de la búsqueda se conectan con nueva información a través del enlace semántico de datos. Este artículo se propone describir el portal de la TIB para medios audiovisuales y las tecnologías de recuperación multimedia así como el valor añadido para bibliotecas y sus usuarios.*

**Palabras clave:** AV-Portal, Metadatos, Recuperación multimedia, Análisis automático de vídeo, Herramientas de búsqueda

---

## **1. ANTECEDENTES**

La Biblioteca Nacional Alemana de Ciencia y Tecnología (TIB) (1) destaca como una de las mayores bibliotecas especializadas del mundo. Está financiada conjuntamente por el gobierno federal y los estados federales (“Länder”) de Alemania. La TIB es miembro de la Asociación Leibniz, una organización que aglutina a 86 instituciones que dirigen investigaciones y proporcionan infraestructura científica. La tarea de la TIB es obtener y archivar de manera exhaustiva la literatura mundial relativa a todas las áreas de la ingeniería, así como de arquitectura, química, tecnología de la información, matemáticas y física. El portal GetInfo de la TIB proporciona acceso a más de 150 millones de conjuntos de datos procedentes de bases de datos especializadas, editores y catálogos de bibliotecas. Además, la TIB participa activamente en un gran número de proyectos enfocados al análisis visual.

El Centro de Competencia para Materiales no textuales en la TIB se ocupa de lograr mejoras fundamentales de las condiciones relativas al acceso y uso de material no textual, como los medios audiovisuales, los objetos 3D y los datos de investigación. El Centro de Competencia recoge sistemáticamente este material y lo conserva como parte del patrimonio cultural. En este contexto la TIB junto con el Hasso-Plattner Institut para la tecnología de software del sistema (HPI) (2) está desarrollando una plataforma Web para medios audiovisuales. El futuro AV-Portal mejorará el acceso y el uso de los vídeos científicos procedentes de los campos de la ingeniería y la ciencia. La TIB está transformando los métodos de análisis multimedia conocidos, como el reconocimiento de escena, habla, texto e imagen, con el fin de mejorar los datos bibliográficos. Los resultados se conectan a nueva información enlazando los datos semánticamente. El objetivo es que para los usuarios sea tan fácil la localización y uso de la creciente cantidad de material no textual, como lo es la búsqueda de medios textuales. En 2011 se desarrolló un prototipo del AV-Portal que funcionó parcialmente; en 2012-2013 siguió un mayor desarrollo y un funcionamiento en beta del sistema y para 2014 está previsto el funcionamiento pleno del portal.

## **2. TRABAJO RELACIONADO**

Ante el número rápidamente creciente de objetos no textuales y la necesidad de indizar los contenidos incluso de secuencias sueltas de películas, una indización de carácter intelectual “manual” es impensable. Por ello se demanda una eficaz extracción automática de metadatos. Además son necesarias nuevas herramientas y tecnologías para mejorar el acceso a los objetos no textuales. Durante los últimos años se han dado distintos pasos en esta dirección.

El proyecto “Mediaglobe” (3) fue financiado por el Ministerio de Economía Federal Alemán como parte del programa de investigación THESEUS (4) desde 2007 hasta 2012. El objetivo del proyecto era desarrollar soluciones que permitieran que los archivos relativos a los medios de comunicación no sólo fueran digitalizados de forma óptima, y su creciente inventario de documentos audiovisuales sobre la historia alemana indizados exhaustivamente y administrados eficientemente, sino también hacerlos accesibles en línea. El socio del proyecto, el Hasso-Plattner Institut para la tecnología de software del sistema, incorporó el desarrollo del análisis automatizado y semántico de los medios de comunicación y la generación de metadatos, así como tecnologías de búsqueda semántica.

Yovisto (5) es un portal de vídeo para grabaciones de conferencias. La atención de la investigación y el desarrollo está puesta en el análisis de vídeo automatizado y en la

integración de los llamados servicios Web 2.0 generados por los usuarios, como el etiquetado, la valoración y los comentarios. Actualmente, el grupo de investigación en web semántica del Hasso-Plattner Institut para la tecnología de software del sistema (HPI) está continuando el proyecto.

ScienceCinema (6) es un portal de vídeo creado por la Oficina de Información Científica y Técnica (OSTI) (7), una de las oficinas del Departamento de Energía de los EE.UU. (8), y la Organización Europea para la Investigación Nuclear (CERN) (9). Usando una innovadora tecnología de Microsoft Research de indización de audio y reconocimiento del habla, ScienceCinema permite a los usuarios buscar palabras y frases específicas habladas en los ficheros de vídeo, considerando que el término de búsqueda esté destacado en los fragmentos de audio.

Voxlead News (10) usa las tecnologías de búsqueda multimedia de Exalead S.A. (11). También busca el contenido hablado de programas de radio y televisión, permitiendo así una navegación innovadora en el vídeo.

En este trabajo nos centramos en la generación de metadatos mejorados mediante el análisis de vídeo automatizado, el usuario se beneficia en el proceso de búsqueda por reconocimiento de entidad nominal, así como al representarlo en un portal Web.

### **3. LOS MÉTODOS DE ANÁLISIS DE VÍDEO**

#### *El caso del AV-Portal*

Un investigador realiza un vídeo y lo sube a través del formulario web de la TIB en el sistema de gestión de medios, donde el vídeo se transcodifica. Los metadatos bibliográficos como autor, título, descripción, etc., los tiene que proporcionar el autor. Luego, el vídeo se procesa para recibir metadatos granulares mejorados, comparativamente buenos, que también se indizarán junto con los metadatos bibliográficos. El índice aumentado posibilita mejorar los resultados de búsqueda. El análisis de vídeo automatizado contiene los siguientes procesos:

El proceso de reconocimiento de escena proporciona una tabla visual del contenido del vídeo. En este caso, el vídeo será escaneado y descompuesto en fragmentos temporales. Los cortes se establecerán automáticamente en los límites de las escenas. Las escenas se dividen además en tomas y subtomas. Cada escena temporal se representa por un fotograma clave en la tabla visual de contenido. Para el reconocimiento automatizado de escena hay que entrenar el algoritmo usado con material de vídeo suficiente y diverso para obtener resultados satisfactorios y correctos. Al igual que en los susodichos portales Yovisto o Medialogbe, las escenas se visualizarán en forma de una tabla visual de contenido. Pinchando en un fotograma clave el usuario puede saltar directamente a la escena seleccionada. La visualización en el portal Web permite al usuario navegar fácilmente a través del vídeo.

El proceso de reconocimiento automatizado de imagen permite la detección de categorías. Las categorías se clasifican como concepto visual en seis áreas temáticas. Éstas son arquitectura, química, tecnología de la información, matemáticas, física e ingeniería, los principales campos de los que se ocupa la TIB. Para cada área temática se ha definido una lista de conceptos específicos. También se han integrado algunos conceptos de referencia cruzada en cada área temática, como por ejemplo, conferencia, congreso, entrevista y vídeos explicativos. El entrenamiento de los conceptos se realizó usando material de vídeo anotado

manualmente. Para esta anotación los expertos de la TIB tuvieron que encontrar imágenes suficientes para cada concepto. Por ejemplo en el área temática de la ingeniería se definió el concepto específico “transporte”. Los fotogramas clave anotados contienen diferentes imágenes de transportes como parte principal del contenido de la escena, que podría analizarse fácilmente por detección de objeto. Sin embargo, el principal problema que apareció en esta parte del proceso fue la definición de los conceptos temáticos específicos. La definición de conceptos para las ciencias aplicadas (por ejemplo, ingeniería y arquitectura) resultó mucho más fácil que para las otras ciencias (por ejemplo, química y física). La mayor parte de las dificultades se dieron en el campo temático de las matemáticas, a causa de escenas de vídeo muy abstractas, que no se podían definir en un concepto. Por tanto, en este caso se usaron principalmente conceptos de referencia cruzada. Como el vídeo es analizado por reconocimiento automatizado de imagen los conceptos detectados se indizan e incluyen como metadatos mejorados. Dentro de la búsqueda facetada del portal el usuario puede restringir fácilmente su búsqueda con las facetas dadas y encontrar vídeos relevantes por concepto.

El proceso de reconocimiento de habla extrae automáticamente el texto hablado del vídeo. El análisis de audio se divide en dos procesos distintos. El análisis estructural distingue entre la palabra hablada y otros sonidos (por ejemplo, música) (12). La segunda parte del reconocimiento de habla automatizado (ASR) es un análisis de habla a texto, donde la estructura auditiva coincide con las palabras (habladas). Sin embargo, la calidad de los resultados depende de la calidad del hablante; los dialectos, los ruidos de fondo y los solapamientos de voz pueden ser problemáticos. Además el desafío es lograr una buena coincidencia, lo cual requiere un previo entrenamiento en el campo del texto hablado. Expertos del TIB seleccionaron y proporcionaron el material de entrenamiento, de modo que pudiera añadirse a cada campo temático el vocabulario específico de la materia. Por tanto, se necesitaron hasta 170 vídeos incluyendo la correspondiente transcripción en alemán y en inglés. Un análisis de audio satisfactorio permite al usuario navegar por el texto hablado del vídeo.

El proceso de reconocimiento de texto extrae información textual de las imágenes de vídeo. El llamado reconocimiento inteligente de caracteres (ICR) contiene procesamiento anterior y posterior de texto así como reconocimiento óptico de caracteres normalizado (OCR) (14). El análisis de procesamiento anterior de texto reconoce y extrae el texto escrito. El OCR normalizado transforma el bloque de texto extraído en información textual. El análisis de procesamiento posterior corrige la información textual usando análisis léxico. La información textual extraída, por ejemplo a partir de diapositivas, también se incluye como metadatos mejorados en la búsqueda facetada.

Cuando termina el análisis automático de vídeo la información textual extraída se enlaza a la base de conocimiento subyacente del AV-Portal usando el proceso de Reconocimiento de Nombres de Entidades (NER). Reconocimiento de Nombres de Entidades significa “localizar y clasificar elementos atómicos [...] en categorías predefinidas tales como nombres, personas, organizaciones, lugares, expresiones de tiempo, cantidades, valores monetarios, etc. (15). Permite explorar el contenido con una profundidad mucho mayor. El proceso necesita el desarrollo de una base de conocimiento tanto en alemán como en inglés. El texto alemán extraído se enlazará con el Fichero de Autoridad Alemán, Gemeinsame Normdatei (GND), que funciona como la base de conocimiento alemana subyacente en el portal. El GND es un vocabulario normalizado usado para catalogar en las bibliotecas alemanas. Proporciona conexiones a términos sinónimos y genéricos/específicos respectivamente. La base de

conocimiento inglesa para el AV-Portal todavía no se ha seleccionado. Las diferentes opciones, como por ejemplo los Encabezamientos de Materia de la Biblioteca del Congreso (LCSH) (16), se han descartado todas a causa de la ausencia de enlace a la base de conocimiento alemana. Sin embargo el enlace es esencial para el proceso de búsqueda mejorado.

Basándose en las tecnologías de análisis de vídeo y la conexión a la base de conocimiento, la TIB puede desarrollar con el HPI y proporcionar un portal de medios audiovisuales, lo que ofrece al usuario nuevas formas de búsqueda y navegación en películas científicas de los campos de la ciencia y la tecnología.

#### **4. EL AV-PORTAL**

Para asegurar la accesibilidad y usabilidad futuras del conocimiento a través del AV-Portal, el desarrollo se ha acompañado de métodos centrados en el usuario. Como un modelo de proceso, éste ofrece varios métodos para el desarrollo de sistemas de información que pueden usarse significativamente en un contexto bibliotecario para desarrollar accesos amigables hacia el usuario. La DIN EN ISO 9241-210 (DIN 2010) (17) sirve de base para este acceso centrado en el usuario. Allí se describe el proceso de diseño de sistemas utilizables basados en el análisis de fase del contexto de uso, la definición de los requisitos, la concepción y diseño/prototipo y la evaluación. Se han usado las siguientes medidas:

- Entrevistas profesionales con representantes de institutos científicos, institutos cinematográficos, bibliotecas y universidades
- Análisis de contexto: investigación en portales audiovisuales de acceso público, extracción automatizada de metadatos, métodos de búsqueda basados en el contenido y visualización
- Desarrollo de un prototipo de baja fidelidad del AV-Portal en base a los resultados
- Grupos de discusión con usuarios de los grupos objetivo (ingeniería y ciencia)
- Desarrollo de un prototipo de alta fidelidad en base a los resultados
- Prueba de usabilidad con 12 usuarios
- Optimización
- Prueba de usabilidad / seguimiento ocular con 30 usuarios

**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: Facebook Twitter

**Startpage**

Like 0 Tweet 0 +1 0 Email 0

**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

[More Information](#)

**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

Figura 1: Página de inicio

En el lado izquierdo de la página de inicio (ver Fig. 1) se muestra una breve descripción. Aquí se explican las tecnologías de recuperación de vídeo que se han usado. En primer lugar, se ha mantenido un escenario de búsqueda por el cual el usuario sabe precisamente lo que está buscando, dando la oportunidad de usar el operador booleano “Y” para conectar los términos de búsqueda. Además, se ha mantenido la opción de navegación para ver qué contenido y funcionalidades ofrece el portal. Luego, en el lado derecho de la página de inicio se visualizan las seis materias de la TIB junto con materias adicionales al final. Sin embargo, los usuarios que quieran acceder a información básica sobre su materia y no sepan mucho sobre los temas relevantes, áreas temáticas y autores pueden usar filtros como “materia” y “productores”. Asimismo se han colocado en la parte superior una lista de control de cambios, la sección de carga de vídeo y el acceso de usuario. El acceso es importante para los usuarios que quieran ver o descargar vídeos de permiso restringido. Habrá una versión alemana y otra inglesa del portal, la versión inglesa aún no se ha completado.

Home Subjects Publisher About AV-Portal Watchlist Video-Upload Login

Roboter

Follow us: f t

Search results

Search for "Roboter" in Getinfo (1315 results)

Filter

Relevance Most popular Recently added Recently viewed

8 Results

The screenshot shows a search results page for 'Roboter'. It features a grid of 8 video results, each with a thumbnail, a title, and a brief description. The titles include 'Einführung: Was braucht man für einen humanoiden Roboter?', 'Fluid Aktoren: Die Natur als Vorbild für Humanoide Roboter', 'Maschinen mit Köpfchen', 'Technik, die versteht', 'Supersymmetrische Teilchen-Kaskaden jenseits des Standardmodells der Teilchenphysik', 'Erfolgreich mit integrativen Produktionstechnologien', 'Einführung: Präzision und die fundamentalen Fragen der Teilchenphysik', and 'Die Monte-Carlo-Simulation für die Teilchenphysik'. The filter sidebar on the right includes categories like Subject (Physics, Engineering), Publisher (Karlsruher Institut für Technologie, Deutsche Forschungsgemeinschaft), Date (2009, 2011, 2012), and License.

Figura 2: Resultados de búsqueda

Los resultados relevantes se visualizan en una lista junto con una miniatura y un fragmento (Fig. 2). Puede comprobarse si los resultados derivan de los metadatos o del análisis de los medios de comunicación como el reconocimiento de habla, texto o imagen. Para limitar la lista de resultados puede usarse una búsqueda facetada. Las facetas *Materia*, *Editor* y *Permiso* provienen de los metadatos estáticos que el autor ha facilitado junto con el vídeo. Las entradas incluidas en las otras facetas se han extraído del análisis automatizado del vídeo. La faceta *Categoría* se obtiene del análisis visual e incluye el género del vídeo, como por ejemplo, entrevista, congreso, toma interior o exterior. Las facetas *Persona*, *Organizaciones*, *Lugares*, y *Otros* conceptos se toman bien del reconocimiento de habla, bien del reconocimiento de texto.

Home Subjects Publisher About AV-Portal
 
 Watchlist Video-Upload Login

Roboter  
Follow us: [f](#) [t](#)

## Video Detail Page

Like 0 Tweet 0 +1 0 Email 0

## Maschinen mit Köpfchen

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

Search ...

ASR  OCR  VCD

00:00  
 Technical drawing Screencast Machine **Roboter**

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

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

■ = Results found ■ = Current segment

Figura 3: Reproductor y metadatos agregados a partir del análisis automatizado de los medios

Al escoger un vídeo interesante de la lista de resultados de búsqueda, se visualiza información detallada (Fig. 3). Se han implantado un reproductor HTML 5 con flash player como apoyo. Al moverse sobre los segmentos se muestra el índice visual. Esto proporciona una vista rápida del contenido del vídeo y facilita el acceso a segmentos concretos. Los segmentos rojos contienen el término de búsqueda, mientras que el segmento amarillo es el segmento actual. Los segmentos se han indizado según el código temporal. El código de tiempo se visualiza en el lado derecho seguido de los metadatos extraídos del análisis automático de los medios. Se agregan y ordenan según color los datos obtenidos del análisis del habla, del texto y visual. De nuevo se destaca el segmento actual. Pinchando en una de las palabras clave el usuario puede saltar al segmento deseado del vídeo. El usuario también puede buscar en los metadatos extraídos usando la entrada de búsqueda.

## 5. CONCLUSIÓN

El suministro, uso e importancia de los medios no textuales están aumentando continuamente pero sólo una pequeña proporción de estos materiales puede buscarse y estudiarse en la actualidad. Para afrontar estos nuevos desafíos las bibliotecas han de abrir sus portales a información no textual, desarrollar nuevas herramientas para indizar, buscar, navegar y visualizar los datos, así como enriquecer los datos con información semántica.



La TIB transforma los métodos de análisis multimedia de última generación para buscar en vídeos, a través del análisis automático con reconocimiento de escena, habla, texto e imagen generando, además, metadatos. Los resultados de la búsqueda se conectan con nueva información enlazando los datos semánticamente.

- Reconocimiento de escena: una tabla visual de contenido proporciona una vista rápida del contenido del vídeo, facilitando el acceso a segmentos concretos.
- Reconocimiento de imagen: basándose en características visuales del vídeo (como la distribución del color), el sistema reconoce automáticamente si es una conferencia, una entrevista o un experimento
- Reconocimiento de habla y texto: tanto la palabra hablada como la escrita en el vídeo (por ejemplo, en logotipos o diapositivas) se reconocen automáticamente. Se destaca el término de búsqueda, permitiendo la navegación en el vídeo
- Búsqueda semántica: añadiendo información semántica a los datos obtenidos del análisis del vídeo, puede realizarse una navegación de exploración en el conjunto, posibilitando conexiones entre los medios audiovisuales

El objetivo es expandir la búsqueda en metadatos bibliográficos, hasta ahora basada en el texto, a una búsqueda de medios y datos cruzados. Al hacer esto, los textos digitales completos con datos numéricos y hechos, otra información de investigación, los medios audiovisuales, las visualizaciones, etc. se integrarán en una interfaz de usuario única. El espacio de búsqueda se amplía, debido a la conexión entre medios audiovisuales y el portal GetInfo de la TIB, que contiene información tal como textos digitales completos, datos numéricos y hechos, así como datos de investigación. Los medios audiovisuales estarán claramente referenciados por la asignación de un Identificador de Objeto Digital (DOI). Las herramientas de búsqueda proporcionadas por el AV-Portal ofrecen escenarios de búsqueda innovadores y nuevas formas de hacer uso de la información.

## 6. REFERENCIAS

- [1] [www.tib.uni-hannover.de](http://www.tib.uni-hannover.de)
- [2] <http://www.hpi.uni-potsdam.de>
- [3] [www.projekt-mediaglobe.de](http://www.projekt-mediaglobe.de)
- [4] [www.theseus-programm.de](http://www.theseus-programm.de)
- [5] [www.yovisto.com](http://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>