

TIB 视听媒介门户：标引与检索的新方式

Chinese Translation of the Original Paper: “TIB’s Portal for audiovisual media: New ways of indexing and retrieval”.

Translated by: 赵丹丹 (Dandan Zhao), National Library of China, Beijing, China.

Janna Neumann

DOI-Service

German National Library of Science and Technology

Hannover, Germany

Janna.Neumann@tib.uni-hannover.de

Margret Plank

CompetenceCenter for nontextual Material

German National Library of Science and Technology

Hannover, Germany

Margret.Plank@tib.uni-hannover.de



This is a Chinese translation of “TIB’s Portal for audiovisual media: New ways of indexing and retrieval” Copyright © 2013 by Janna Neumann & 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/>

摘要：

德国国家科技图书馆（TIB）正在开发视听媒介的网络平台，即将投入使用的视听门户优化了科技视频的检索，如计算机动画、讲座和会议录像。通过情景、语音、文本和图像识别的自动视频分析，TIB 视听门户提供了视频内容检索的新方法。检索结果通过数据的语义关联与新知识相联。本文旨在介绍 TIB 的视听媒介门户、多媒体检索技术以及对图书馆和图书馆用户的附加价值。

关键词：视听门户、元数据、多媒体检索、自动视频分析、检索工具

1 背景

德国国家科技图书馆（TIB）¹位列全球最大的专业图书馆之一，由联邦政府和联邦州共同资助。TIB 是莱布尼茨学会的成员（Leibniz-Association），该协会会员共有 86 家机构，从事研究和提供科学基础设施。TIB 的任务是全面收录和保存世界各地与工程学、建筑学、化学、信

息技术、数学和物理学相关的文献。TIB 的门户 GetInfo 提供 150,000,000 多个数据集的检索，这些数据集来自专业数据库、出版商和图书馆目录。此外，TIB 还积极参与了许多以可视化分析为中心的项目。

TIB 非文本文献能力中心（Competence Centre for non-textual Materials）致力于改善视听媒介、3D 实物和研究数据等非文本文献的检索和使用。文献在能力中心进行系统采集，并保存为文化遗产。在此基础上，TIB 与哈索·普拉特纳软件研究所（HPI, Hasso-Plattner Institut for software system technology GmbH）²正在共同开发视听媒介的网络平台。未来的视听媒介（AV）门户将优化工程和科学领域的科技视频的检索和使用。TIB 正在将已知的多媒体分析方法如情景、语音、文本和图像识别用于加强书目元数据，并通过数据的语义关联将结果与新知识相连。目的是使用户可以像现在获取文本文献一样方便查找和使用数量日益增多的非文本文献。2011 年研发出了可以实现部分功能的 AV 门户原型；2012-2013 年进一步开发，对系统进行测试运行；计划 2014 年实现正式运行。

2 相关工作

面对非文本实物数量的急骤增多和影片目次标引的必要性的增强，人工手动标引是无法想像的。这就推动了对于有效的自动元数据提取的需求。此外也需要新的工具和技术以提高对非文本实物的检索。过去的几年中在这一方面取得了一些进展。

2007-2012 年进行的“Mediaglobe”项目³是由德国联邦经济部资助的，作为 THESEUS 研究项目⁴的一部分。这一项目的目标是开发解决方案，使媒介档案馆不但实现最优化地数字化、全面地标引以及有效地管理日益增长的德国历史方面的视听文件库，而且可以支持在线获取。项目合作者 HPI 加入了自动化语义媒体分析、元数据生成及语义搜索技术的开发。

Yovisto⁵是讲座录像的视频门户。主要研究和开发自动视频分析和用户生成的 Web 2.0 服务的整合，如标签、评价和注释。目前这一项目由 HPI 的语义网研究小组继续开展。

ScienceCinema⁶视频门户是由科技信息办公室（OSTI⁷，美国能源部⁸业务处之一）和欧洲核子研究组织（CERN）⁹共同创建的。应用了微软研究院的创新性的音频标引和语音识别技术，ScienceCinema 允许用户检索视频文件中口述的特定单词和短语，检索词在检索出的视频片段中突出显示。

Voxlead News¹⁰应用了来自 Exalead S.A.¹¹的多媒体检索技术。支持对广播电视节目的语音内容的检索，实现了视频内的创新性导航。

本文主要探讨的是通过自动视频分析加强型元数据的生成，以及通过命名实体识别与其在网络门户中的体现用户在检索过程中的获益。

3 视频分析方法

AV 门户使用实例

研究者制作视频后，通过 TIB 的网络上传到媒体资产管理系统，在这一系统对视频进行代码转换。作者、题名、描述等书目元数据需要作者提供。之后视频经过处理得到相对详尽的加强型元数据，与书目元数据一同编入索引。扩展后的索引可以改善搜索结果。自动视频分析包含以下过程：

情景识别过程为视频生成可视化目次。在这一过程中，视频经过扫描，分解为基于时间的片段。剪切处自动设置为情景分界处。情景又再细分为镜头和分镜头。每一段基于时间的情景表

现为可视化目次中的一个关键影格。在自动情景识别中使用的算法需要经由大量的各种类型的视频文献的测试，以获得准确的令人满意的结果。在上文提到的 Yovisto 和 Mediaglobe 门户中，情景实现了可视化，形式为可视化目次。点击关键影格，用户便可直接跳到选定的情景。网络门户中的可视化使得用户可以轻松浏览视频。

自动图像识别过程可以对类型进行检测，类型被归纳为六大学科范围的视觉概念。六大学科为建筑学、化学、信息技术、数学、物理学和工程学，是 TIB 关注的主要领域。每个学科范围均定义了一些特定概念。跨学科概念在各学科范围内进行了整合，如讲座、会议、采访和截屏视频。概念的测试是通过手动标注的视频文献实现的。为了进行标注，TIB 的专家们必须为每个概念找到足够多的图像。例如在工程学学科范围中定义了“shipping”（航运）这一特定概念。标注的关键影格中包含不同航运的图像作为情景内容的主要部分，通过对象检测可以很容易加以分析。然而在这一过程中出现的主要问题是特定学科概念的定义。应用科学的概念定义（如工程学和建筑学）要比其他科学的定义（如化学和物理学）更为容易。最困难的是数学领域概念的定义，因为都是非常抽象的视频情景，无法定义为一个概念。因此这种情况下主要使用的是跨学科概念。视频通过自动图像识别分析，检测到的概念被编入索引，成为加强型元数据。在门户的分面搜索中，用户通过给出的分面可以很容易缩小检索范围，并通过概念找到相关视频。

在语音识别过程中自动提取视频中的语音文本。音频分析分为两个过程。一是结构分析，将口头语言和其他声音（如音乐）相区分¹²。二是自动语音识别（ASR），是一种从语音到文本的分析¹³，将听觉内容匹配为（口述的）文字。识别结果的质量取决于说话人的说话质量，方言、背景噪音和声音重叠都会产生问题。这一过程的挑战在于实现优质的匹配，这就需要进行语音文本的先期测试。测试文献是由 TIB 专家挑选和提供的，这样可以将学科特定词汇添加到各学科领域中。因此需要 170 多部视频及其对应的德文和英文文本。成功的音频分析可以实现用户在视频的语音文本中进行导航。

文本识别过程是提取视频图像中的文本信息。智能字符识别（ICR，intelligent character recognition）包含文本预处理、后处理和标准的光学字符识别（OCR，optical character recognition）¹⁴。文本预处理分析是识别和提取书面文本。标准的 OCR 将提取的文本块转换为文本信息。后处理分析利用词法分析修正文本信息。提取的文本信息如幻灯片中的信息也在分面搜索中列为加强型元数据。

自动视频分析完成后，提取的文本信息通过命名实体识别（NER，Named Entity Recognition）关联到 AV 门户的基础知识库。命名实体识别是指“将基本元素定位和归纳为[.....]预先定义的类别，如名称、个人、机构、地点以及时间、数量和货币值的表达等”¹⁵。可以实现内容的更深入查找。这一过程需要开发德文和英文的知识库。提取的德文文本将关联到门户的德文知识库，即德国规范文档（GND，Gemeinsame Normdatei）。GND 是规范词表，用于德国图书馆的编目中，可以分别关联到同义术语和上一级/下一级术语。AV 门户的英文知识库还没有选定。已经排除了美国国会主题标引（LCSH）¹⁶等可选项，因为它们无法与德国知识库进行关联。然而为了增强检索性能这一关联是不可或缺的。

在视频分析技术和与知识库进行关联的基础上，TIB 与 HPI 合作开发，建立了视听媒介门户，为用户提供检索和浏览科技领域的影片的新方法。

4 AV 门户

为了保障 AV 门户知识的未来可获取性和可使用性，门户的开发采用了以用户为中心的方法。作为过程模型，提供了几种信息系统开发的方法，可以用于图书馆中开发用户友好的方式。DIN EN ISO 9241-210（DIN 2010）¹⁷是这一以用户为中心的方法的基础。基于对使用环境的相

位分析，对要求、概念和设计/原型设计的定义及评估，描述了可利用系统的设计过程。以下为采取的方法：

- 专家与来自科学研究院、电影协会、图书馆和大学的代表进行会谈
- 环境分析：研究已面向公众开放的 AV 门户、自动元数据提取、基于内容的检索方法和可视化
- 在研究结果基础上开发 AV 门户的低保真原型
- 与来自对象群体（工程学和科学）的用户进行焦点小组讨论
- 在研究结果基础上开发 AV 门户的高保真原型
- 面向 12 位用户进行可用性测试
- 最优化
- 面向 30 位用户进行可用性测试/视线追踪

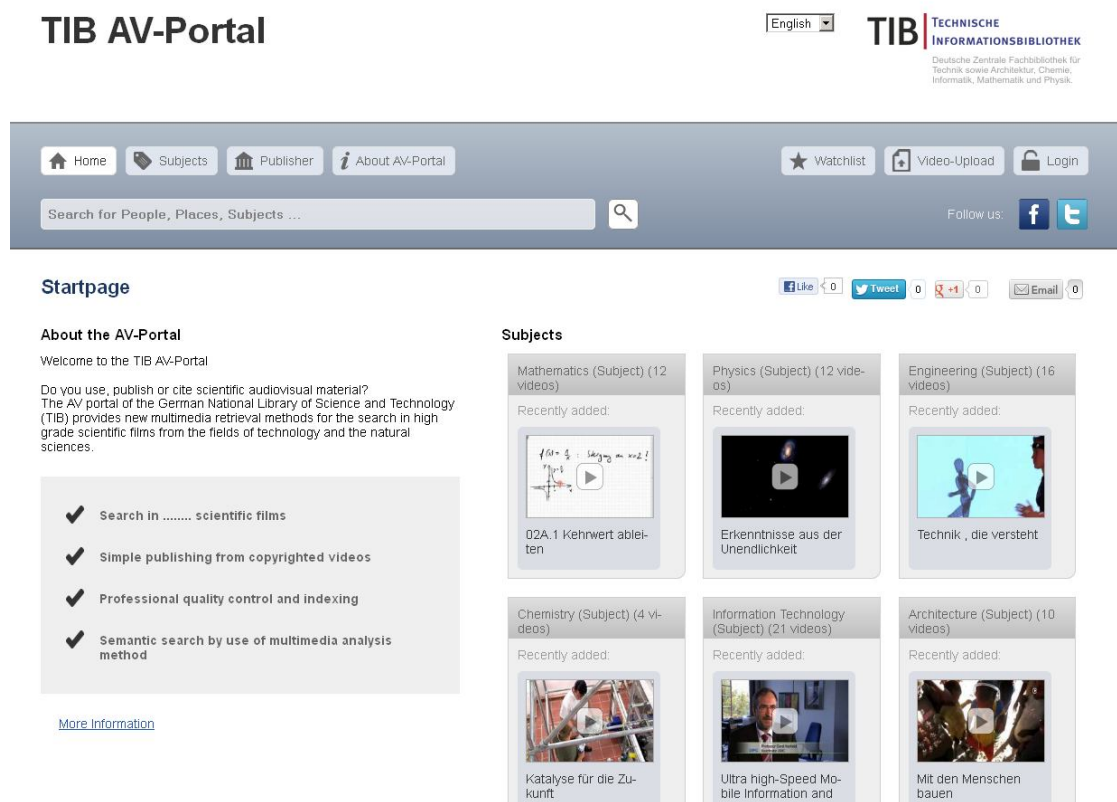


图 1 起始页

起始页面的左侧（见图 1）有一段简短的描述，解释了门户使用的视频检索技术。首先支持用户使用布尔“AND”连接多个检索词进行精确检索。此外还支持通过浏览查看门户提供的内容和功能。在起始页面的右侧为 TIB 主要涵盖的六大学科，其它学科在页面的下方。然而，对于那些想获取所在学科的基本信息，却不了解相关主题、学科领域和作者的用户可以使用“subject（学科）”和“producers（制作者）”等过滤词。在门户的顶部设置了查看列表、视频上传和登录功能。对于希望观看和下载有授权限制的视频的用户来说，登录功能是不可少的。这一门户将有德文和英文两个版本，英文版尚未完成。

[Home](#)
[Subjects](#)
[Publisher](#)
[About AV-Portal](#)

[★ Watchlist](#)
[Video-Upload](#)
[Login](#)


[Follow us: !\[\]\(8be75a20d635c380cd7a52c5c7bbbed5_img.jpg\) !\[\]\(92a9c66d52fa00484c508cd82aded8f9_img.jpg\)](#)

Search results

Search for "Roboter" in GetInfo (1315 results)


[Relevance](#)
[Most popular](#)
[Recently added](#)
[Recently viewed](#)

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




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 ge...


Metadata: **2** Analysis: **10**

★ Add to Watchlist




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 ↓

图2 检索结果

相匹配的视频显示为缩微图和视频片段（见图 2）。可以对匹配是源自元数据还是源自语音、文本和图像识别的媒体分析进行核查。可以使用分面搜索缩小检索结果的范围。分面中的 *Subject*（学科），*Publisher*（出版商）和 *License*（授权）来自于静态元数据，是作者与视频一同发布的。其他分面中的条目是通过自动视频分析提取的。*Category*（类型）分面源自可视化分析，包含的视频类型如采访、会议、内景拍摄或外景拍摄。分面中的 *Person*（个人），*Organisations*（机构），*Places*（地点）和 *Other*（其他）概念源自语音或文本识别。



Video Detail Page



Maschinen mit Köpfchen

图 3 播放器与自动媒体分析获得的元数据的聚合

点击检索结果列表中感兴趣的视频，便可以显示详细信息（见图 3）。门户应用了 HTML 5 播放器和作为备用的 flash 播放器。通过移动片段可以获得可视索引，提供视频内容的快速概览，便于特定片段的获取。红色片段中包含了检索词，而黄色片段为当前正在播放的片段。片段是以时间码为基础进行索引的。时间码显示在页面的右侧，后随自动媒体分析时提取的元数据。从语音、文本和视觉分析中获取的数据聚合在一起，并用颜色编码。当前播放的视频片段突出显示。点击其中的关键词用户可以跳到想要观看的片段。用户也可以通过检索条目在提取的元数据中进行检索。

5 结语

非文本媒介的数量、使用和重要性都在与日俱增，但这些文献中只有很少一部分可以实现检索。面临这些新的挑战，图书馆必须开放门户容纳非文本信息，开发标引、检索、浏览和显示数据的新工具，并用语义信息丰富这些数据。

TIB 将先进的多媒体分析方法应用于视频内容检索，通过情景、语音、文本和图像识别的自动分析生成补充的元数据。检索结果通过数据的语义关联与新知识相连。

- 情景识别：可视化目次提供视频内容的快速概览，便于特定片段的获取。
- 图像识别：以视频中的视觉特征（如颜色分布）为基础，系统自动识别视频为讲座、采访或实验。
- 语音和文本识别：视频中的口头语言和书写语言（如标志或幻灯片）都可以自动识别。检索词突出显示，方便视频内部的导航。

- 语义检索：通过将语义信息添加到视频分析得出的数据中，可以实现对资源库的导航，支持视听媒介间的关联。

AV 门户的目标是将现有的对书目元数据的文本检索扩展为媒介和跨数据的检索。这样，有数值数据和事实的数字全文、其他研究信息、视听媒介、可视化等都可以整合到用户界面。由于视听媒介与 TIB 门户 GetInfo (GetInfo 中收纳了数字全文、数值数据和事实及研究数据等) 间的关联，检索空间得以拓宽。通过配置数字对象标识符 (DOI, Digital Object Identifier) 视听媒介可以作为参考引用。AV 门户的检索工具提供了创新性的检索情景和挖掘知识的新方法。

6 参考文献:

-
- [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 audiominig 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>