
Englischsprachige Erweiterung des TIB|AV-Portals Ein GND-DBpedia-Mapping zur Gewinnung eines englischen Begriffssystems

Sven Strobel, Technische Informationsbibliothek Hannover

Zusammenfassung:

Die Videos des TIB|AV-Portals werden mit insgesamt 63.356 GND-Sachbegriffen aus Naturwissenschaft und Technik automatisch verschlagwortet. Neben den deutschsprachigen Videos verfügt das TIB|AV-Portal auch über zahlreiche englischsprachige Videos. Die GND enthält zu den in der TIB|AV-Portal-Wissensbasis verwendeten Sachbegriffen nur sehr wenige englische Bezeichner. Es fehlt demnach ein englisches Indexierungsvokabular, mit dem die englischsprachigen Videos automatisch verschlagwortet werden können. Die Lösung dieses Problems sieht wie folgt aus: Die englischen Bezeichner sollen über ein Mapping der GND-Sachbegriffe auf andere Datensätze gewonnen werden, die eine englische Übersetzung der Begriffe enthalten. Die verwendeten Mappingstrategien nutzen die DBpedia, LCSH, MACS-Ergebnisse sowie den WTI-Thesaurus. Am Ende wurde für 35.025 GND-Sachbegriffe (mindestens) ein englischer Bezeichner ermittelt. Diese englischen Bezeichner können für die automatische Verschlagwortung der englischsprachigen Videos unmittelbar herangezogen werden. 11.694 GND-Sachbegriffe konnten zwar nicht ins Englische ‚übersetzt‘, aber immerhin mit einem Oberbegriff assoziiert werden, der eine englische Übersetzung hat. Diese Assoziation dient der Erweiterung der Suchergebnisse.

Summary:

The videos in the TIB|AV-Portal are automatically indexed with a total of 63,356 GND subject headings in the field of science and technology. Besides the German-language videos, the TIB|AV-Portal also includes numerous English-language videos. However, the GND (Integrated Authority File) contains very few English labels for the subject headings used in the TIB|AV-Portal knowledge base. Thus, there is no English indexing vocabulary for the automatic indexing of English-language videos. The solution to this problem is as follows: English labels are determined by mapping the GND subject headings to other data sets which provide an English translation of these terms. The mapping strategies which have been applied make use of the DBpedia, LCSH, MACS results, and the WTI Thesaurus. As a result, (at least) one English label was determined for 35,025 GND subject headings. These English labels can be used directly for the automatic indexing of English-language videos. 11,694 GND subject headings could not be directly ‘translated’ into English, but at least it was possible to relate them to a broader term which has an English translation. This association helps to expand the search results.

Zitierfähiger Link (DOI): [10.5282/o-bib/2014H1S197-204](https://doi.org/10.5282/o-bib/2014H1S197-204)

Autorenidentifikation: Strobel, Sven: GND 136445365

1. Einleitung

Der Beitrag befasst sich im Kern damit, wie ein englisches Indexierungsvokabular für die Verschlagwortung der englischsprachigen Videos des TIB|AV-Portals gewonnen werden kann.¹ Kapitel 2 stellt zunächst das TIB|AV-Portal vor. Kapitel 3 behandelt die automatische Verschlagwortung der Videos des TIB|AV-Portals. In Kapitel 4 wird das Problem der Verschlagwortung englischsprachiger Videos vorgestellt. Anschließend wird die Lösung des Problems ausführlich erörtert, d.h. das Mapping der GND-Sachbegriffe auf andere Normdatensätze, die über eine englische Übersetzung der Begriffe verfügen. Zu diesen Normdatensätzen gehören die DBpedia, LCSH, MACS-Ergebnisse und der WTI-Thesaurus. Der Beitrag schließt mit einer Zusammenfassung der Ergebnisse und einem Ausblick auf geplante Vorhaben.

2. Das TIB|AV-Portal

Das TIB|AV-Portal² wurde vom Kompetenzzentrum für nicht-textuelle Materialien (KNM)³ an der Technischen Informationsbibliothek (TIB)⁴ in Kooperation mit dem Hasso-Plattner-Institut für Softwaresystemtechnik⁵ entwickelt. Das KNM arbeitet an grundlegenden Verbesserungen der Zugangs- und Nutzungsbedingungen für audiovisuelle Medien, 3D-Objekte und Forschungsdaten aus Technik und Naturwissenschaft.

Das TIB|AV-Portal optimiert den Zugang zu wissenschaftlichen Videos wie z.B. Computervisualisierungen, Lernmaterialien, Simulationen, Experimenten, Interviews, Vorlesungs- oder Konferenzaufzeichnungen. Sein Sammlungsprofil orientiert sich an den sechs TIB-Kernfächern: Technik sowie Architektur, Chemie, Informatik, Mathematik und Physik. Das TIB|AV-Portal hat derzeit 650 Stunden Videomaterial archiviert und enthält darüber hinaus 1.800 Filmnachweise mit externen Links. Am 29. April 2014 ist das Portal online gegangen.

Ein wesentliches Merkmal des Portals ist, dass es eine Reihe von Multimedia-Retrieval-Techniken einsetzt.⁶ Die Prozesskette beinhaltet den Ingest der AV-Medien, eine Szenenerkennung (Shot Boundary Detection), Spracherkennung (Speech to Text), Texterkennung (OCR), Bilderkennung (Visual Concept Detection) sowie automatische Verschlagwortung (Named Entity Recognition).⁷

1 Der vorliegende Beitrag ist die Verschriftlichung eines Vortrags auf dem 103. Deutschen Bibliothekartag in Bremen.

2 <https://av.getinfo.de> (25.10.2014).

3 <http://www.tib-hannover.de/de/dienstleistungen/kompetenzzentrum-fuer-nicht-textuelle-materialien-knm> (25.10.2014).

4 <http://www.tib-hannover.de/de/die-tib> (25.10.2014).

5 <http://www.hpi.uni-potsdam.de> (25.10.2014).

6 Vgl. Plank, Margret; Neumann, Janna: TIB's portal for audiovisual media. New ways of indexing and retrieval. In: IFLA journal 40 (2014), S. 17-23, http://www.ifla.org/files/assets/hq/publications/ifla-journal/ifla-journal-40-1_2014.pdf (25.10.2014).

7 Vgl. Sack, Harald; Plank, Margret: AV-Portal. The German National Library of Science and Technology's semantic video portal. In: ERCIM news Nr. 96 (Januar 2014), S. 33-34, <http://ercim-news.ercim.eu/en96> (25.10.2014). Zur Bilderkennung vgl. Hentschel, Christian; Blümel, Ina; Sack, Harald: Automatic annotation of scientific video material based on visual concept detection. In: Lindstaedt, Stefanie (Hg.): Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies, New York: ACM, 2013, Artikel Nr. 16, <http://dx.doi.org/10.1145/2494188.2494213> (25.10.2014).

Der Schwerpunkt dieses Beitrags liegt auf der automatischen Verschlagwortung des Portals, die im nächsten Kapitel erläutert wird.

3. Automatische Verschlagwortung der Videos mit Hilfe der GND

Sprach- und Texterkennung sind der automatischen Verschlagwortung vorgelagert. Mit Speech to Text wird die gesprochene Sprache im Video notiert und in Form eines Audiotranskripts gespeichert. Die OCR-Analyse erkennt die geschriebene Sprache im Video – z.B. Text auf Folien – und repräsentiert diese in einem OCR-Transkript. Die eigentliche Verschlagwortung des Videos leistet die Named Entity Recognition (NER). NER bezeichnet die Zuordnung von Begriffen aus einem Referenzvokabular zu analysierten Textinhalten. Im Fall des TIB|AV-Portals besteht das Referenzvokabular aus Sachbegriffen der Gemeinsamen Normdatei (GND); die analysierten Textinhalte sind die Audio- und OCR-Transkripte. Die NER extrahiert demnach GND-Sachbegriffe aus den Audio- und OCR-Transkripten und verschlagwortet damit das Video. Die Verschlagwortung verläuft auf der Segmentebene, so dass eine zielgenaue, segmentbasierte Suche innerhalb des Videos möglich ist (vgl. Abb.). Auf der rechten Seite der Abbildung befinden sich die GND-Sachbegriffe, mit denen die einzelnen Videosegmente automatisch verschlagwortet wurden.

The screenshot displays the TIB|AV-Portal interface. At the top, there is a navigation bar with links for Home, Fächer, Herausgeber, and Über AV-Portal, along with utility icons for Merkleliste, Hochladen, Einloggen, and Registrieren. A search bar contains the word 'schaum'. Below the search bar, there are social media icons for Facebook and Twitter, and a 'Folgen Sie uns:' section. The main content area shows a video player with a red progress bar and a play button. To the right of the video player is a sidebar titled 'Automatisierte Medienanalyse (i)'. This sidebar has tabs for 'Erkannte Entitäten' and 'Audiotranskript'. Below the tabs is a search input field and three checked filter boxes: 'Spracherkennung', 'Texterkennung', and 'Bildinhalt'. The sidebar displays a list of time-coded segments with associated GND terms: 00:00 (Schaum, Computeranimation, Hannover), 00:21 (Zugbeanspruchung, Nährstoff, Lebensmittel, Nuklearfaktor 3, Biochemiker, Lösung, Käse, Experiment innen, Stickstoffkürnung, Pharmazeutische Industrie, Filter, Surfactant-Faktor, Enzym, Chemischer Prozess), 01:22 (Experiment innen, Schaum, Enzym, Lösung), 02:46 (Experiment innen, Setzen <Verfahrenstechnik>, Besprechung/Interview, Lösung), and 03:13 (Blitzschlagsyndrom, Kr).

Abb.: Detailansicht des TIB|AV-Portals

Die GND-Datensätze wurden als Linked Data unter CC0-Lizenz publiziert.⁸ Die Normdatensätze sind durch Relationen miteinander verknüpft, so dass eine semantische Suche darauf aufgebaut werden kann. Die Bezeichner der GND sind vorwiegend auf Deutsch. Die Verschlagwortung der Videos verläuft fachbezogen. Das heißt, jedes der sechs TIB-Fächer hat seinen eigenen Fachausschnitt aus der GND, mit denen die Videos annotiert werden. Insgesamt enthält die TIB|AV-Portal-Wissensbasis 63.356 Sachbegriffe.⁹

Die GND-Sachbegriffe muss man als Entitäten verstehen. Was heißt das? Entitäten beschreiben Dinge, d.h. Gegenstände, Individuen, Sachverhalte, Ereignisse etc., die in einem konkreten oder abstrakten Sinne existieren.¹⁰ Entitäten sind die Bausteine einer Ontologie und als solche durch Relationen (Oberbegriff, Unterbegriff, verwandter Begriff etc.) mit anderen Entitäten verknüpft. Entitäten haben Bezeichner, mit denen sie benannt werden. Im Fall der GND sind die Bezeichner sprachliche Zeichen wie *Personenkraftwagen* oder *Thermodynamik*.

4. Mapping der GND-Entitäten auf die DBpedia und andere Normdaten

Neben den deutschsprachigen Videos verfügt das TIB|AV-Portal auch über zahlreiche englischsprachige Videos. Die GND enthält zu den in der TIB|AV-Portal-Wissensbasis verwendeten Sachbegriffen aus Technik und Naturwissenschaft nur sehr wenige englische Bezeichner. Infolgedessen können die englischsprachigen Videos nicht verschlagwortet werden. Das wiederum hat zur Konsequenz, dass keine segmentbasierte Suche innerhalb englischsprachiger Videos mit Hilfe von GND-Sachbegriffen möglich ist. Die Frage ist: Wie können englische Bezeichner für die GND-Sachbegriffe der TIB|AV-Portal-Wissensbasis gewonnen werden?

Die englischen Bezeichner sollen über ein Mapping der GND-Sachbegriffe auf andere Datensätze ermittelt werden. Bei den verschiedenen Mappingstrategien geht es stets darum, dass die deutschen Bezeichner der GND-Sachbegriffe in einem anderen Datensatz nachgeschlagen werden, der eine englische Übersetzung des Begriffs enthält. Bei einem erfolgreichen Mapping können die englischen Bezeichner für die GND-Sachbegriffe extrahiert werden.¹¹

Für das Mapping werden folgende Datensätze verwendet: DBpedia, LCSH (Library of Congress Subject Headings), Ergebnisse des MACS-Projekts (Multi Lingual Access to Subjects) sowie der Thesaurus für Technik und Management des WTI-Frankfurt.¹² Das Mapping wurde vom Hasso-Plattner-Institut

8 <http://www.dnb.de/lds> (25.10.2014).

9 Vgl. Strobel, Sven; Plank, Margret: Semantische Suche nach wissenschaftlichen Videos. Automatische Verschlagwortung durch Named Entity Recognition. In: Zeitschrift für Bibliothekswesen und Bibliographie 61 (2014) H. 4/5, S. 255-259.

10 Vgl. Precht, Peter; Burkard, Franz-Peter (Hg.): Metzler Philosophie-Lexikon. Begriffe und Definitionen. 2. Aufl. Stuttgart/Weimar: Metzler, 1999, S. 135.

11 Vgl. Strobel, Sven: Wie das TIB|AV-Portal Englisch lernte. Eine englische Übersetzung für Sachbegriffe der AV-Portal-Wissensbasis. <http://blogs.tib.eu/wp/tib/2014/06/12/wie-das-tibav-portal-englisch-lernte-eine-englische-uebersetzung-fuer-sachbegriffe-der-av-portal-wissensbasis/> (26.10.2014).

12 Dbpedia: <http://de.dbpedia.org>; LCSH: <http://id.loc.gov/authorities/subjects.html>; MACS: http://www.dnb.de/DE/Wir/Kooperation/MACS/macs_node.html; WTI-Thesaurus: <http://www.wti-frankfurt.de/index.php/produkte-thesaurus> (09.12.2014).

für Softwaresystemtechnik im Auftrag der TIB durchgeführt.

4.1 Mapping der GND-Entitäten auf die DBpedia¹³

Die DBpedia enthält strukturierte Informationen aus der Wikipedia wie Infoboxen, Tabellen und Weblinks, die als Linked Data zur Verfügung stehen. Beim Mapping auf die DBpedia werden zunächst die Haupt- und Alternativbezeichner des GND-Sachbegriffs, z.B. *Advent_{gnd}* (Vorweihnachtszeit), in einem Lexikon nachgeschlagen, das die Haupt- und Alternativbezeichner der deutschen DBpedia-Entitäten enthält. Der Hauptbezeichner der DBpedia-Entitäten entspricht dem Wert der Property *dfs:label*; Alternativbezeichner sind die Bezeichner der Redirects, Begriffserklärungsseiten und Ankertexte in den korrespondierenden Wikipedia-Artikeln.

Meist werden mehrere DBpedia-Kandidaten gefunden, auf die die Abbildung möglich ist, also etwa: *Advent_{dbp}* (Vorweihnachtszeit), *Advent_{dbp}* (Computer), *Advent_{dbp}* (Band), *Advent_{dbp}* (Herausgeber) etc. In dem Fall muss eine Auflösung der Mehrdeutigkeit (Disambiguierung) erfolgen. Der Disambiguierungsalgorithmus gleicht in einer Kookkurrenzanalyse die Kontextinformationen des GND-Sachbegriffs mit den Kontextinformationen der DBpedia-Kandidaten ab. Zu den Kontextinformationen der GND-Entität gehören Haupt- und Alternativbezeichner, Fachbereich, Definition, verwandte Begriffe und Oberbegriffe aus der GND. Zu den Kontextinformationen der DBpedia-Entitäten gehören die korrespondierenden Wikipedia-Artikel.

Vereinfacht gesagt, gewinnt derjenige DBpedia-Kandidat das Mapping, dessen Kontextinformation die größte Übereinstimmung mit der Kontextinformation des GND-Sachbegriffs aufweist. In unserem Beispiel ist das die deutsche DBpedia-Entität *Advent_{dbp}* (Vorweihnachtszeit). Im nächsten Schritt erfolgt über den Interlanguage Link (owl:sameAs) die Abbildung der deutschen DBpedia-Entität *Advent_{dbp}* auf die englische DBpedia-Entität *advent_{dbp}*. Damit kann der Bezeichner *advent* der englischen DBpedia-Entität extrahiert werden, und der GND-Sachbegriff ist erfolgreich ins Englische 'übersetzt' worden.¹⁴ Auf diese Weise konnten 28.691 (45 %) GND-Sachbegriffe der TIB|AV-Portal-Wissensbasis (insg. 63.356 Sachbegriffe) auf englische DBpedia-Entitäten gemappt werden.

Die Kookkurrenzanalyse basiert auf der folgenden Formel:

$$\text{Score} = \text{Anzahl} \times \frac{|W_{\text{dbp}} K(\text{gnd})|}{|K(\text{gnd})|}$$

Anzahl gibt an, wie oft die Begriffe des Kontexts der GND-Entität insgesamt im Kontext des DBpedia-Kandidaten (Wikipedia-Artikel) vorkommen. $|W_{\text{dbp}} K(\text{gnd})|$ repräsentiert die Größe der Schnittmenge der Begriffe des GND-Kontexts und aller Begriffe, die im Wikipedia-Artikel des DBpedia-Kandidaten vorkommen. $|K(\text{gnd})|$ steht für die Größe des Kontexts der GND-Entität. Der DBpedia-Kandidat,

13 Die Darstellung des GND-DBpedia-Mappings in Kapitel 4.1 basiert im Wesentlichen auf den Projektberichten des Hasso-Plattner-Instituts (Projektleiter: Harald Sack).

14 Vgl. Steinmetz, Nadine; Sack, Harald: Cross-Lingual semantic mapping of authority files. Präsentation auf der Konferenz „Semantic Web in Libraries“. SWIB 2013, Hamburg. Video zur Präsentation: <http://www.scivee.tv/node/61558> (26.10.2014).

der nach dieser Berechnung den höchsten Score erzielt, wird als der Kandidat mit der höchsten Wahrscheinlichkeit für ein korrektes Mapping identifiziert.

Die Ergebnisse des Algorithmus für das Mapping der GND-Entitäten auf die deutsche DBpedia wurden qualitativ evaluiert. Zu diesem Zweck wurde händisch ein Evaluationsdatensatz erstellt, der 180 intellektuell überprüfte Mappings von GND-Entitäten der TIB|AV-Portal-Wissensbasis auf die entsprechenden deutschen DBpedia-Entitäten enthält. Der Algorithmus konnte 90 % der GND-Entitäten des Evaluationsdatensatzes mappen (Recall: 90 %). 92 % der GND-Entitäten wurden dabei korrekt auf die deutsche DBpedia abgebildet (Precision: 92 %).

4.2 Mapping der GND-Entitäten auf die LCSH

Der Linked-Data-Dump der LCSH enthält Mappings von LCSH-Entitäten in die GND (LCSH nach GND), ebenso wie Mappings auf andere Vokabularien (z.B. RAMEAU). Bei dieser Mappingstrategie müssen die GND-Entitäten der TIB|AV-Portal-Wissensbasis unter den LCSH-Mappings identifiziert werden. Dies geschieht über die GND-ID (URI des Datensatzes). Bei erfolgreicher Identifikation kann der englische Bezeichner der verlinkten LCSH-Entität für die GND-Entität extrahiert werden. Insgesamt konnten darüber 11.169 (18 %) der GND-Sachbegriffe der TIB|AV-Portal-Wissensbasis auf LCSH-Entitäten abgebildet werden.

4.3 Mapping der GND-Entitäten auf die MACS-Ergebnisse

Im Projekt MACS wurden u.a. manuelle Verlinkungen zwischen Sachbegriffen der GND und LCSH (GND nach LCSH) erstellt. Zum Zeitpunkt des Mappings waren die MACS-Ergebnisse noch nicht im GND-Dump veröffentlicht. Die TIB konnte in Absprache mit der Deutschen Nationalbibliothek diese Daten vorab mit Hilfe der WinIBW beziehen und für das Mapping nutzen. Auch bei dieser Mappingstrategie müssen die GND-Entitäten der TIB|AV-Portal-Wissensbasis unter den MACS-Mappings identifiziert werden. Danach kann der englische Bezeichner der verlinkten LCSH-Entität für die GND-Entität extrahiert werden. Über diese Methode konnten 13.692 (22 %) GND-Sachbegriffe der Wissensbasis den entsprechenden LCSH-Entitäten zugeordnet werden.

4.4 Mapping der GND-Entitäten auf den WTI-Thesaurus

Der Thesaurus für Technik und Management des WTI-Frankfurt ist ein maschinenlesbarer Thesaurus, der von der TIB lizenziert wurde. Der WTI-Thesaurus enthält Begriffssätze mit deutschen Bezeichnern sowie (mindestens) einem englischen Bezeichner. Die Haupt- und Alternativbezeichner der GND-Entitäten werden im WTI-Thesaurus nachgeschlagen, potenzielle Mapping-Kandidaten werden identifiziert und disambiguiert. Am Ende konnte für 14.719 (23 %) GND-Sachbegriffe der TIB|AV-Portal-Wissensbasis mindestens ein englischer Bezeichner aus dem WTI-Thesaurus ermittelt werden.

5. Ergebnisse

Das Gesamtergebnis der verschiedenen Mappingstrategien zur Gewinnung englischer Bezeichner findet sich in der folgenden Tabelle:

Tab. 1: Ermittelte englische Bezeichner

GND-Sachbegriffe	GND-Sachbegriffe mit englischem Bezeichner		GND-Sachbegriffe ohne englischen Bezeichner, aber mit verknüpftem ‚übersetzten‘ Oberbegriff	
63.356	35.025	55 %	11.694	19 %

Insgesamt enthält die TIB|AV-Portal-Wissensbasis mit den sechs GND-Fachauschnitten 63.356 GND-Sachbegriffe. 35.025 (55 %) GND-Sachbegriffe haben (mindestens) einen englischen Bezeichner ermittelt bekommen. (Teilweise konnten über die verschiedenen Mappingstrategien mehrere englische Bezeichner für die GND-Sachbegriffe extrahiert werden.) Diese englischen Bezeichner können unmittelbar zur automatischen Verschlagwortung der englischsprachigen Videos herangezogen werden.

Für 11.694 (19 %) GND-Sachbegriffe gilt, dass für sie zwar kein englischer Bezeichner ermittelt werden konnte, sie aber immerhin mit einem Oberbegriff assoziiert wurden, der eine englische Übersetzung hat. Diese Assoziation ermöglicht es, dass bei der Suche nach einem ‚übersetzten‘ Oberbegriff, z.B. *waste heat* (*Abwärme_{gnd}*), auch Suchergebnisse zum nicht-übersetzbaren Unterbegriff *Industrieabwärme_{gnd}* zurückgeliefert werden.

6. Ausblick

Die Mappingergebnisse sind in Form von RDF in einem Triplestore abgelegt. Bislang sind diese Ergebnisse nur intern verfügbar. Für die Zukunft ist geplant, diese Daten als Linked Open Data zu publizieren, um deren Sichtbarkeit im Web of Data zu erhöhen und eine Nachnutzbarkeit in anderen Projekten zu ermöglichen.

Außerdem soll die TIB|AV-Portal-Wissensbasis regelmäßig aktualisiert werden. Es müssen zunächst die Updates der GND, DBpedia und LCSH zu den für die Wissensbasis relevanten Fachauschnitten bezogen werden. Danach müssen die GND-Sachbegriffe erneut auf die aktuellen Versionen der DBpedia und LCSH sowie auf die statischen Versionen des WTI-Thesaurus und der MACS-Ergebnisse gemappt werden. Auf diese Weise wird die Aktualität der Begriffe gewährleistet.

Literaturverzeichnis

- Hentschel, Christian; Blümel, Ina; Sack, Harald: Automatic annotation of scientific video material based on visual concept detection. In: Lindstaedt, Stefanie (Hg.): Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies, New York: ACM, 2013, Artikel Nr. 16, <http://dx.doi.org/10.1145/2494188.2494213> (25.10.2014).

- Plank, Margret; Neumann, Janna: TIB's portal for audiovisual media. New ways of indexing and retrieval. In: IFLA journal 40 (2014), S. 17-23, http://www.ifla.org/files/assets/hq/publications/ifla-journal/ifla-journal-40-1_2014.pdf (25.10.2014).
- Prechtel, Peter; Burkard, Franz-Peter (Hrsg.): Metzler Philosophie-Lexikon. Begriffe und Definitionen. 2. Aufl. Stuttgart/Weimar: Metzler, 1999.
- Sack, Harald; Plank, Margret: AV-Portal. The German National Library of Science and Technology's semantic video portal. In: ERCIM news Nr. 96 (Januar 2014), S. 33-34, <http://ercim-news.ercim.eu/en96> (25.10.2014).
- Steinmetz, Nadine; Sack, Harald: Cross-Lingual semantic mapping of authority files. Präsentation auf der Konferenz „Semantic Web in Libraries“. SWIB 2013, Hamburg. Video zur Präsentation: <http://www.scivee.tv/node/61558> (26.10.2014).
- Strobel, Sven: Wie das TIB|AV-Portal Englisch lernte. Eine englische Übersetzung für Sachbegriffe der AV-Portal-Wissensbasis. <http://blogs.tib.eu/wp/tib/2014/06/12/wie-das-tibav-portal-englisch-lernte-eine-englische-uebersetzung-fuer-sachbegriffe-der-av-portal-wissensbasis/> (26.10.2014).
- Strobel, Sven; Plank, Margret: Semantische Suche nach wissenschaftlichen Videos. Automatische Verschlagwortung durch Named Entity Recognition. In: Zeitschrift für Bibliothekswesen und Bibliographie 61 (2014) H. 4/5, S. 255-259.