

Evaluation einer Audio Mining-Suche im Kontext der 'Social Connected TV'-Plattform*

Tobias Simon¹, Joachim Köhler², Sven Pagel¹

Fachbereich Wirtschaft, Hochschule Mainz¹

NetMedia, Fraunhofer-Institut Intelligente Analyse- und Informationssysteme IAIS²

1 Einleitung

Mit steigender Kapazität der Internetbandbreite (Paul & Wolf 2014, S. 64) und verbesserten Möglichkeiten der Datenkomprimierung (Oh & Hua 2000, S. 415), nimmt auch die Bedeutung von digitalen Videos eine immer wichtigere Rolle in einem breiten Spektrum von Multimedia-Anwendungen ein (Fan et al 2004, S. 974; TNS Infratest 2014, S. 55 ff.). Zu diesen Anwendungen zählen beispielsweise Videokonferenzen, Video on Demand (VoD), Internet Protocol Television (IPTV), digitale Bibliotheken, öffentliche Informationssysteme oder E-Commerce-Anwendungen (Oh & Hua 2000, S. 415; Vranješ et al 2013, S. 1). Durch die steigende Verfügbarkeit von digitalen Videoaufnahmegegeräten steigt auch die Anzahl der von Nutzern selbst erzeugten Videoinhalten auf den entsprechenden Videoplattformen im Internet (Ling Shao et al 2014, S. 504). Auf der Videoplattform YouTube werden nach Angaben der Betreiber durchschnittlich pro Minute 100 Minuten Videomaterial hochgeladen (Google Inc. 2014). Dieses exponentielle Wachstum im Bereich von Online Videos und das damit verbundene steigende Interesse der Nutzer an Aktivitäten, die mit diesen Videos in Verbindung stehen, ist ein konstantes Phänomen, welches über die gesamte letzte Dekade beobachtet werden kann (Liu et al 2013, S. 1).

Der schnelle Anstieg der Datenbestände in diesem Bereich bringt allerdings auch Probleme mit sich. Hier stellt sich die Frage, wie Nutzer auf diese riesigen Datensammlungen möglichst effizient zugreifen können und die gesuchten Inhalte möglichst schnell finden (Ling Shao et

*Reprint: Pagel, Sven; Köhler, Joachim; Simon, Tobias: Evaluation einer Audio Mining-Suche im Kontext der „Social Connected TV“-Plattform. Update 22 Hochschule Mainz – SS 2016 – Forschung und Technik. S.50-55. ISSN (1861-3152)

al 2014, S. 504). Eine mögliche Antwort auf diese Frage bietet die hier thematisierte Technologie Audio Mining.

Die vorliegende Arbeit ist in Zusammenarbeit mit dem Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS, Abteilung NetMedia, in Sankt Augustin entstanden. Im Rahmen des Forschungsprojekts „Social Connected TV“-Plattform (SCTVP) arbeitete das Fraunhofer IAIS unter anderem an der Audio Mining Technologie. Mit Hilfe dieser Technologie können große Datenbestände von Videomaterial automatisiert, indiziert und für Nutzer durchsuchbar gemacht werden. Die Evaluation einer Audio Mining-Suche im Kontext der „Social Connected TV“-Plattform soll in diesem Beitrag vorgestellt werden.

1.1 „Social Connected TV“-Plattform

Die SCTVP ist neben der „Smart City Services“- sowie der „Pervasive Games“-Plattform eine von drei innovativen Informations- und Kommunikationsplattformen, die im Rahmen des Forschungs- und Entwicklungsprojekts FIcontent entwickelt wurden. FIcontent ist ein Bestandteil des EU-finanzierten „Future Internet Public Private Partnership“ (FI-PPP) Programms. (Cabel et al 2014, S. 1; FIcontent 2014a, S. 1 f.)

Die drei Hauptanwendungsgebiete der SCTVP sind Anwendungsszenarien mit Interaktion zwischen mehreren Bildschirmen, ein verbessertes personalisiertes Fernseherlebnis sowie User-Tracking inklusive der entsprechenden Datenschutzmechanismen (FIcontent 2014b; Krauss et al 2014, S. 2).

Wie in Abbildung 1 zu sehen ist, besteht die SCTVP im Kern aus zwei Ebenen. Den Generic Enablers (GE) und Specific Enablers (SE). SEs sind unterschiedliche Content-Zentrierte Softwarekomponenten, die mithilfe von offenen Programmierschnittstellen (Application Programming Interface – API) für Softwareentwickler sowie für kleinere und mittlere Unternehmen (KMUs) zur Verfügung gestellt werden. Generic Enablers sind technologische Bausteine die durch das FI-WARE¹ Projekt bereitgestellt werden. Sie bilden die technische Grundlage für einige der Specific Enablers oder können direkt in Anwendungsszenarien implementiert werden. (Krauss et al 2014, S. 2)

¹ <http://www.fi-ware.org/>

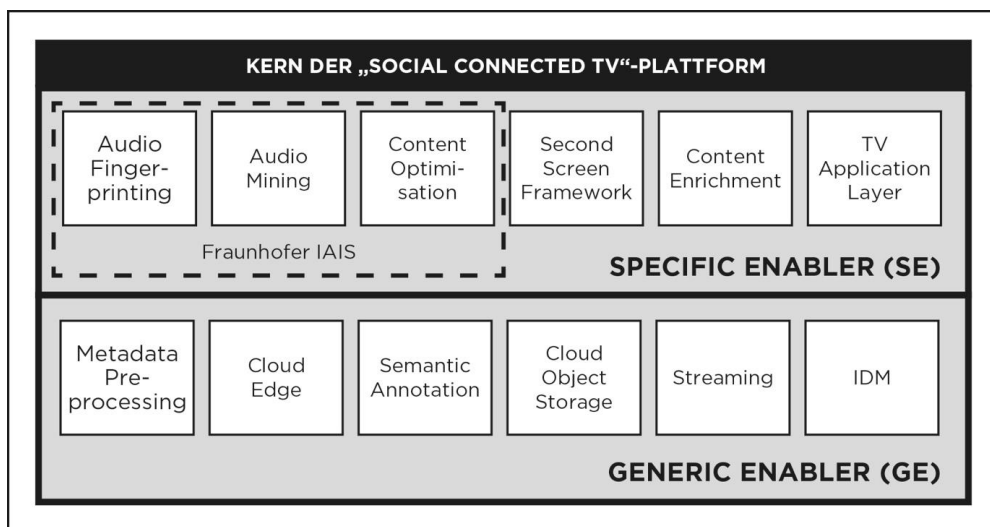


Abbildung 1: Kern der SCTVP [eigene Darstellung in Anlehnung an Krauss et al, 2014]

Eingesetzt werden die Content-zentrierten Technologien in Softwaresystemen und Anwendungen, welche zur Erschließung und Anreicherung von unstrukturierten Medien-Inhalten verwendet werden. Die Erschließung bzw. Anreicherung kann sowohl automatisch als auch manuell erfolgen. Dabei werden die Inhalte sowohl mit inhaltlichen als auch mit strukturellen Metadaten versehen und können mit weiteren Datenquellen verknüpft werden. (Eble & Winkler 2014, S. 233)

1.2 Audio Mining

Im Fokus dieser Zusammenarbeit zwischen Hochschule Mainz und Fraunhofer IAIS steht der Specific Enabler (SE) Audio Mining. Dieser wird im Bereich der Indexierung und Durchsuchung von Multimedia Datenbeständen eingesetzt. Der SE verarbeitet sowohl Audio als auch Video Dateien. Die Dateien werden zunächst in Sprach und Sprecher Segmente untergliedert, anschließend wird mit Hilfe von Spracherkennung gesprochene Sprache in digitalen Text umgewandelt. Die von diesem SE ausgegebenen textuellen Informationen können dann zur Indexierung der Multimediadaten verwendet werden. (FIcontent 2014c)

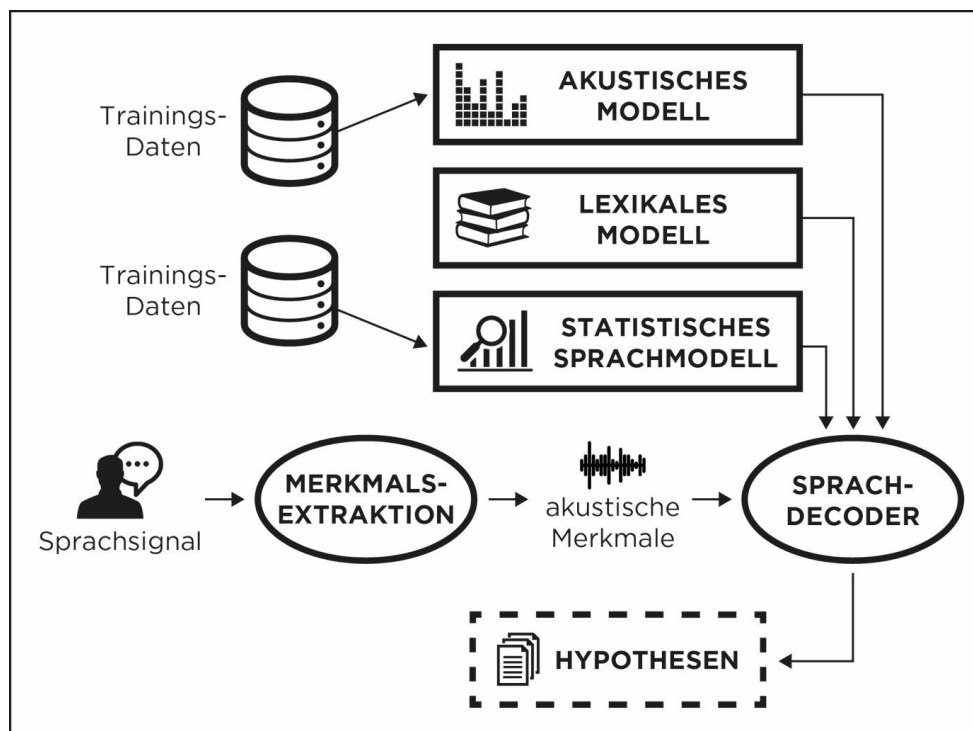


Abbildung 2: Architektur eines ASR Systems [Eigene Darstellung in Anlehnung an Besacier et al 2014, S. 90]

Die Erfassung und Weiterverarbeitung der Multimediadaten kann in vier Abschnitte untergliedert werden. Dies sind die Erfassung und Analyse der Daten, Extraktion von Schlüsselbegriffen, Segmentierung der Daten und das Finden von ähnlichen Inhalten.

Die Analyse der Daten wird mit einem Automatic Speech Recognition (ASR) System durchgeführt. Ein ASR System besteht aus drei wesentlichen Komponenten: Akustische Modellierung, Lexikale Modellierung und Sprachliche Modellierung (Young 2008, S. 540). Die Architektur eines solches System ist in Abbildung 2 dargestellt.

2 Forschungsdesign

2.1 Evaluationsgegenstand

Gegenstand der Evaluation ist die Suche in einer Musteranwendung des in Kapitel Audio Mining vorgestellten SE Audio Mining, welcher von Fraunhofer IAIS im Rahmen der „Social Connected TV“-Plattform entwickelt wurde. Die Web-Anwendung bietet eine Suchmaske, mit der die Nutzer eine Video-Datenbank des Rundfunks Berlin-Brandenburg (rbb) mit insgesamt 245 Videos durchsuchen können. Die Besonderheit liegt bei dieser Anwendung in den

vorgehaltenen Videos, die alle mit dem Audio Mining-Verfahren automatisiert aufbereitet wurden. Bei der Aufbereitung wurde von jedem Video ein vollständiges Transskript erstellt. Aus diesen Transskripten wurden wiederum zu jedem Video passende Schlüsselwörter generiert. Die Nutzer können somit nicht nur nach den Titeln suchen, sondern auch gezielt die Inhalte der Videos durchsuchen. Nach Absenden der Suchanfrage gibt die Anwendung alle Videos aus, in welchen der Suchbegriff als gesprochenes Wort vorkommt. Klickt der Nutzer auf ein Suchergebnis, startet das entsprechende Video genau in jenem Segment, in dem der Suchbegriff genannt wird. Kommt ein Begriff in einem Videobeitrag häufiger vor, werden in der Ergebnisliste pro Video alle Positionen, an denen der Begriff genannt wird, separat aufgeführt. Im Folgenden wird diese Suche als Audio Mining-Suche bezeichnet.

2.2 Evaluationsziele und Forschungsfragen

Die Evaluation verfolgt im Wesentlichen zwei Hauptziele: zum einen die Überprüfung, ob die Verwendung von Audio Mining für die Nutzer einen Mehrwert darstellt und wie dieser bewertet wird. Zum anderen soll die Evaluation Ansätze liefern, wie für den Nutzer ein größerer Mehrwert generiert werden kann. Um diese Ziele zu erreichen, werden die folgenden Forschungsfragen gestellt:

F 1: Entsteht durch die Audio Mining-Funktion ein Mehrwert für die Endanwender?

F 1.1: Sind die Nutzer in der Lage, mit Hilfe der Suchfunktion Inhalte in den Videos zu finden, die über den Titel der Videos nicht ersichtlich sind und die sie ohne die Audio Mining-Technologie nicht gefunden hätten?

F 1.2: Finden Nutzer, die gezielt nach Inhalten suchen, die Ergebnisse mit Hilfe der Audio Mining-gestützten Suche besser (Indikatoren: Geschwindigkeit, Treffsicherheit, Anzahl der Versuche)?

F 2: Wie kann der Mehrwert für die Nutzer vergrößert werden?

2.3 Methodik und Ablauf

Um die, in Kapitel 2.2, aufgestellten Forschungsfragen zu beantworten, wurde eine qualitative, explorative Studie durchgeführt. Dabei wurden die in Tabelle 1 näher beschriebenen Untersuchungsmethoden Interview, Retrospective Thinking Aloud (RTA), Befragung und Screenrecording-Analyse verwendet.

UNTERSUCHUNGS-METHODE	BESCHREIBUNG	ZIEL
Qualitatives Interview / RTA	Befragung der Probanden anhand eines Interview/RTA-Leitfadens	Erkenntnisse aus den Aussagen und aus den Gedanken der Probanden zu generieren

Befragung	Befragung der Probanden mit Hilfe eines Fragebogens	Schnelles Abfragen von Soziodemografischen Daten
Screenrecording-Analyse	Analyse der während des Tests aufgezeichneten Bildschirmaufnahme	Analytische Auswertung der gestellten Aufgaben

Tabelle 1: Untersuchungsmethoden

AUFGABE 1	Bitte finden Sie Videoinhalte über den Schriftsteller Rolf Schneider.
AUFGABE 2	Bitte finden Sie Videoinhalte über den Smogalarm in Berlin.
AUFGABE 3	Bitte finden Sie Videoinhalte über die RAF.
AUFGABE 4	Bitte finden Sie Videoinhalte über die Kommunalwahlen in der DDR.

Tabelle 2: Rechercheaufgaben der Probanden

Zu Beginn der Untersuchung bekamen die Probanden vier Aufgaben gestellt (Siehe Tabelle 2). Zur Bearbeitung der Aufgaben wurden den Probanden zwei unterschiedliche Suchmasken bereitgestellt. Beide Suchen griffen auf die gleiche Datenbank aus Kap. 2.1 zu. Die erste Suche durchsuchte dabei alle zur Verfügung stehenden Metadaten wie den Transskripten oder den Schlüsselwörtern der Videos. Die zweite Suche hingegen beschränkte sich bei der Suche auf die Titel der Videos. Nachdem die jeweiligen Teilnehmer die vier Aufgaben abgeschlossen hatten, wurde mit ihnen ein qualitatives Leitfadeninterview durchgeführt und aufgezeichnet. Optional konnte der Testleiter vor dem Interview noch ein Retrospective Think Aloud (RTA) durchführen; also ein nachträgliches Lautes Denken. Bei dieser Untersuchungsmethode werden dem Probanden die relevanten Ausschnitte aus der Screenrecording-Aufzeichnung gezeigt. Mit gezielten Fragen versucht der Testleiter gemeinsam mit dem Probanden dessen Gedankengänge während der Aufnahme zu rekonstruieren und zu verbalisieren (Jaspers 2009, S. 623). Abschließend wurden mittels Fragebogen die Soziodemografischen und die Zahlungsbereitschaft für ein Audio-Mining Feature der Teilnehmer erhoben. Das Setting des Versuchs ist in Abbildung 3 dargestellt.

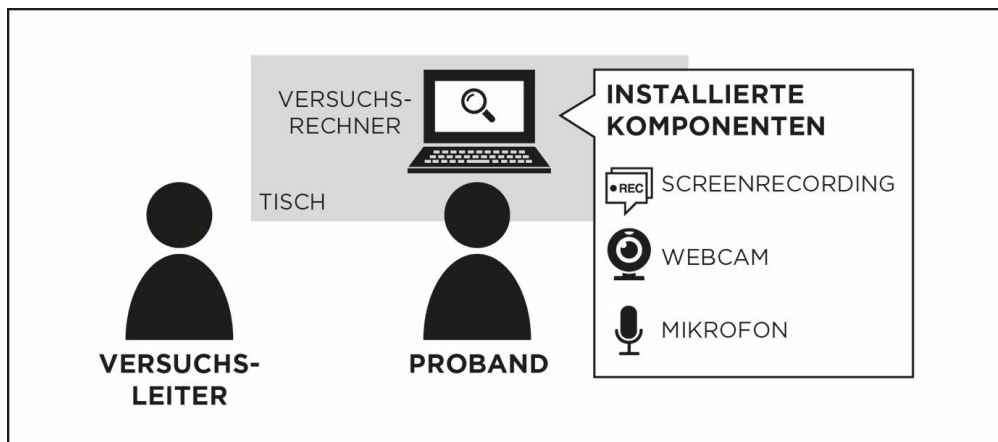


Abbildung 3: Testsetting und installierte Komponenten

2.4 Probanden

Als Probanden wurden Studenten aus dem Studiengang Medien, IT & Management der Hochschule Mainz akquiriert. Durch den Medienbezug dieses Studienganges ist die Recherche in einem Videoportal für diese Gruppe ein sowohl relevanter als auch bekannter Aufgabenbereich. Es handelte sich bei der Auswahl der Probanden also nicht um eine rein zufällige Stichprobe, sondern um ein bewusst gesteuertes Theoretical Sampling (Kelle & Kluge 2010, S. 48; Robinson 2014, S. 34). Auf Grund des Vorwissens in diesem Bereich eignete sich dieser Personenkreis besonders um mit qualitativen Untersuchungsmethoden interessante und aufschlussreiche Erkenntnisse zu gewinnen (Jaspers 2009, S. 623). Insgesamt nahmen zwölf Teilnehmer an der Untersuchung teil.

3 Interpretation der Ergebnisse

3.1 Audio Mining-Nutzer lösen mehr Aufgaben

Wie in Forschungsfrage F 1 formuliert, ist ein Ziel dieser Arbeit herauszufinden, ob die Audio Mining-Funktion einen Mehrwert für den Anwender darstellt. Um diese Frage zu beantworten, wird zunächst die Forschungsfrage F 1.1 „Finden die Anwender Videos, die sie ohne Audio Mining-Funktion nicht gefunden hätten?“ betrachtet. Die Ergebnisse der Untersuchung zeigen, dass, wenn der Titel des gesuchten Videos keine übereinstimmenden Wörter mit dem Suchauftrag aufweist, deutlich mehr Probanden mit der Audio Mining-Suche ein passendes Video finden als Probanden mit der Standardsuche. Bei Aufgabe 3 konnten alle sechs Probanden mit der Audio Mining-Suche ein passendes Video finden. Mit der Standardsuche schafften dies lediglich zwei der sechs Versuchspersonen. Die sechs Probanden mit der Audio Mining-Suche haben alle bereits mit dem ersten Versuch ein passendes Video gefunden. Die

beiden Versuchspersonen, die mit der Standardsuche ein passendes Video gefunden haben, müssen jeweils fünf unterschiedliche Suchphrasen ausprobieren, bis sie die Aufgabe lösen können. Die Forschungsfrage F 1.1 kann demnach mit „ja“ beantwortet werden. Die Anwender können mit Hilfe der Audio Mining-Suche Videos finden, die sie mit einer einfachen Suche nicht gefunden hätten. Hierdurch entsteht für den Nutzer ein klarer Mehrwert; in Form einer besseren Trefferquote kann dieser auch als Matchingwert bezeichnet werden. Durch die Tatsache, dass bei den Interviews vier der Probanden genau diesen Vorteil der Audio Mining-Suche betonen, wird dieser Mehrwert noch einmal unterstrichen. Allerdings trifft diese These nicht ohne Einschränkung zu. Betrachtet man Aufgabe 2, so ist hier der gleiche Effekt in die andere Richtung zu erkennen, wenn auch nicht so stark ausgeprägt. Hier können alle Probanden mit der Standardsuche, ein passendes Video finden. Mit der Audio Mining-Suche gelingt dies nur vier der sechs Testpersonen. Zurückzuführen ist dies auf die große Menge an Ergebnissen bei der Audio Mining-Suche. Da die Probanden bei dieser Suche alle bei jedem Versuch das Schlüsselwort „Berlin“ mit angegeben haben, werden vom System alle 192 Videos ausgegeben, in denen das Wort „Berlin“ gesprochen wird. Die beiden gesuchten Videos befinden sich dabei auf den Plätzen 29 und 112. Sie wurden von Probanden in der Menge der Suchergebnisse schlichtweg übersehen. Durch die Menge der Suchergebnisse kann sich der Mehrwert für den Nutzer wieder relativieren.

3.2 Audio Mining-Nutzer finden Videos effizienter

Bei der Forschungsfrage F 1.2 geht es darum, ob die gesuchten Videos mit der Audio Mining-Suche besser gefunden werden. Als Indikatoren wird hier die Anzahl der Versuche, die benötigte Zeit pro Aufgabe und die Zeit, die der Proband benötigt um ein richtiges Video als ein solches zu identifizieren, gemessen. Bei fast allen Aufgaben benötigen die Teilnehmer mit der Audio Mining-Suche weniger oder genauso viele Versuche wie mit der Standardsuche. Lediglich bei Aufgabe 2 benötigen die Versuchspersonen im Mittel einen Versuch mehr. Insgesamt werden bei der Audio Mining-Suche durchschnittlich 1,21 Versuche weniger benötigt um das Ziel zu erreichen.

Bei den Aufgaben 1 und 3 können die Probanden mit der Audio Mining-Suche die Videos schneller identifizieren als bei der Standardsuche. Bei den Aufgaben 2 und 4 ist es genau umgekehrt. Mögliche Ursache für das bessere Abschneiden der Standardsuche bei diesen beiden Aufgaben könnten die Anzahl der ausgegebenen Videos oder die Beschriftung der Videos sein. Bei Aufgabe 4 beispielsweise wurden am häufigste die beiden Begriffe „Kommunalwahl“ und „DDR“ in das Suchfeld eingegeben. Diese erzeugen bei der Audio Mining-Suche 121 Ergebnisse. Bei der Standardsuche werden mit den gleichen Begriffen lediglich fünf Ergebnisse ausgegeben, unter anderem das Video „Kommunalwahlen in der DDR“. Ein kurzer Blick auf die restlichen vier Titel verrät dem Probanden an dieser Stelle bereits, dass es nur dieses Video sein kann. Dies unterstützen auch die Aussagen der Probanden in den Interviews. Insgesamt können die Videos, mit der Audio Mining-Suche, durchschnittlich 3 Sekunden schneller identifiziert werden als mit der Standardsuche. Dies liegt vor allem daran, dass man beim Klicken auf ein Suchergebnis direkt an die entsprechende Stelle innerhalb des Videos geleitet wird. So kann der Anwender sehr schnell beurteilen, ob

das gesuchte Schlüsselwort auch im gewünschten Kontext verwendet wird. Auch dies wird durch die Aussagen der Teilnehmer in den Interviews bestätigt.

Insgesamt kann damit auch Forschungsfrage F 1.2 mit „ja“ beantwortet werden. Der Mehrwert entsteht für den Nutzer hier durch eine effizientere Suche, bei welcher er weniger Suchanfragen benötigt um das Ziel zu erreichen. Ein weiterer Mehrwert wird durch die „Sprünge zu Segment“-Funktion geschaffen. Mit dieser können die Nutzer die Videos schneller identifizieren und gleich an die für ihre Suche relevante Stelle im Video gelangen. Beide genannten Mehrwerte können in die Kategorie der Strukturierungswerte eingeordnet werden. Werden allerdings zu viele Ergebnisse ausgegeben, wirkt sich dies negativ auf die Effizienz aus, sodass bei solchen Suchaufträgen die Standardsuche besser geeignet ist.

3.3 Audio Mining als Mehrwert für den Endanwender

Nachdem die beiden Forschungsfragen F1.1 und F1.2 in den Kapiteln 3.1 und 3.2 beantwortet wurden, kann nun auch Forschungsfrage F1 beantwortet werden. Da in den beiden vorangegangenen Kapiteln bereits die Mehrwerte der Audio Mining-Suche aufgezeigt wurden, lautet die Antwort auf F1: Ja. Durch Audio Mining entsteht ein Mehrwert für den Endanwender. Dieser Mehrwert besteht als Strukturierungswert und durch die bessere Trefferquote als Matchingwert.

3.4 Größerer Mehrwert durch besseres Ranking

Nachdem in Forschungsfrage F1 der Audio Mining-Suche bereits ein Mehrwert bestätigt wurde, zielt Forschungsfrage F2 darauf ab, wie dieser Mehrwert noch gesteigert werden kann. Einen Ansatzpunkt zur Verbesserung des Selektionswerts bietet das Ranking der Suchergebnisse. Als Ranking wird die Reihenfolge der Suchergebnisse bezeichnet. In den Interviews äußerten sich alle Probanden negativ über das Ranking. Dies hat mehrere Ursachen. Beispielsweise ist die Reihenfolge der Suchergebnisse nicht nachvollziehbar. Hierbei hätten einige Probanden erwartet, dass Videos umso weiter oben ausgegeben werden, je öfter das Suchwort darin gefunden wird. Geht man davon aus, dass die Relevanz der Videos mit der Häufigkeit des Schlüsselbegriffs steigt, würden so die relevantesten Ergebnisse ganz oben stehen. Dies ist nicht der Fall. Gibt man zum Beispiel die Suchbegriffe „Rolf“ und „Schneider“ ein, wird in den ersten beiden Videos jeweils einmal eines der beiden Schlüsselwörter gefunden. In dem Video auf Platz drei kommt das Suchwort „Schneider“ hingegen sechsmal vor. Eine alphabetische Sortierung nach dem Titel oder der Video ID kann auch nicht festgestellt werden. Gerade bei Suchen mit vielen Ergebnissen, wie beispielsweise bei Eingabe der Suchbegriffe „Smogalarm“ und „Berlin“ (192 Treffer), wird die fehlende Sortierung zu einem Problem.

Ein weiterer wichtiger Punkt bei der Reihenfolge der Suchergebnisse ist, dass die Suchbegriffe getrennt voneinander behandelt werden. Bei der Suche nach „Rolf“ und „Schneider“ sollten Ergebnisse in denen beide Schlüsselwörter vorkommen an oberster Stelle stehen, da diese Videos vermutlich eine größere Relevanz für den Suchenden haben. Teilweise haben die Probanden sogar versucht dieses Verhalten der Suchmaske zu erzwingen. Die Teilnehmer

versuchten Google-Suchoperatoren² wie z.B. das Anführungszeichen zu verwenden, um nur Ergebnisse mit allen eingegebenen Suchbegriffen zu erhalten. Dies führte bei den Versuchen der Probanden nicht zum gewünschten Ergebnis. Auch diese Verbesserung des Rankings würde gerade bei Suchen mit vielen Ergebnissen den Suchprozess für den Nutzer vereinfachen.

Um abschließend F2 zu beantworten, kann festgehalten werden, dass durch eine Verbesserung des Rankings, vor allem bei Suchen mit vielen Ergebnissen, der Mehrwert für den Endanwender, durch die Steigerung des Selektionswerts, erheblich vergrößert werden kann.

4 Fazit

Zusammenfassend kann festgehalten werden, dass in dieser qualitativen Nutzerstudie mit 12 Teilnehmern ein Mehrwert der Audio Mining-Suche festgestellt werden konnte. Dieser besteht vor allem in einem höheren Strukturierungswert und einem besseren Matchingwert als bei der Standardsuche. Durch, dass finden von Inhalten die nicht im Titel benannt sind und somit nicht über die Standardsuche gefunden werden wird der Matchingwert deutlich verbessert. Durch die Einbettung der Treffer im jeweiligen Video-Kontext („Springe zu Segment-Funktion“) können die Ergebnisse schneller eingeordnet werden was eine Steigerung des Strukturierungswertes zur Folge hat. Durch die Optimierung des Rankings der Suchergebnisse könnte der Selektionswert der Suche noch weiter gesteigert werden. Hier gibt es demnach noch Optimierungspotential mit dem der Mehrwert der Audio Mining-Suche noch weiter gesteigert werden kann.

² <https://support.google.com/websearch/answer/2466433?hl=de>

Literatur

- Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014). Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, 56, 85–100. doi:10.1016/j.specom.2013.07.008
- Cabel, N., Elkhatib, Y., Seeliger, R., Ziegler, C., & Eble, M. (2014). Social Connected TV Platform by FIcontent: An Open Toolbox for Creative Developers. Gehalten auf der TVX'14. http://files.figshare.com/1505326/D_100.pdf. Zugegriffen: 24. September 2014
- Eble, M., & Winkler, T. (2014). Digitale Wertketten für Social Connected TV: Wertbeiträge von Content-Technologies in der Multimedia-Produktion. In H. Rau (Hrsg.), *Digitale Dämmerung: Die Entmaterialisierung der Medienwirtschaft* (Bd. 7, S. 229–239). Baden-Baden: Nomos.
- Fan, J., Luo, H. L., & Elmagarmid, A. K. (2004). Concept-Oriented Indexing of Video Databases: Toward Semantic Sensitive Retrieval and Browsing. *IEEE Transactions on Image Processing*, 13(7), 974–992. doi:10.1109/TIP.2004.827232
- FIcontent. (2014a). FI content - Driving innovation at the crossroads of content, media, networks & creativity. http://mediafi.org/wp-content/uploads/2014/02/FIcontent_brochure.pdf. Zugegriffen: 10. Juni 2014
- FIcontent. (2014b). Social Connected TV - FIcontent Open Platform. *FIcontent*. <http://mediafi.org/open-platforms/social-connected-tv/>. Zugegriffen: 7. Oktober 2014
- FIcontent. (2014c, September 15). Introduction to the Audio Mining SE. *FIcontent Wiki*. Wiki. <http://wiki.mediafi.org/doku.php/ficontent.socialtv.enabler.audiomining>. Zugegriffen: 15. Oktober 2014
- Google Inc. (2014). YouTube - Statistiken. *Statistiken*. <https://www.youtube.com/yt/press/de/statistics.html>. Zugegriffen: 18. November 2014
- Jaspers, M. W. M. (2009). The Think Aloud Method and User Interface Design. In P. Zaphiris & C. S. Ang (Hrsg.), *Human computer interaction: concepts, methodologies, tools, and applications* (S. 622–628). Hershey, PA: Information Science Reference.
- Kelle, U., & Kluge, S. (2010). *Vom Einzelfall Zum Typus Fallvergleich Und Fallkontrastierung in Der Qualitativen Sozialforschung*. Vs Verlag Fur Sozialwissenschaften.
- Krauss, C., Seeliger, R., Wilson, A., & Arbanowski, S. (2014). Enriched personalized multi-screen content for social connected TV. Gehalten auf der TVX '14 ACM International Conference on Interactive Experiences for TV and Online Video, Newcastle Upon Tyne, United Kingdom. <http://mediafi.org/wp-content/uploads/2014/07/WP-108.pdf>. Zugegriffen: 6. Oktober 2014
- Ling Shao, Jones, S., & Xuelong Li. (2014). Efficient Search and Localization of Human Actions in Video Databases. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(3), 504–512. doi:10.1109/TCSVT.2013.2276700
- Liu, J., Huang, Z., Cai, H., Shen, H. T., Ngo, C. W., & Wang, W. (2013). Near-duplicate video retrieval: Current research and future trends. *ACM Computing Surveys*, 45(4), 1–23. doi:10.1145/2501654.2501658
- Oh, J., & Hua, K. A. (2000). Efficient and Cost-effective Techniques for Browsing and Indexing Large Video Databases. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data* (S. 415–426). New York, NY, USA: ACM. doi:10.1145/342009.335436

- Paul, L., & Wolf, M. (2014). *D21 - Digital - Index 2014*. TNS Infratest. <http://www.initiated21.de/portfolio/d21-digital-index-2014/>. Zugegriffen: 18. November 2014
- Robinson, O. C. (2014). Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide. *Qualitative Research in Psychology*, *11*(1), 25–41.
- TNS Infratest. (2014). *Digitalisierungsbericht 2014 - Daten und Fakten*. http://www.die-medienanstalten.de/fileadmin/Download/Publikationen/Digitalisierungsbericht/2014/Digitalisierungsbbericht_2014_Web.pdf. Zugegriffen: 9. September 2014
- Vranješ, M., Rimac-Drlje, S., & Grgić, K. (2013). Review of objective video quality metrics and performance comparison using different databases. *Signal Processing: Image Communication*, *28*(1), 1–19. doi:10.1016/j.image.2012.10.003
- Young, S. (2008). HMMs and Related Speech Recognition Technologies. In P. J. B. Dr, P. M. M. Sondhi, & P. Y. (Arden) H. Dr (Hrsg.), *Springer Handbook of Speech Processing* (S. 539–558). Springer Berlin Heidelberg. http://link.springer.com/referenceworkentry/10.1007/978-3-540-49127-9_27. Zugegriffen: 21. Oktober 2014

Autoren

Tobias Simon, B.Sc.



Tobias Simon ist Mediengestalter für Digital- und Printmedien und hat im Dezember 2014 sein berufsbegleitendes Studium an der Hochschule Mainz im Studiengang „Medien, IT und Management“ mit dem Bachelor of Science abgeschlossen.

Von 2010 bis 2014 war er Abteilungsleiter und Ausbilder bei der mittelständischen Full-Service-Agentur andré Blickkontakte Werbegesellschaft mbH in Meisenheim. Seit Anfang 2014 ist Herr Simon im Team von Prof. Dr. Sven Pagel für Mediengestaltung und Usability zuständig.

Prof. Dr. Sven Pagel



Sven Pagel lehrt an der Hochschule Mainz Wirtschaftsinformatik und Medienmanagement. Als Studiengangleiter verantwortet er das Bachelor-Programm „Medien, IT und Management“ und ist Akademischer Leiter des Beobachtungslabors. Zudem ist er Lehrbeauftragter der Universität Düsseldorf. Seit 2014 ist er Sprecher der „Fachgruppe Medienökonomie“ der Deutschen Gesellschaft für Publizistik und Kommunikationswissenschaft (DGPK). Sein Forschungsinteresse gilt der digitalen Bewegtbildkommunikation in Internetmedien und der Web-Usability-Forschung.

Dr.-Ing. Joachim Köhler



Joachim Köhler leitet die Abteilung NetMedia im Fraunhofer-Institut Intelligente Analyse- und Informationssysteme IAIS und verantwortet das Geschäftsfeld „Content Technologies and Services“. Die von ihm geleiteten Forschungs- und Entwicklungsarbeiten umfassen Lösungen und Systeme zur Erschließung und zum Management von digitalen Inhalten. Dr. Joachim Köhler koordinierte das EU Projekt LinkedTV, in dem innovative Technologien zur Konvergenz von Web und TV erforscht werden. Darüber hinaus ist er in zahlreichen nationalen und internationalen Medienprojekten aktiv.

Für die Content-Industrie (Verlage, Medienhäuser, TV/Rundfunk) hat er zahlreiche Vorträge und Workshops gehalten. Seine Forschungs- und Interessengebiete liegen

in dem Bereich der Content Technologien (innovative Verfahren zur Erschließung von multimedialen Inhalten, Big Data und

Sprachverarbeitung). Er war maßgeblich am Aufbau der Deutschen Digitalen Bibliothek durch das Fraunhofer IAIS beteiligt.