

DIE ARD MINING-PLATTFORM

KÜNSTLICHE INTELLIGENZ IM PRODUKTIVEN EINSATZ FÜR DIE AUTOMATISIERTE ERSCHLIESSUNG IM MULTIMEDIALEN PRODUKTIONSPROZESS

DIRK MARONI, JOACHIM KÖHLER, JENS FISSELER, SVEN BECKER

Das Thema Künstliche Intelligenz ist der wesentliche Treiber für die Digitalisierung und Automatisierung in Unternehmen – komplette Arbeitsprozesse und Unternehmensmodelle werden mittels KI transformiert. Dies betrifft ganz besonders auch die Medienlandschaft und die Prozesse in den Rundfunkanstalten.

► The topic of artificial intelligence is the main driver for digitization and automation in companies – complete work processes and company models are transformed by means of AI. This is especially true for the media landscape and processes in broadcasting companies.

1. Einleitung und Kontext

Im Zuge der Modernisierung seines Archivsystems beauftragte der WDR das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS bereits im Jahr 2015 mit einer Technologiestudie, um zukünftige Medienarchivsysteme zu bewerten. Die Ergebnisse dieser Studie wurden auf der Produktions- und Technik-Konferenz den Direktorinnen und Direktoren von ARD, ZDF und Deutschlandradio im Mai 2016 vorgestellt. Bereits zu diesem Zeitpunkt lautete die Empfehlung des Fraunhofer IAIS, den Einsatz von automatisierten Verfahren zur Erschließung von Mediendaten zu forcieren und dabei den Schwerpunkt auf maschinelle Verfahren zu legen. Damals war der Begriff Künstliche Intelligenz in diesem Kontext noch nicht gebräuchlich, wenngleich die technischen Inhalte der Studie genau die aktuellen Trends vorhersagten. Der Durchbruch und Fortschritt im Bereich der Künstlichen Intelligenz erfolgte hauptsächlich in mediennahen Bereichen – beispielsweise Bilderkennung, Spracherkennung und Textverstehen. Gerade hier haben Medienunternehmen die idealen Voraussetzungen, um die datenhungrigen Methoden der KI mit Trainingsdaten zu füttern und zum Erfolg zu führen. Der Bedarf ist groß, denn es liegen immer größere Mengen von Mediendateien (Videos, Audios oder Texte) vor, die den Redakteuren und Journalisten rechtzeitig mit den notwendigen Metadaten in Produktionsprozessen zur Verfügung stehen sollen.

Die Methoden der automatisierten Medienerschließung waren bereits Gegenstand zahlreicher Forschungsprojekte in den letzten 20 Jahren. In vielen davon konnte das

Fraunhofer IAIS die Grundlagen dafür schaffen, Forschungsergebnisse in die Anwendung zu bringen und produktiv einsetzbare Methoden zu entwickeln:

- Ergebnis des EU-Projekts Vitalas (2006 bis 2009) war ein Videoretrieval-System, das sowohl Methoden der Logo-Erkennung, Spracherkennung und Objekterkennung beinhaltete und auf Daten des Institut National de l'Audiovisuel (kurz INA) angewendet wurde.
- In Zusammenarbeit mit der BBC und der Deutschen Welle entwickelte das Fraunhofer-Team im EU-Projekt AXES ein multimodales Retrievalsystem.
- Das Bundesministerium für Wirtschaft und Energie BMWi förderte 2007 bis 2011 das Forschungsprojekt Theseus-Contentus, hier wurde eine semantische Multi-mediasuchmaschine entwickelt.
- In dem BMBF-Projekt MediaGrid wurden von 2007 bis 2012 erstmals cloudbasierte Ansätze zur Medienanalyse erprobt und prototypisch aufgebaut.

All diese Projekte haben zum Aufbau von Kompetenzen und Technologien im Bereich der KI-basierten Analyse von Mediendateien beigetragen, die in Rundfunkanstalten real im Einsatz sind: Der Schritt von Forschungsprototypen zu produktiv einsetzbaren Systemen ist wichtig, um die Verbreitung und den tatsächlichen Einsatz der erforschten Technologien sicherzustellen.

Als erstes produktives Ergebnis stellte das Fraunhofer IAIS Medienunternehmen sein Audio-Mining-System bereit. Es transkribiert und segmentiert in hoher Qualität Video- und Audiodateien mittels einer Speech-to-Text-Technologie und ist mittlerweile in der kompletten ARD, bei Deutschlandradio sowie im ZDF im Einsatz – es hat den Sprung von der Forschung hin zu einem produktiven System in die Anwendung geschafft. Die Audio-Mining-Technologie ist allerdings nur ein Baustein im gesamten Prozess der Medienerschließung.

Den technologischen Trends folgend begann der WDR in strategischer Partnerschaft mit dem Fraunhofer IAIS 2017 mit dem Aufbau des Media Data Hubs (MDH), einem Mediendatensystem, welches den zentralen Zugriff auf Daten aus verschiedenen Systemen ermöglicht. Wesentliche Bestandteile des Media Data Hubs sind die durchgängige crossmediale Recherche und eine Mining-Plattform zur automatischen Erschließung von Medieninhalten. Seit 2019 ist der Media Data Hub Teil von medas, dem zentralen ARD Metadaten- und der zentralen ARD Mining-Plattform. Die medas-Komponenten werden gemeinschaftlich von den Entwicklungspartnern hr, SWR und WDR für die SAD Kooperationspartner entwickelt. Mit medas verfolgen die in der Kooperation beteiligten Rundfunkanstalten die

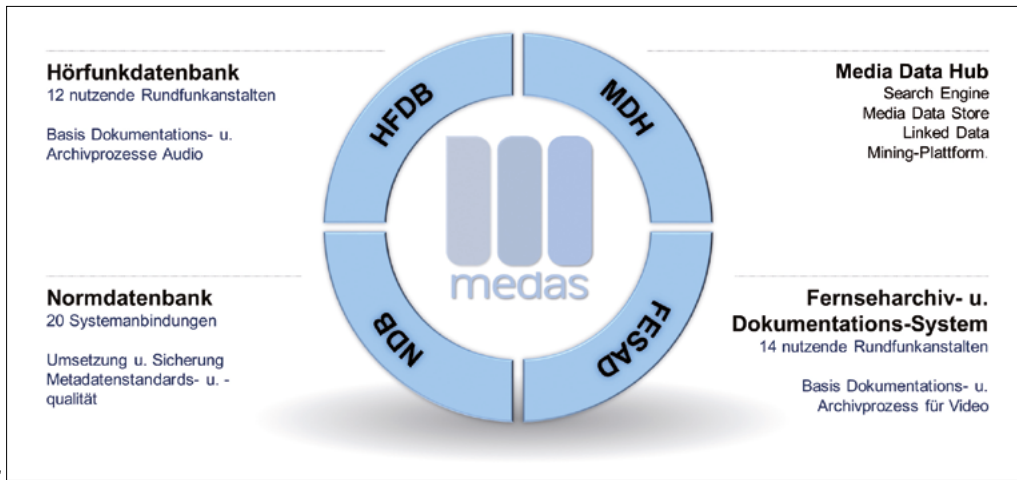


Abbildung 1: Hauptkomponenten von medas

Vision einer übergreifenden Recherche über alle Systeme des Produktionsworkflows sowie die einheitliche automatische Erschließung des Contents aller beteiligten Rundfunkanstalten bereits ab der Phase Produktion. Abbildung 1 zeigt die Hauptsystemkomponenten von medas: die Hörfunkdatenbank (HFDB), die Normdatenbank (NDB), den Media Data Hub (MDH) und das Fernseharchiv- und Dokumentationssystem (FESAD).

Schwerpunkt der Entwicklungen des Fraunhofer IAIS sind aktuell die fortwährende Ergänzung neuer Mining-Services auf der ARD Mining-Plattform und der Aufbau eines automatischen Trainings für die Text-Mining-Services. Die nachfolgenden Abschnitte beschreiben den technischen Aufbau der ARD Mining-Plattform und ihrer Services, deren Eigenschaften und Vorteile sowie Anwendungsbeispiele.

2. Technische Beschreibung der ARD Mining-Plattform und ihrer Services

Die ARD Mining-Plattform ist als modular aufgebautes, Microservice-basiertes System realisiert und beinhaltet zahlreiche Mining-Services, die über eine zentrale Workflow Engine verwaltet werden, siehe Abbildung 2. Die Kommunikation mit der Plattform erfolgt über eine REST-Schnittstelle. Über diese kann die Mining-Plattform mit der Analyse von Mediendateien beauftragt werden, wobei neben der zu verarbeitenden Essenz die Art der durchzuführenden Analyse angegeben werden muss. Die Mediendateien müssen dabei entweder durch die Mining-Plattform heruntergeladen werden oder, falls es sich um einen Text handelt, auch als Teil des Auftrags mitgeschickt werden können.

Die eigentliche Medienanalyse erfolgt durch einzelne sogenannte Mining-Services. Diese können entweder als eigenständige containervirtualisierte Services innerhalb der ARD Mining-Plattform laufen oder in Form externer Services angebunden werden. Über die innerhalb der ARD Mining-Plattform laufenden Mining-Services hat die Plattform eine engere Kontrolle und stellt darüber hinaus Funktionalität zur Verfügung, welche die Implementierung dieser internen Mining-Services erleichtert. Deswegen sind zum Beispiel alle Text-Mining-Services wie Named-Entity-Recognition (NER), Keyword-Extraction (KWE), Semantic-Tagging (SEM, ein Ansatz zur intelligenten Verschlagwortung) und Topic-Modeling als interne Mining-Services implementiert. Manche Verfahren zur Medienanalyse lassen sich aber nur mit großem Aufwand in die Plattform integrieren oder werden sogar bei Drittanbietern gehostet. Diese lassen sich dann als externe Mining-Services an die Plattform anschließen, wie es in Abbildung 2 für das Audio-Mining-System dargestellt ist.

Die einzelnen Schritte zur Analyse einer Mediendatei werden durch eine Workflow-Komponente gesteuert. Anfangs wurde hierfür eine selbstentwickelte Workflow-Komponente verwendet, derzeit erfolgt der Umstieg auf eine BPMN-Workflow-Engine. Diese bringt zahlreiche Funktionalitäten mit, die sonst aufwendig implementiert werden müssten, wie zum Beispiel die Speicherung von Workflow-Zuständen, was die Wiederanlauffähigkeit der ARD Mining-Plattform verbessert.

Abbildung 3 zeigt beispielhaft den Workflow zur Extraktion von Schlüsselwörtern aus einem Text. Zunächst wird die Mediendatei (in diesem Fall der Text) in die ARD Mining-Platt-

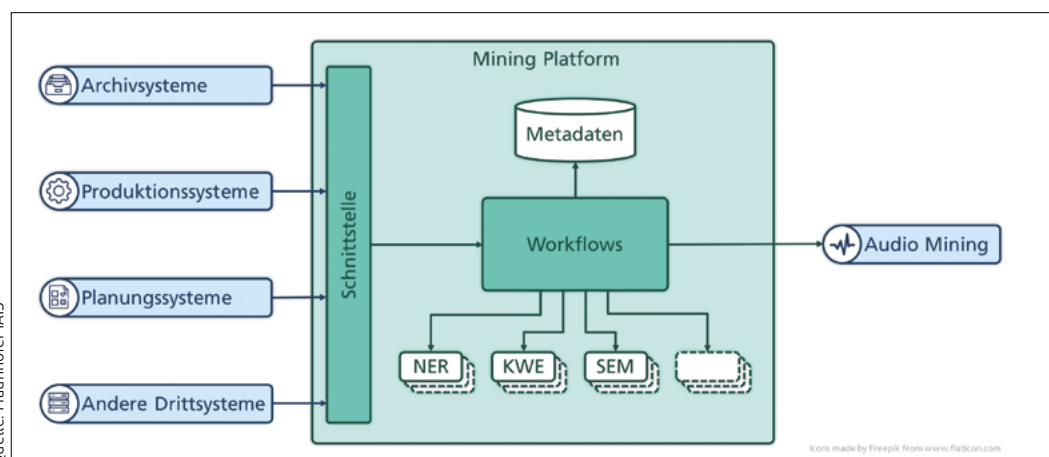
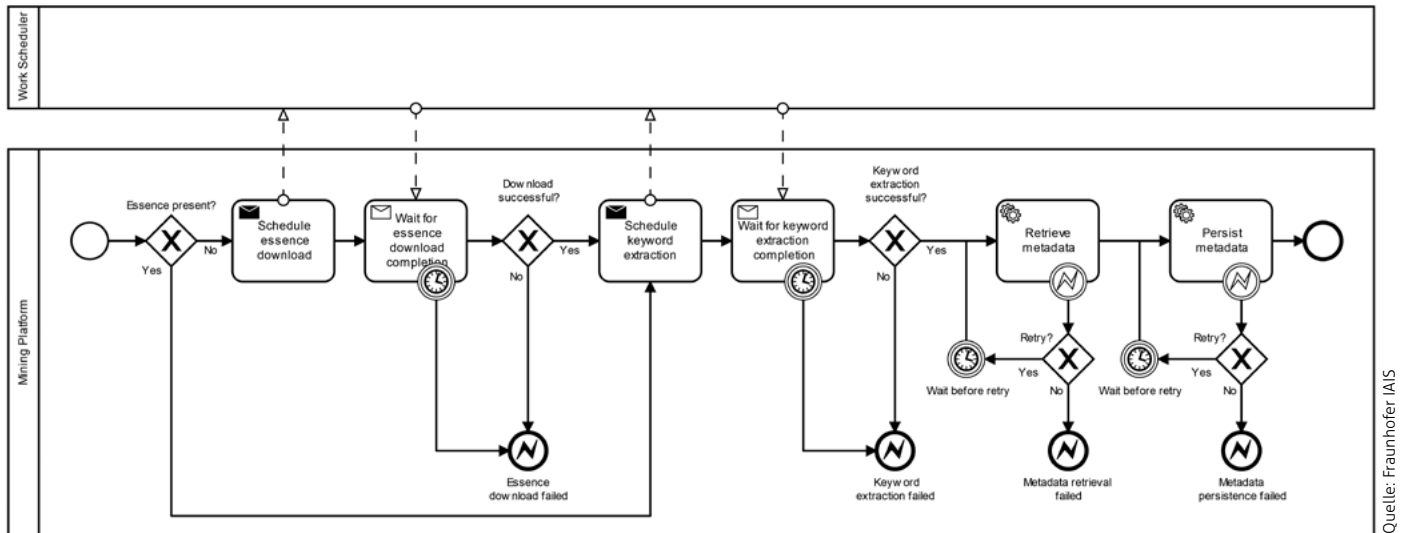


Abbildung 2: Schematischer Aufbau der ARD Mining-Plattform

Quelle: ARD

Quelle: Fraunhofer IAIS



Quelle: Fraunhofer IAI

Abbildung 3: Der Workflow zur Keyword-Extraction

form heruntergeladen, sofern diese nicht bereits vorliegt. Im nächsten Schritt wird die eigentliche Keyword-Extraction angestoßen und anschließend auf das Ergebnis gewartet. Liegt das Ergebnis im Mining-Service vor, so wird es heruntergeladen und anschließend im Metadaten-Speicher abgelegt.

Neben den eigentlichen Verarbeitungsschritten enthält der in Abbildung 3 dargestellte Workflow BPMN-Konstrukte, welche mit Hilfe von Wiederholungen und Zeitbegrenzungen eine robuste und transparente Ablaufsteuerung realisieren.

Der hier vorgestellte Workflow analysiert eine Mediendatei nur mit einem Mining-Service. Ein crossmedialer Ansatz zur Medienanalyse sollte aber die Kombination verschiedener Mining-Technologien zur umfänglichen Erschließung verschiedener Medienarten unterstützen. So könnte zum Beispiel ein Audiotranskript mit Text-Mining-Verfahren weiter erschlossen werden, eine Bildanalyse mit einer Audioanalyse kombiniert werden (zum Beispiel Face-Recognition mit Sprechererkennung) oder verschiedene Bildanalyseverfahren miteinander verknüpft werden (Face-Recognition und Bauchbindenerkennung). Die Kombination verschiedener Mining-Verfahren kann die Erschließungsergebnisse verbessern und ist mit dem verwendeten Workflow-Konzept ohne Weiteres umsetzbar.

3. Eigenschaften und Vorteile der ARD Mining-Plattform

Nachfolgend werden einige Eigenschaften und Vorteile der ARD Mining-Plattform erläutert, insbesondere in Bezug auf vorhandene Cloud-Services.

Trainierbarkeit: Sämtliche KI-Module zum Text-Mining, zur Sprachanalyse und zur Videoerkennung können mit den Daten und Wissensquellen der Rundfunkanstalten trainiert, adaptiert und optimiert werden. So lassen sich beispielsweise Modelle für die Sprechererkennung mittels 5 Minuten Sprachmaterial eines Sprechers trainieren, so dass eine auf die jeweilige Rundfunkanstalt zugeschnittene Sprechererkennung erfolgen kann. Somit lassen sich semantische O-Ton-Suchen realisieren, wie beispielsweise „Was hat Frau Merkel (Sprecherin) über den Brexit (Suchbegriff, Thema) gesagt?“. Die Möglichkeit, eigene Modelle zu erstellen, hat zum einen den Vorteil, dass zum Beispiel Sprecher- oder Gesichter gezielt auf den jeweiligen Anwendungsfall hin von den Rundfunkanstalten selbst erstellt und verwendet werden können.

Zum anderen erlaubt ein eigenständiges Training mit den Daten der Rundfunkanstalten eine bessere und zielgerichtete Erkennungsleistung.

Integrierbarkeit: Die ARD-Mining-Plattform lässt sich von verschiedenen IT-Anwendungen der ARD im multimedialen Produktionsprozess prinzipiell nutzen. Die Anbindung erfolgt über REST-Schnittstellen. Im Archivumfeld sind bereits FESAD und HFDB an die ARD Mining-Plattform angeschlossen. Die Schnittstellen sind in einem API-Dokument ausführlich beschrieben. Ziel ist es, die Mining-Nutzung möglichst an den Anfang der multimedialen Produktionskette zu verlagern, um schon zu Produktionsbeginn hilfreiche Metadaten automatisiert zu generieren.

Private Cloud durch Betrieb im IVZ: Die ARD Mining-Plattform wird im Informationsverarbeitungszentrum (IVZ) von ARD und Deutschlandradio als Managed Service betrieben und steht somit sämtlichen Rundfunkanstalten der Kooperation zur Verfügung. Dieses Set-up ist daher eher als On-Premises bzw. Private Cloud zu bezeichnen. Mit dieser Konfiguration ergeben sich folgende Vorteile:

- **Datensicherheit:** Es wird gewährleistet, dass die ARD-Inhalte nicht die IT-Infrastrukturen der ARD verlassen und somit ein DSGVO-konformer Betrieb möglich ist.
- **Anpassbarkeit:** Die Module der ARD Mining-Plattform lassen sich auf die Anforderungen der ARD optimieren und das für die KI-Module notwendige Training der Modelle mit ARD-spezifischen Daten durchführen.
- **Geringer Datentransfer:** Die audiovisuellen Daten verbleiben in der IT-Infrastruktur der ARD, so dass ein aufwendiger Datentransfer auf ein oder mehrere externe Cloudplattformen vermieden wird.

Grundsätzlich lassen sich für Spezialanforderungen bei der automatischen Erschließung (zum Beispiel maschinelle Übersetzung für arabische Texte) Cloud-Dienste von verschiedenen Anbietern wie AWS, Microsoft oder Google in die Mining-Plattform integrieren und nutzen. Dies erfordert sowohl Erweiterungen bei der technischen Architektur als auch im Betriebskonzept.

4. Nutzung und Anwendungsfälle für die Dienste der ARD Mining-Plattform

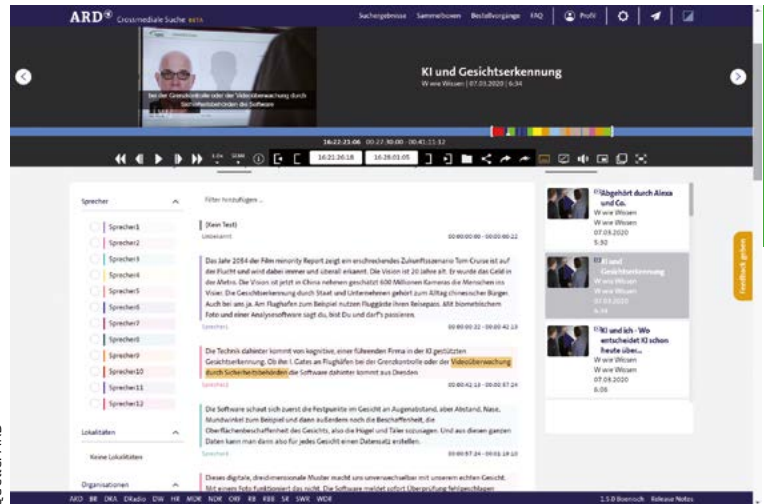
Die ARD Mining-Plattform ist Bestandteil des medas-Systems und wird somit zunächst im Kontext der crossmedialen Suchanwendung genutzt. Die medas-Entwicklung zu einem

crossmedialen Mediendatensystem erfolgt im Rahmen der Kooperation von ARD und SAD, so dass eine weite Verbreitung der Mining-Plattform vorgesehen ist. Zentraler Einstieg von medas ist eine übergreifende crossmediale Suche, die alle Audio- und Videobeiträge sowie Presstexte in den digitalen Archiven der Rundfunkanstalten durchforstet. Software mit Künstlicher Intelligenz ermöglicht die Anreicherung der audiovisuellen Inhalte mit automatisch generierten Informationen. Neue Beiträge werden sofort durch durchsuchbare Metadaten ergänzt, die Themen und Inhalte sind so schneller auffindbar. Journalisten finden zeitnah Zitate und Stichworte in den Beiträgen, denn das Mining-System funktioniert wie eine Suchmaschine für audiovisuelle Inhalte, die die Tonspur als Text durchsuchbar macht, Sprecher erkennt und zukünftig auch Gesichter identifiziert. In Abbildung 4 ist nachfolgend die automatisch generierte Transkription in der crossmedialen Suchanwendung der ARD zu sehen.

Die Inbetriebnahme des vorgestellten Systems wurde zunächst für die vollautomatische Erschließung des im Produktionsbetrieb der ARD tagtäglich zu archivierenden Medienmaterials konzipiert. Dies umfasst die produzierten Hörfunk- und Fernsehbeiträge sowie die im Produktionsprozess entstandenen und als archivierungswürdig befundenen Begleitmaterialien und die von den Landesrundfunkanstalten erworbenen Nachrichtenmeldungen und Presseartikel. In der aktuellen Ausbaustufe kann das System täglich 2000 Stunden Audio- und Videomaterial und 100.000 Textdokumente automatisch erschließen.

In einem weiteren Schritt sollen sämtliche in den Archiven digital enthaltenen Materialien mit den verfügbaren Mining-Verfahren für Video-, Audio- und Text-Mining erschlossen werden, so dass die crossmediale Recherche auf die vollautomatisch generierten Daten über den gesamten Archivbestand zugreifen kann. Wurde der Vorgang der massenweisen Digitalisierung von Tonbändern, Filmrollen und Magnetbändern als Retrodigitalisierung bezeichnet, so nennen wir dieses Vorhaben „Retro-Mining“.

Die automatischen Erschließungsverfahren können jedoch nicht nur die Recherche über Archive aufwerten, sie sollen ebenfalls für die bessere Handhabung von Rohmaterial eingesetzt werden. So können zum Beispiel umfangreiche Video- und Audio-Rohdaten schneller gesichtet werden, gezielt O-Töne von ausgewählten Sprechern identifiziert oder Informationsbezüge sowie -differenzen zwischen unterschiedlichen Datenquellen aufgespürt werden. Diese Beispiele sind inspiriert durch bislang identifizierte Bedarfe



Quelle: ARD

Abbildung 4: Automatische Transkription in der ARD Crossmedialen Suche von medas

von interessierten Nutzergruppen. Zahlreiche weitere Anwendungsfälle sind denkbar und können nach Bedarf maßgeschneidert umgesetzt werden.

Künftig können auch weitere Datenquellen in die vollautomatische Erschließung und die crossmediale Recherche eingebunden werden, zum Beispiel Bilddatenbanken, Online-Archive oder andere Online-Plattformen und Social Media. Mit jeder Erweiterung des angebundnen Datenbestands wird die Reichweite der zentralen Informationsrecherche verbessert.

5. Zusammenfassung und Ausblick

Die ARD Mining-Plattform in medas bietet aktuell schon zahlreiche Dienste zum Text-Mining und zur Sprachanalyse (zum Beispiel Audio-Mining) und ist bereits im produktiven Einsatz für die inhaltliche Erschließung von audiovisuellen Daten. In der weiteren Entwicklung geht es darum, weitere Dienste zur automatischen Erschließung von AV-Inhalten in die Mining-Plattform zu integrieren. Besonders im Bereich Video-Mining sollen vorhandene Technologien genutzt und neue Verfahren implementiert werden. Dabei werden Synergien zu dem aktuell vom Bundesministerium für Wirtschaft und Energie BMWi geförderten Projekt AI4MediaData genutzt: Dort erforscht das Fraunhofer-Team gemeinsam mit Partnern, wie Wissen über Medieninhalte

Quelle: WDR




DR. DIRK MARONI

ist seit 2016 Leiter Informationsmanagement im Westdeutschen Rundfunk

➤ www.wdr.de

Quelle: Fraunhofer IAIS



DR. JOACHIM KÖHLER

ist seit 1999 wissenschaftlicher Mitarbeiter und seit 2002 Leiter der Abteilung „NetMedia“ am Fraunhofer IAIS in Sankt Augustin

➤ www.iais.fraunhofer.de

Quelle: Fraunhofer IAIS




DR. JENS FISSELER

ist seit 2018 Leiter der Fachgruppe Software & Content Analysis am Fraunhofer IAIS in Sankt Augustin

➤ www.iais.fraunhofer.de

Quelle: Fraunhofer IAIS



DIPL.-ING. (FH) SVEN BECKER

ist seit 1999 wissenschaftlicher Mitarbeiter am Fraunhofer IAIS in Sankt Augustin und leitet seit 2014 die Projekte der Kooperation von WDR und Fraunhofer IAIS

➤ www.iais.fraunhofer.de

FORSCHUNG

und deren Nutzung generiert und Metadaten mit Nutzungsdaten verknüpft werden können. (<https://ai4mediadata.com/>).

Für die Weiterentwicklung und Nutzungs-Optimierung der Mining-Dienste ist eine enge Abstimmung des technischen Entwicklungsteams mit Redakteuren und Dokumentaren notwendig. Durch die stetigen Fortschritte im Bereich des maschinellen Lernens gilt es, diese neuen Technologien nach einer vorhergehenden Evaluierung möglichst effizient in die ARD Mining-Plattform zu integrieren. Die Identifikation, Erprobung, Anpassung und Integration von KI-basierten

Technologien und Machine-Learning-Verfahren ist die Aufgabe des Fraunhofer IAIS in Zusammenarbeit mit Experten der ARD. Das Institut, das zu den führenden Wissenschaftsinstituten in den Gebieten Künstliche Intelligenz, Maschinelles Lernen und Big Data in Deutschland und Europa zählt, legt hierbei großen Wert auf die sorgfältige Auswahl verfügbarer KI Technologien (bevorzugt als Open-Source-Implementierung), der Optimierung der KI-Verfahren und der Verwendung speziell erstellter Modelle (trainiert auf Daten des Kunden oder auf sorgfältig ausgewählten Daten, die am besten für den Anwendungsfall geeignet sind), um für jeden Bedarf die bestmöglichen Ergebnisse erzielen zu können. ➔

MEDIEN-MINING MIT UNTERSCHIEDLICHEM REIFEGRAD

So vielfältig der Einsatz von Künstlicher Intelligenz (KI) in der Medienproduktion ist, so unterschiedlich gestaltet sich auch der Reifegrad der einzelnen Verfahren. Welche Rolle Trainingsdaten bei Text-Mining, Audio-Mining und Video-Mining spielen und was für die ARD Mining-Plattform zukünftig geplant ist, erläutern Dr. Dirk Maroni vom WDR und Dr. Joachim Köhler vom Fraunhofer IAIS im Gespräch mit FKT.

FKT: Warum ist es notwendig, für die Erschließung von Mediendaten automatische Verfahren einzusetzen?

Dr. Maroni: Die automatische Erschließung von Mediendaten ermöglicht schnellere Prozesse in Redaktion und Produktion sowie eine Bewältigung der deutlich gestiegenen Materialmengen in den Archiven. Insbesondere im Newsroom sind die Anforderungen an die Geschwindigkeit der Prozesse durch die Digitalisierung und die damit größere Konkurrenz – Google, Twitter & Co. – deutlich gestiegen.

FKT: Hier kommt Künstliche Intelligenz (KI) ins Spiel. Welche Methoden gibt es, um Mediendaten automatisch zu erschließen?

Dr. Köhler: Das hängt vom Medium ab. Bei Texten stehen Verschlagwortung, semantische Beschreibung und die Definition von Themen-Klassen im Zentrum. Im Audio-Bereich ist das Ziel, gesprochenes Wort in Text umzuwandeln (Speech-to-text) und den Sprecher zu erkennen. Die automatische Schnitt-Generierung steht wiederum im Video-Bereich im Mittelpunkt, also die Erstellung von Schlüsselszenen (key scenes). Hier geht es aber auch darum, Personen und Objekte zu erkennen. Das sind die typischen Verfahren des maschinellen Lernens.

Dr. Maroni: Die drei Verfahren – Text-Mining, Audio-Mining und Video-Mining – haben dabei unterschiedliche Rei-

fegrade. Text-Mining und Audio-Mining können wir in der Praxis sehr gut einsetzen. Video-Mining-Verfahren sind differenzierter zu betrachten: Bei der Gesichtserkennung sind wir sehr weit. Verfahren für die automatische Beschreibung von Konzepten in Videos haben heute aber nicht den Reifegrad, um sie in einem produktiven System zu verwenden. Interessanter Aspekt dabei: Sowohl Video-Mining- als auch Audio-Mining-Verfahren können nachgelagert mit Text-Mining-Verfahren bearbeitet und somit verbessert werden.

FKT: Warum gibt es diese unterschiedlichen Reifegrade?

Dr. Köhler: Bei Text-Mining und Audio-Mining existieren relativ viele annotierte Trainingsdaten für das Lernen dieser Modelle. Beim Video-Mining sind die Komplexität des Trainings und die Datenmengen wesentlich größer. Hier benötigen wir viele annotierte Trainingsdaten, die gibt es bei der Videoerkennung nur für Objekte. So genannte Objekt-Erkennen liefern sehr gute Ergebnisse. Geht es aber um Gesichtserkennung, dann muss das System sehr aufwendig trainiert werden.

Dr. Maroni: In der Tat, Trainingsmaterial für Videos zu extrahieren und bereitzustellen, ist extrem aufwendig. Bei der Gesichtserkennung benötigen wir Daten aus allen möglichen Positionen eines Gesichtes – und dies auch über einen gewissen Zeitraum hinweg, weil sich Menschen verändern. Das Ergebnis der Videoanalyse lässt sich mit der Kombination verschiedener KI-Services verbessern – etwa durch die zusätzliche Auswertung von Bauchbinden und durch die Erkennung der Sprecherstimme.

FKT: Aktuell können Sie pro Tag 2.000 Stunden Audio- und Videomaterial und 100.000 Textdokumente mit der ARD Mining-Plattform automatisch erschließen. Ist da noch Luft nach oben?

Dr. Maroni: Ja. Lassen Sie mich dies am Beispiel Audio-Mining erklären: Bei Speech-to-Text haben wir 66 pa-

Quelle: Fraunhofer IAIS

**DR. JOACHIM KÖHLER**

Leiter der Abteilung „NetMedia“ am Fraunhofer IAIS

Quelle: WDR

**DR. DIRK MARONI**

Leiter Informationsmanagement im WDR

parallele Kanäle, die so genannten „worker“. Diese 66 Audios lassen sich parallel bearbeiten und die Anzahl der „worker“ können beliebig skaliert werden. Das ist nur eine Frage von Hardware. Das gleiche gilt auch für Text-Mining-Services, die wir in ihrer Mikroservice-Architektur beliebig skalieren können. Im Moment geschieht dies noch manuell, in Zukunft dynamisch.

Dr. Köhler: Die Plattform ist auch erweiterbar hinsichtlich neuer Methoden, die jetzt noch in der Erprobung sind und künftig produktiv verwendet werden können. Da wird sich noch einiges tun.

Dr. Maroni: Wir arbeiten auch daran, die Mining-Services in der Qualität zu verbessern und die Durchlaufzeiten zu verkürzen. Das ist bei der Analyse von Archivmaterial nicht so entscheidend, aber insbesondere im News-Bereich müssen die Verfahren wesentlich schneller sein.

FKT: Mit der Plattform unterstützen Sie die Arbeit der Redakteure und Archivare. Wie sehen typische Anwendungsszenarien aus?

Dr. Köhler: Mit den Mining-Verfahren ist eine viel schnellere und viel tiefere Erschließung von Inhalten möglich. Bisher war es im Hörfunk etwa so: Ein Beitrag wurde abgehört, eine Zusammenfassung geschrieben und bestimmte Zeitmarken gesetzt. Das hat gedauert. Durch Audio-Mining können Redakteure nun direkt über eine Wortsuche exakt an die gewünschte Position springen. Wie Journalisten das im Detail nutzen, hängt aber nicht zuletzt von der Suchanfrage ab. Von den Dokumentaren erfahren wir, dass sich die Art und Weise der Suche ändert – wegen der nun vorhandenen vielen Metadaten. Ein weiterer Vorteil bei der Recherche sind neue Empfehlungsfunktionen, die auf ähnliche Beiträge verweisen.

Dr. Maroni: Ein weiteres Beispiel aus dem Bereich TV-Redaktion/Produktion: Bereits auf der Rückfahrt nach einem Außendreh schickt der Reporter das Material zum Sender. Dort läuft es automatisch durch die Mining-Plattform, mit Speech-to-Text und bestenfalls auch Gesichtserkennung. Ist der Autor dann im Sender, liegt das Material bereits auf-

bereitet vor: Wo hat der Interviewpartner was gesagt? Über die Textsuche kann der Reporter einen bestimmten Bereich markieren, das Material direkt abklammern, schneiden, bearbeiten und automatisch untertiteln. Dafür kann er wieder KI einsetzen.

Wichtig aber ist auch: Die Mining-Plattform kann nicht alles leisten, sie muss auch immer noch in andere Workflows integriert werden. Wir arbeiten an Schnittstellen zu VPMS-Systemen, zu Dira, zu Untertitelungs-Tools. Der Media Data Hub ist ein wichtiges Element in dieser Kette, aber nicht das einzige.

FKT: Aktueller Schwerpunkt der Forschungsarbeit bei Fraunhofer ist es, die Mining-Services auch fortwährend zu ergänzen. Herr Dr. Köhler, woran arbeitet Ihr Institut im Moment?

Dr. Köhler: Beim Text-Mining ist ein neuer Service in Vorbereitung, der sich mit Disambiguierung bzw. Entitäten-Linking befasst. Wenn eine Entität – ein Name zum Beispiel – erkannt wird, dann muss auch der richtige Link auf externe Quellen gesetzt werden. Das ist bei vielen Entitäten gar nicht so leicht, weil es Mehrdeutigkeiten gibt. Ein anderer neuer Dienst im Audio-Bereich ist das Erkennen von Sprachen. Wir haben häufig die Situation, dass gerade in internationalen Beiträgen fremdsprachige O-Töne eingespielt werden. Diese lassen sich mit deutschen Modellen natürlich nicht erkennen.

Im Video-Mining ist – wie erwähnt – der Reifegrad noch nicht so ausgeprägt. Hier arbeiten wir weiter an einer Reihe von Verfahren wie Gesichtserkennung, Objekterkennung und Konzepterkennung. Zudem haben wir kürzlich das Projekt AI4MediaData gestartet. In den nächsten drei Jahren werden wir hier gezielt Mining-Verfahren für Videomaterial entwickeln.

Herr Dr. Maroni, Herr Dr. Köhler,
vielen Dank für das Gespräch.

Interview: Martin Braun