

Audio Mining für die Geistes- und Kulturwissenschaften: Nutzungsszenarien und Herausforderungen

Köhler, Joachim

joachim.koehler@iais.fraunhofer.de
Fraunhofer IAIS

Leh, Almut

almut.leh@fernuni-hagen.de
FernUniversität in Hagen

Himmelman, Nikolaus

sprachwissenschaft@uni-koeln.de
Universität zu Köln, Deutschland

Rau, Felix

f.rau@uni-koeln.de
Universität zu Köln, Deutschland

Workshopleiterinnen und -leiter

- Nikolaus P. Himmelman, Universität zu Köln,
- Joachim Köhler, Fraunhofer IAIS,
- Almut Leh, FernUniversität in Hagen,
- Felix Rau, Universität zu Köln,

Ort

Der Workshop findet in Raum S16 des Seminargebäudes (Gebäude 106) der Universität zu Köln statt.

Programm

Zum Workshop gibt es Beiträge, die primär aus Entwicklerperspektive das Thema angehen, und solche, die eher eine Anwenderperspektive einnehmen, wobei das keine scharfe Trennung ist.

Aus Entwicklerperspektive berichten Alexandre Arkhipov vom Hamburger Zentrum für Sprachkorpora, Christoph Draxler vom Bayerisches Schallarchiv (LMU), Jens Gorisch vom IDS Mannheim, Joachim Köhler vom Fraunhofer IAIS sowie Burkhard Meyer-Sickendiek und Hussein Hussein von der FU Berlin über Audiomining Werkzeuge, die bei Ihnen entwickelt wurden und werden.

Aus der Anwenderperspektive stellen Thomas Beckers vom WDR, Thorsten Dresing von der Firma *audiotranskription* (Marburg), Almut Leh von der FernUniversität Hagen sowie Anna-Maria Götz, Annabelle Petschow & Ruth Rosenberger von der Stiftung Haus der Geschichte der Bundesrepublik Deutschland (Bonn) Herausforderungen für die automatische Spracherkennung in unterschiedlichen Anwendungsbereichen dar.

Ein detailliertes Programm ist Ende Januar über die Webseite des KA3 Projekts zugänglich. Im Anschluss an die Beiträge gibt es jeweils Gelegenheit zu Fragen und kurzer Diskussion. Am Schluss des Workshops gibt es eine kurze Abschlussdiskussion. Workshopssprache ist Deutsch, Beiträge können aber auch auf Englisch präsentiert werden.

Weitere Details zum Thema

Moderne intelligente Analyse- und Auswertungsmethoden für Audiodaten ermöglichen effektivere Arbeitsweisen und neue Fragestellungen in den Geistes- und Kulturwissenschaften. Dieser Workshop soll Fortschritte, Herausforderungen und Perspektiven im Audio Mining präsentieren und Forscherinnen und Forscher aus dem Bereich der Sprachtechnologie mit (potentiellen) Nutzerinnen und Nutzern aus den Geistes- und Kulturwissenschaften zusammenbringen.

Audiovisuelle (multimodale) Primärdaten spielen eine zunehmend größer werdende Rolle in den Geistes- und Kulturwissenschaften. Einfacher und kostengünstiger werdende Aufnahme- und Speicherungstechniken für Audio- und Videodaten ermöglichen die Erhebung von immer umfangreicheren Datensets. Die wissenschaftliche Analyse und Auswertung dieser Daten basiert aber weiterhin überwiegend auf Transkripten, also konvertiert ins schriftsprachliche Medium. Dabei gehen wesentliche Merkmale dieses Datentyps (Sprechmelodie, Stimmqualitäten, Gestik, Mimik etc.) verloren. Selbst wenn die Primärdaten als audiovisuelle Daten analysiert werden geschieht dies meist durch arbeitsintensive manuelle Segmentierung und Annotation.

Von zunehmender Bedeutung ist auch die Nachnutzung audiovisueller Daten, namentlich qualitativer Interviews, wie sie in fast allen Geistes- und Kulturwissenschaften eine Rolle spielen. AV-Daten stellen schon allein aufgrund ihres großen Umfangs besondere Herausforderungen dar für den flexiblen und systematischen Zugriff wie auch eine zweckmäßige langfristige Speicherung. Es bedarf einer Arbeitsumgebung, in der flexibel online auf den gesamten Datensatz zugegriffen werden kann und je nach Bedarf Unterkorpora definiert werden können. Für die einschlägigen Archive, Medienzentren und Dokumentationsstellen ermöglicht das Audio Mining eine nutzerspezifische Recherche in der Gesamtdatenmenge bei direktem Zugriff auf das Audiosignal und sekundengenaue Trefferanzeige. Dadurch wird die Ermittlung und

Bereitstellung relevanter Daten für Sekundäranalysen erheblich verbessert.

Der Einsatz von intelligenten Analyse- und Auswertungsmethoden für multimodale Daten gewinnt vor diesem Hintergrund in den letzten Jahren stark an Bedeutung. Fortschritte im Bereich des maschinellen Lernens ermöglichen stärker automatisierte Arten der Datenanalyse und führen zu einer weniger zeit- und arbeitsintensiven Aufbereitung. Diese Analysemethoden operieren darüber hinaus auf den audiovisuellen Primärdaten und nicht auf textuellen Derivaten.

Im Vergleich zu Textdaten sind die Möglichkeiten der Analyse und Auswertung von audiovisuellen Daten noch weniger verbreitet. Besonders bei der Sprachanalyse von Audio- und Videoaufzeichnungen gibt es aber in jüngster Zeit bemerkenswerte Fortschritte. So lassen sich inzwischen umfangreiche Sprachkorpora automatisch analysieren. Die Sprachaufnahmen können mittels statistischer Verfahren in kleinere Segmente unterteilt werden, Sprecher erkannt und Sprache in Text umgewandelt werden. Diese Techniken werden unter dem Begriff Audio Mining zusammengefasst und ermöglichen den exakten Zugriff auf einzelnen Begriffe, Abschnitte und Ereignisse in Audiodaten.

Technologien und Anwendungen des Audio Minings sind daher für verschiedenste Bereiche der Geistes- und Kulturwissenschaften von großem Interesse. So lassen sich zum Beispiel Sprachaufnahmen automatisch vorsegmentieren, so dass diese anschließend mit gängigen Tools (z.B. ELAN und EXMARaLDA) weiter verarbeitet werden können. Diese Anwendung von Audio Mining ist besonders für die Forschung in Linguistik und Gesprächsanalyse sowie anderen Bereichen, die sich mit Interaktion und Sprache beschäftigen, interessant. Für Oral History, Ethnologie und andere vor allem an inhaltlichen Aspekten der Sprachdaten interessierte Fachgebiete lassen sich die hier typischerweise sehr langen Interviews automatisch segmentieren und transkribieren. Biographische, narrative Interviews sind das Produkt eines kommunikativen Geschehens und damit hoch komplexe Quellen mit vielfältigen Ansatzpunkten für die Interpretation. Aus forschungspraktischen Gründen wird die Analyse von Interviewdaten bisher meist auf das schriftsprachliche Medium reduziert. Das Audio Mining eröffnet hier ganz neue Dimensionen und Fragestellungen. Während die Fallzahlen bei konventionellen Analysemethoden meist bei um die 30 Interviews liegen, können mit technischer Unterstützung viel größere Fallzahlen bearbeitet und somit unter vielfältigen vergleichenden Fragestellungen auch quantitativ ausgewertet werden. Gleichzeitig bieten die Werkzeuge auch der qualitativen Analyse neue Dimensionen, indem sowohl sprachliche wie nicht-sprachliche Aspekte der Kommunikation differenziert erfasst und somit für Forschungsfragen zugänglich gemacht werden können. Um diese Potentiale realisieren zu können, besteht allerdings Forschungsbedarf.

Für die Sprachtechnologien stellen die Audiodaten vieler Fachrichtungen der Digital Humanities interessante Herausforderungen dar. Denn während die Spracherkennung bei Sprachaufnahmen aus dem Nachrichtenbereich (optimale Aufnahmebedingungen und Aufnahmetechnik, artikulierte Hochsprache professioneller Sprecher) inzwischen gute Ergebnisse liefert, stellen zum Beispiel Zeitzeugeninterviews noch eine erhebliche Herausforderung dar. Ursachen sind die oft schlechte Qualität der Audiodaten bedingt durch Modalitäten der Aufzeichnung (Rauschen, Übersteuerung, geringe Lautstärke durch schlechte Platzierung des Mikrophons), Alterungsprozesse der Magnetbänder bis zur Digitalisierung und Fehlentscheidungen bei der Wahl des Audioformats sowie die Charakteristiken der Spontansprache (undeutliche Aussprache, schnelles Sprechen und Dialektfärbung). Bei den Aufnahmen aus Linguistik und Interaktionsforschung handelt es sich oft um Daten aus wenig erforschten Sprachen, zu denen es kaum Ressourcen und wenig andere linguistische Information gibt. Aber selbst Daten aus gut erforschten Sprachen wie dem Deutschen können für die Analyseverfahren anspruchsvoll sein. Datensets bestehen in beiden Fällen häufig aus natürlicher, spontaner Sprache mit mehreren Gesprächsteilnehmern und entsprechenden Sprechüberlappungen. Werkzeuge und Verfahren, die an Daten aus geplanter Sprache trainiert wurden, stoßen bei diesen Aufnahmen schnell an ihre Grenzen. Da diese Daten nicht unter Studiobedingungen erhoben werden können, ist darüber hinaus die Aufnahmequalität zumeist deutlich geringer als bei Aufnahmen aus Hörfunk und Fernsehen. Die im Vergleich zu Daten auf dem Rundfunk recht kleinen Datensets erschweren dabei die Anwendung von intelligenten Analyse- und Auswertungsverfahren noch weiter. So bestehen eine Vielzahl von Forschungsthemen im Bereich der Sprachanalyse, wie beispielsweise die Detektion von überlappenden Sprachsegmenten, eine robuste Sprechersegmentierung von kurzen Dialogsequenzen, Diarisierung von Aufnahmen für die Analyse von Turn-Takings, die Erkennung von Sprechern, Erkennung von Dialekten, robuste Transkription von Sprachdaten hinsichtlich Hintergrundgeräusche und Raumhall, Code-Switching in gesprochener Sprache, Erkennung von Pausen und gegebenenfalls Satzzeichen.

Weitere Informationen zu den Ausrichtern des Workshops

Der Workshop wird von dem an der Universität zu Köln angesiedelten BMBF-Zentrum »Kölner Zentrum Analyse und Archivierung von AV-Daten« (KA³) ausgerichtet. Ziele des über drei Jahre geförderten Projektes sind die Erforschung und Entwicklung von Werkzeugen und Services zur akustischen Analyse von AV-Daten, die Einrichtung einer Nutzerplattform für

die Analyse und Archivierung von AV-Daten sowie die Erprobung und Anwendung der Methoden in ausgewählten Pilotprojekten im Bereich der Oral History/Biografieforschung und der Linguistik/Interkulturelle Kommunikation. Besondere Aufmerksamkeit gilt den miteinander zusammenhängenden Problemen der interaktionsbezogenen Strukturierung und der effizienten Bereitstellung und Archivierung von audiovisuellen Daten. Im Rahmen des Verbundprojektes wird ein fach- und standortübergreifendes Zentrum für die Analyse und Archivierung audiovisueller Daten mit den drei Komponenten Analyse, Archivierung/Publikation und Schulung/Beratung aufgebaut, das die im Projekt entwickelten Werkzeuge und Services interessierten Forscherinnen und Forschern zur Verfügung stellt. Das geplante Zentrum ist institutionell eingegliedert in das umfassender angelegte Kölner Data Center for the Humanities (DCH), das informationstechnologische Unterstützung für alle Datentypen in den Geisteswissenschaften anbietet und erforscht. Das Projekt KA³ ist eine Kooperation des Instituts für Linguistik (N. P. Himmelmann), dem Regionalen Rechenzentrum (U. Lang) und dem Data Center for the Humanities (A. Witt und B. Mathiak) der Universität zu Köln, dem Fraunhofer# Institut für Intelligente Analyse# und Informationssysteme, Sankt Augustin (J. Köhler), dem Institut für Geschichte und Biographie der FernUniversität Hagen (A. Leh) und dem Max-Planck-Institut für Psycholinguistik (P. Trilsbeek).

Nikolaus P. Himmelmann lehrt Allgemeine Sprachwissenschaft an der Universität zu Köln. Zentrales Forschungsgebiet ist sprachliche Diversität und was diese uns über menschliche Kognition und Gesellschaftlichkeit lehrt. Er hat entscheidend zur Entwicklung des Konzepts digitaler Sprachdokumentationen und Spracharchive beigetragen.

Joachim Köhler leitet den Bereich Spracherkennung am Fraunhofer Institut IAIS. Zentrale Arbeitsgebiete sind maschinelles Lernen, Mustererkennung, Deep Learning, Information Retrieval, Medieninformationssysteme, Metadaten sowie Linked Data.

Almuth Leh leitet das Archiv ‚Deutsches Gedächtnis‘ an der FernUniversität Hagen und forscht zu deutscher Mentalitätsgeschichte im 20. Jahrhundert mit Einzelarbeiten zum Naturschutz in Nordrhein-Westfalen, Gewerkschafterinnen, Soldaten im Zweiten Weltkrieg, Wehrmachtsjustiz, Universitätsgeschichte. Weitere Interessensgebiete sind forschungsethische und methodische Fragen der Oral History und die Archivierung lebensgeschichtlicher Interviews.

Felix Rau ist Projektmitarbeiter im Projekt »Kölner Zentrum für Analyse und Archivierung audiovisueller Daten« (KA³) und koordiniert die Fachspezifische Arbeitsgruppe »Linguistische Feldforschung, Ethnologie, Sprachtypologie« (Leitung: Himmelmann) in CLARIN-D. Er ist Feldlinguist mit einem Schwerpunkt auf den Munda-Zweig der austroasiatischen Sprachen und ist ein Archivmanager am Language Archive Cologne.