

Themenvorschläge für BA Arbeiten Phonetik - Medieninformatik

Christoph Draxler

Institut für Phonetik und Sprachverarbeitung

LMU München

`draxler@phonetik.uni-muenchen.de`

30.04.2015

Die folgenden Themenvorschläge für BA-Abschlussarbeiten sollen von Studenten der Medieninformatik oder Phonetik und Sprachverarbeitung bearbeitet werden. Die fachliche Betreuung der Arbeiten erfolgt durch PD Dr. Christoph Draxler am Institut für Phonetik und Sprachverarbeitung, formal läuft die Arbeit als BA-Abschlussarbeit für die Medieninformatik bzw. Phonetik (je nach Studiengang des Studenten).

Visualisierung der Metadaten von Fotos

Digitale Fotos und ihre Dateien enthalten eine Vielzahl unterschiedlicher Daten: Bilddaten, Verarbeitungsschritte, inhaltliche und technische Metadaten. Häufig werden diese Daten bereits genutzt, z.B. geografische Daten für die örtliche Zuordnung, Autorenangaben für die Abrechnung, usw. – das sind naheliegende Anwendungen. Was aber kann man mit den erhobenen Daten noch anfangen? Kann man sie z.B. zum Verbessern der eigenen Fotos heranziehen, kann man die Zusammenstellung seiner Ausrüstung optimieren?

In dieser BA-Arbeit sollen die technisch-inhaltlichen Metadaten digitaler Fotos, also z.B. Brennweite, Belichtungszeit, ISO-Empfindlichkeit usw. von Fotobibliotheken visualisiert und in Bezug zu Bewertungskriterien gesetzt werden. Diese Visualisierungen sollen dem Fotografen helfen, verborgene Gemeinsamkeiten seiner Fotos zu erkennen: sind meine Nachtfotos wirklich die besten Bilder? Warum werden meine Telefotos überwiegend schlecht beurteilt? Habe ich einen typischen Fotostil und kann man erkennen, welches meine Bilder sind?

Kontakt: `draxler@phonetik.uni-muenchen.de`, Tel: 2180-2807

Ein Bankautomat für Analphabeten

In vielen Ländern Afrikas ist die Analphabetenrate immer noch sehr hoch. Gleichwohl haben die Menschen Bedarf an Dienstleistungen wie z.B. Bankgeschäften. Außerdem nutzen sie moderne Kommunikationsmittel wie Mobiltelefone und das Internet - auf teilweise sehr innovative und spannende Weise.

In dieser BA-Arbeit soll ein Prototyp eines Bankterminals für Analphabeten entwickelt werden. Das Terminal soll sowohl auf mobilen Geräten wie auch in klassischen Bankautomaten verwendet werden können – auf dem Mobilgerät können z.B. Überweisungen getätigt werden, am Automaten Ein- und Auszahlungen. Es soll ein intuitives und sicheres Bedienkonzept entwickelt werden, Text, Grafik und Sprachein- oder -ausgabe sinnvoll kombiniert. Dieses Konzept soll in einem Demonstrator für eine bislang nicht sprachtechnologisch unterstützte Sprache, z.B. Ewe in Togo, implementiert und evaluiert werden.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Visualisierung von Reiseblogs

Viele Reisende schreiben Blogs, in denen sie ihre Routen und Erlebnisse für ein breites Publikum schildern. Diese Blogs enthalten vielfältige Referenzen, z.B. auf Orte, Ereignisse, Personen und Datums- und Zeitangaben, und oft auch Bilder, Audio und Video. Neben ihrem Unterhaltungswert dienen diese Blogs häufig auch als Informationsquelle für andere Reisende.

In dieser BA-Arbeit soll eine Visualisierung von Reiseblogs entwickelt werden, in der mehrere Blogs, Reiseführer und andere Informationsquellen verknüpft werden. Diese Visualisierung soll einen explorativen Zugriff auf die Ausgangsdaten ermöglichen und so eine multimediale und -dimensionale eigene Reisewelt aufspannen.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Mobiler Therapie-Assistent für Jugendliche mit Leseschwäche

Zahlreiche Studien belegen einen hochsignifikanten Zusammenhang zwischen Leseflüssigkeit/Lesegeschwindigkeit und Leseverstehen. Ferner wird im kompetenten Leseprozess die automatische Worterkennung als das entscheidende Kriterium für das flüssige Lesen erachtet.

In aktuellen computerbasierten Assistenzprogrammen bleibt meist unberücksichtigt, dass die Lesefähigkeit nicht nur das lexikalische Erkennen, sondern noch weitere Lesefertigkeiten wie das automatische Erkennen prosodischer und grammatischer Strukturen auf Wort- und Satzebene voraussetzt. Sehr wahrscheinlich ist es sogar so, dass lexikalisches Erkennen nur dann hoch automatisiert erfolgen kann, wenn die prosodischen und grammatischen Strukturen auf Wortebene und später auch auf Satzebene dekodiert werden können.

In dieser BA-Arbeit werden bestehende Assistenzprogramme ermittelt und auf ihre technische Umsetzung sowie die inhaltliche Eignung analysiert. Darauf aufbauend wird anhand fachlicher Vorgaben einer erfahrenen Therapeutin eine Prototyp-App entwickelt, mit der Kinder und Jugendliche selbständig das flüssige Lesen durch sprachsystematische Spiele üben und verbessern können.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Mobiler Therapie-Assistent für Aphasie-Patienten

Infolge einer Hirnschädigung (z.B. Schlaganfall oder Schädel-Hirn-Trauma) kommt es häufig zu einer Sprachstörung ("Aphasie"). Ein gezieltes Training sprachlicher Funktionen kann die Kommunikation dieser Patienten deutlich verbessern und ihnen somit wieder eine bessere Teilhabe am Leben ermöglichen. In der Aphasie-Therapie werden verschiedene sprachliche Übungen zum Sprechen, Verstehen, Lesen und Schreiben durchgeführt. Grundlage dieser Therapien sind umfangreiche, an die Bedürfnisse jedes Patienten angepasste Materialsammlungen.

In dieser BA-Arbeit soll ein web-basiertes System für ausgewählte Übungen für die Aphasie-Therapie entwickelt werden. Das System soll sowohl für den Einsatz in Therapiestunden als auch zur eigenverantwortlichen Übung zuhause geeignet sein, es soll den Verlauf der Therapie dokumentieren und eine flexible Anpassung an individuelle Patienten erlauben. Das System soll sowohl am PC als auch auf mobilen Endgeräten, z.B. Tablets nutzbar sein.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Computerunterstützte Dramenanalyse

In den Digital Humanities wird der Computer zum zentralen Werkzeug in der wissenschaftlichen Arbeit. In den Literaturwissenschaften gibt es seit langem Versuche literarische Werke, z.B. Romane, Dramen, Lyrik mit mathematischen Modellen zu beschreiben. Mit diesen Modellen können Untersuchungen der Autorschaft, des Stils oder der Struktur dieser Werke durchgeführt werden. In neuerer Zeit werden diese Ansätze wieder verstärkt verfolgt, um z.B. große computerlesbare Korpora literarischer Werke aufzubauen um sie der Forschung zugänglich zu machen.

In dieser BA-Arbeit soll ein System zur Computerunterstützten Dramenanalyse als Webanwendung realisiert werden. Dieses System soll als eine offene Plattform für die Dramenanalyse gestaltet werden, die sowohl den Aufbau einer zentralen Dramen-Sammlung als auch individuelle Arbeitsbereiche unterstützt.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Optimale web-basierte Transkription langer Audiodateien

In den Archiven sprachverarbeitender Forscher und Institutionen lagern große Mengen an Sprachaufnahmen, die für wissenschaftliche Untersuchungen sehr interessant wären, aber noch nicht ausreichend aufbereitet sind, um sie für diese Untersuchungen zu nutzen. Grundlage aller weiteren Nutzungen ist eine mindestens orthographische Transkription – diese kann auch von ungeübten Personen im Crowdsourcing erstellt werden.

In dieser BA-Arbeit soll eine Crowdsourcing-Lösung entwickelt werden, die digitalisierte Sprachaufnahmen langer Dauer zur Transkription anbietet. Um

diese Transkription möglichst effizient durchführen zu können, soll das Sprachsignal in geeigneter Weise geschnitten (z.B. in so kurze Einheiten, dass ein mehrfaches Anhören nicht notwendig ist) und für die Transkription in optimaler Weise präsentiert werden. Die Transkription soll sowohl auf mobilen Geräten als auch auf normalen Rechnern möglich sein, und die Einzeltranskripte müssen zu einem Gesamttranskript einer Aufnahme zusammengefügt werden können, und es muss der Beitrag jedes einzelnen Transkribierers erkennbar bleiben (z.B. für Abrechnungszwecke).

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Bereitstellung von Sprachmaterial zum Trainieren von MAUS

MAUS ist ein forced-alignment System zur automatischen Segmentierung von Sprachaufnahmen. Damit MAUS eine Sprache gut segmentieren kann ist ein Training der akustischen Modelle notwendig – für jede zu segmentierende Sprache gibt es ein eigenes Sprachmodell. Dieses Training erfordert eine Mindestmenge an Sprachaufnahmen in der jeweiligen Sprache, und unterschiedliches Sprachmaterial, z.B. gesprochene Ziffern, Zahlen, Datums- und Zeitangaben, Buchstabierungen, Eigennamen, und phonetisch reiche Sätze.

Ziel dieser BA-Arbeit ist es, für einzelne Sprachen dieses mindestnotwendige Sprachmaterial zu definieren und in Form von Aufnahmeskripten in WikiSpeech abzulegen. Damit ist es dann über das Web möglich, in kurzer Zeit ausreichend viele Sprecher in einer Sprache aufzunehmen, und diese Aufnahmen können dann dazu dienen, MAUS auf diese Sprache zu trainieren.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Grafisches UI für openSMILE

Das open source Paket **openSMILE** ist eine Sammlung verschiedener Algorithmen zur Berechnung akustischer Merkmale, vorzugsweise für Sprachaufnahmen. Es hat eine außerordentlich große Anzahl an Konfigurationsparametern, die den Umgang mit dem Paket recht umständlich machen.

In dieser BA-Arbeit soll ein grafisches UI für openSMILE entwickelt werden, das einem Benutzer die Verwendung von openSMILE vereinfacht. Für die verschiedenen Algorithmen sollen sinnvolle Standardeinstellungen vorgegeben und anwendungsabhängig nur die Optionen verfügbar sein, die wirklich notwendig sind. Dies kann je nach Aufgabenstellung unterschiedlich sein – ein Wizard kann hier weiterhelfen.

Nach der Verarbeitung durch openSMILE sollen die Ergebnisse in geeigneter Weise visualisiert bzw. für den Export (z.B. in Statistik-Programme) formatiert werden.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Klassifikation und zeitlicher Verlauf von Versprechern in der Ph@ttSessionz Sprachdatenbank

Ph@ttSessionz ist eine große Sprachdatenbank mit den Stimmen Jugendlicher. Sie enthält sowohl gelesene als auch spontan gesprochene Sprache. In der Sprachdatenbank gibt es auch jede Menge Äußerungen mit Auffälligkeiten: Aussprachefehler, Versprecher, Reparaturen, usw.

In dieser BA-Arbeit sollen die in den Transkriptionen der Ph@ttSessionz-Sprachdatenbank notierten Auffälligkeiten a) klassifiziert und b) in ihrem zeitlichen Verlauf untersucht werden. Beim zeitlichen Verlauf interessiert vor allem die Frage, ob es auf nonverbalen Ebenen Hinweise auf bevorstehende Fehler auf der verbalen Ebene gibt – mit anderen Worten: hört man schon, ob gleich ein Fehler kommt? Diese Frage soll u.a. mit einem Perzeptionsexperiment empirisch untersucht werden.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Editor für akustische Perzeptionsexperimente

Percy ist ein Framework für online Perzeptionsexperimente. In einem akustischen Perzeptionsexperiment hören Versuchspersonen sprachliche Stimuli und geben ihre Beurteilung ein; aufgezeichnet werden neben den Benutzereingaben auch noch weitere Daten wie Reaktionszeit, Häufigkeit des Abspielens, usw.

In dieser BA-Arbeit soll ein intuitiv zu bedienender grafischer Editor für die Gestaltung von online Perzeptionsexperimenten als Webdienst entwickelt und implementiert werden.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Game with a Purpose: Audioannotation als online-Spiel

Die Annotation von Audiodaten ist sehr zeitaufwendig, selbst wenn nur der Wortlaut niedergeschrieben werden soll. Diese wortwörtliche Transkription ist aber notwendig, um in sinnvoller Weise auf den Inhalt der Äußerung zugreifen zu können, z.B. über eine Wortsuche.

In dieser BA-Arbeit soll die Annotation von Audioaufnahmen als online-Spiel realisiert werden. Wie kann man das Spiel so attraktiv gestalten, dass viele Spieler auch über längere Zeit mitmachen, wie kann man eine möglichst hohe Qualität der Eingaben gewährleisten, und wie kann man es so implementieren, dass es auf möglichst vielen und auch mobilen Endgeräten läuft? Die Implementation des Spiels soll so flexibel sein, dass man statt Audio z.B. auch Bilder oder Videodaten annotiert.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Interaktiver Aussprache-Tutor für Fremdsprachen

Ziel ist die Entwicklung eines Anwendungsprogramms, das das Erlernen der Aussprache der Zielsprache unterstützt. Dazu wird eine Äußerung des Lerner aufgenommen, und der Lerner hört sie sich im Vergleich zu einer gespeicherten muttersprachlichen Äußerung an. Dabei markiert er in einem graphischen Interface die Stellen in der Äußerung, an denen der a) entweder einen großen Unterschied zur Vorgabe, oder b) keinen Unterschied zur Vorgabe gehört hat.

Diese Eingabe wird vom Programm genutzt um weitere Lernschritte vorzuschlagen. So können z.B. Stellen, die häufig falsch ausgesprochen werden, wiederholt angeboten werden.

Zu jedem Zeitpunkt soll eine übersichtliche Darstellung des Lernfortschritts möglich sein (z.B. Kurven-Diagramm), außerdem soll eine Verwaltung der aufgenommenen Audiodateien möglich sein (z.B. um mit früheren Aufnahmen zu vergleichen usw.).

Implementiert wird das Programm in Java, wobei die Audio-Module des IPS zur Verfügung stehen. Die Software soll, auch wenn es eine standalone Version wird, als Client-Server System konzipiert und implementiert werden, damit eine spätere Portierung auf eine web-basierte Anwendung einfach realisierbar wird.

Grundlage soll ein japanisches Lehrbuch, möglichst mit Hörbuch sein.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Bild-basiertes Katalogisierungssystem

In vielen Kellern und Speichern lagern große Mengen an mehr oder weniger sortierten bild-basierten Dokumenten, z.B. lose Briefmarken, Fotosammlungen, usw. Eine Erfassung dieser Dokumente ist sehr zeitaufwendig und wird daher immer wieder verschoben.

In dieser BA-Arbeit soll ein Katalogisierungssystem auf der Basis von Bild- und Texterkennung entwickelt werden, das speziell für den Einsatz im privaten Bereich gedacht ist. Es soll mit handelsüblichen oder eingebauten Kameras Dokumente erfassen, diese mit bereits gespeicherten Dokumenten (lokal oder im WWW) vergleichen und dann in einer Datenbank ablegen. Besonderes Gewicht soll dabei auf eine möglichst intuitive und effiziente Bedienung des Systems, z.B. mit Sprach- oder Gestensteuerung auf mobilen Geräten, gelegt werden.

Verkettung von Audioverarbeitung und Annotation von Sprachdaten-Diensten

Die Sprachdatenbank WikiSpeech erlaubt die Aufnahme, Annotation und Nutzung von Sprachdaten. WikiSpeech lässt sich über eine grafische Oberfläche, direkt im SQL Terminal oder über Java-Programme mittels JPA bedienen.

In dieser BA-Arbeit soll der Zugriff auf die WikiSpeech-Sprachdatenbank entsprechend der REST-Architektur neu konzipiert und implementiert werden. Insbesondere soll dabei die Möglichkeit der Verkettung von Arbeitsschritten zu Arbeitsabläufen und der Austausch von einzelnen Komponenten durch andere berücksichtigt werden.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Abfragesprache für Mehrebenen-Annotationen von Sprachdatenbanken

Formal lassen sich Annotationen von Sprachdatenbanken als gerichtete zyklische Graphen beschreiben, in dieser Form auch speichern, und mit Graph-basierten Abfragesprachen durchsuchen. Für die Nutzer von Sprachdatenbanken ist diese Darstellung jedoch zu abstrakt – die Organisation von Annotationen in separaten Ebenen mit festgelegten quantitativen Beziehungen zwischen diesen Ebenen folgt einer logischen und dem Anwendungsgebiet angepassten Struktur. Diese Struktur ist dem Anwender vertraut, da sie seine Sicht auf die Daten implementiert, und daher soll die Abfragesprache diese Struktur widerspiegeln.

In dieser BA-Arbeit soll eine Datenbankabfragesprache entwickelt werden, die die fachspezifische Organisation der Sprachdatenbank berücksichtigt, die wichtigsten benötigten Operatoren implementiert und syntaktisch einfach ist. Diese Abfragesprache wird dann zur Ausführung nach SQL, XQuery oder eine sonstige Datenbankabfragesprache übersetzt.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Explorative Abfragen in Sprachdatenbanken

Die WikiSpeech-Sprachdatenbank enthält symbolische und Sprachdaten vieler einzelner Sprachdatenbanken. Aktuell setzt die Nutzung entweder SQL-Kenntnisse voraus, oder sie ist auf nur sehr eingeschränkte Weise über simple Web-Interfaces zugänglich.

In dieser BA-Arbeit soll eine interaktive und explorative Benutzerschnittstelle entwickelt werden, die sowohl für Einsteiger als auch Experten geeignet sein und möglichst flexible Abfragen ermöglichen soll. Basis für diese Abfragesprache ist das ER-Modell der Datenbank; die Eingabe von Attribut-Wert-Paaren und deren logische Verknüpfung führt sofort zu einer grafischen Kennzeichnung oder Ausgabe der zu erwartenden Datensätze bzw. -menge.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Modellierung von Annotationen von Sprachdaten im Graph-Datenmodell

Im Graph-Datenmodell werden Elemente der abzubildenden Welt und ihre Beziehungen untereinander direkt in Form von binären Graphen dargestellt. Von dieser Modellierung erhofft man sich eine flexible und effiziente Datenstruktur auch für sehr große Datenbestände. Außerdem sollen Abfragen sehr einfach formuliert werden können.

In dieser BA-Arbeit soll eine große phonetische Sprachdatenbank mit dem Graph-Datenmodell dargestellt und in z.B. Neo4J implementiert werden. Typische Fragestellungen der Phonetik sollen in DB-Abfragen realisiert und die

Effizienz ihrer Berechnung, vor allem im Vergleich zu relationalen Datenbanksystemen, untersucht werden.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807

Zugriff auf eine große Sprachdatenbank mittels SRU/CQL

Die Sprachdatenbank WikiSpeech des Instituts für Phonetik enthält Sprecherangaben, Projektbeschreibungen, Audio- und Annotationsdaten. Sie soll nun über ein SRU (Search and Retrieval via URL) Interface mit der Abfragesprache CQL erreichbar sein.

In dieser BA-Arbeit soll die Abbildung der CQL-Sprachkonstrukte auf die relationale Datenstruktur (in PostgreSQL) bzw. die Java-Klassen der WikiSpeech Datenbank entsprechend der definierten Kompatibilitätsstufen entworfen und implementiert werden.

Kontakt: draxler@phonetik.uni-muenchen.de, Tel: 2180-2807