Suche in aufgezeichneten Vorträgen und Vorlesungen

Wolfgang Hürst

Institut für Informatik, Universität Freiburg, D-79110 Freiburg
huerst@informatik.uni-freiburg.de

Abstract: Durch den immer stärker werdenden Trend, Vorträge und Vorlesungen automatisch aufzuzeichnen und Studenten zur späteren Verwendung zur Verfügung zu stellen, entsteht auch ein zunehmender Bedarf an Suchverfahren für die derart entstandenen Dokumente. Dieser Artikel geht der Frage nach, wie sich eine Suchmaschine für aufgezeichnete Vorträge und Vorlesungen realisieren lässt, was die speziellen Probleme und Besonderheiten sind und welche Lösungsansätze hierfür existieren. Neben einer detaillierten Einführung in die Problematik werden erste Ergebnisse präsentiert und mögliche künftige Forschungsthemen aufgezeigt.

1. Einführung

Billigere Hardware sowie komfortable, einfach zu benutzende Präsentationssoftware haben dazu beigetragen, dass der Computer in Hörsäle und Unterrichtsräume Einzug gehalten hat. Neben dem reinen Nutzen zu Präsentationszwecken wird die betreffende Hardware auch immer mehr dazu verwendet, Vorträge und Vorlesungen aufzuzeichnen und für die spätere Verwendung zu konservieren [1]. Im Rahmen von Verbundprojekten, wie z.B. VIROR[1] oder ULI[2], sind auf diese Weise bereits zahlreiche Vorträge sowie komplette Kurse und Vorlesungen aufgezeichnet worden. Mit Hilfe komfortabler Aufzeichnungsverfahren und Programme, wie z.B. Lecturnity[3] oder Camtasia[4], lassen sich Vorlesungen auf einfache Art und Weise nahezu vollautomatisch und mit minimalem Mehraufwand aufzeichnen. Im wesentlichen werden dabei die verwendeten Folien samt der darauf während des Vortrags gemachten Annotationen sowie die Stimme des Vortragenden aufgezeichnet. Das Aufzeichnen weiterer Datenströme, z.B. eines Videobilds des Dozenten, ist zwar bei vielen Systemen prinzipiell möglich, wird in vorliegendem Beitrag jedoch nicht berücksichtigt, da diese Datenströme für den hier behandelten Schwerpunkt nur eine sehr geringe Relevanz haben (vgl. Kapitel 2).

Aufgrund dieser einfachen und kostengünstigen Möglichkeit zur Produktion multimedialer Dokumente wird die Aufzeichnung von Vorlesungen nicht nur im Rahmen geförderter Projekte betrieben, sondern scheint sich nach und nach auch im Alltagsbetrieb durchzusetzen. Beispielsweise gehen an unserer Fakultät immer mehr Professoren dazu über, ihre Vorlesungen routinemäßig aufzuzeichnen und sie dann den

[1] VIROR – Virtuelle Hochschule Oberrhein: *http://www.viror.de*
[2] ULI – Universitärer Lehrverbund Informatik: *http://www.uli-campus.de*
[3] imc Autorentools – Lecturnity: *http://www.im-c.de/homepage/autorentools_lecturnity.htm*
[4] TechSmith, Camtasia: *http://www.camtasia.com*

Studenten über das lokale Universitätsnetz zur Verfügung zu stellen. Studenten nutzen diese Dokumente insbesondere zum Wiederholen und Vertiefen des Stoffes beim Anfertigen von Übungsaufgaben oder zur Prüfungsvorbereitung. Neben einer derartigen Kurzzeitnutzung der Aufzeichnungen bietet es sich jedoch auch an, diese Dokumente potentiellen Nutzern langfristig zugreifbar zu machen, da die entstandenen Daten eine immense Wissensbasis bilden, welche insbesondere zur späteren Vertiefung einzelner Themengebiete oder für die Weiterbildung (Stichwort „Live-long learning“) interessante Perspektiven bietet. Unter anderem aus diesem Grund gab und gibt es verstärkt Ansätze, Vorlesungsaufzeichnungen (sowie andere multimediale Unterrichtsmaterialien) mit Hilfe von Metadaten auszuzeichnen, zu verwalten und in digitalen Repositories zur Verfügung zu stellen, um somit ein späteres Suchen und Finden der darin enthaltenen Information zu ermöglichen (siehe z.B. die Merlot[5] oder die ARIADNE[6] Web-Site). Aufgrund unserer Erfahrung ist es jedoch für eine komfortable Nutzbarkeit der Daten auch zwingend erforderlich, zusätzlich zu einer Verwaltung und Suche auf der Basis von Metadaten die Möglichkeit einer Suche auf einer detaillierteren Ebene anzubieten, vergleichbar der Funktionalität von Web-Suchmaschinen, die ebenfalls nicht nur auf der Basis von Metadaten operieren, sondern eine tiefergehende Inhaltsanalyse betreiben. Diese These wird u.a. von einer in [2] präsentierten Studie gestützt, bei der die Nutzung aufgezeichneter Vorlesungen durch Studenten untersucht wurde und welche herausfand, dass diese Dokumente sehr selektiv verwendet werden. Benutzer betrachten in der Regel nicht die gesamte Aufzeichnung in voller Länge, sondern nur einzelne, ausgewählte Teile davon. Aufgrund einer relativ großen Gesamtlänge, die teilweise über einer Stunde liegt, findet sich bestimmte Information häufig über verschiedene Positionen eines gesamten Dokuments verstreut und einzelne Dokumente können unterschiedliche thematische Bereiche oder Schwerpunkte abdecken. Daher ist es notwendig, den Benutzern eine gezieltere Suche und einen verfeinerteren Zugriff auf ausgewählte Teile der Dokumente zu ermöglichen, als dies auf der Ebene von Metadaten möglich ist.

Die Realisierung einer detaillierten Suche in Lehrmaterialien allgemein – nicht nur in Vortragsaufzeichnungen – betrachten wir als essentiell für eine langfristige und nachhaltige Nutzung der erstellten Dokumente. Die dabei zugrundeliegenden Daten unterscheiden sich teilweise signifikant von anderweitigen Dokumenten, so dass hierfür eine Anpassung und Weiterentwicklung existierender Standardverfahren zur Suche notwendig ist. Ziel dieses Artikels ist es, einen Überblick über spezielle Probleme bei der Suche in aufgezeichneten Vorträgen zu geben, über erste Lösungsansätze zu berichten und mögliche Bereiche für zukünftige Forschung zu skizzieren. Kapitel 2 beschreibt, warum Standardansätze zur Indizierung und Suche hier nicht immer verwendet werden können und statt dessen spezielle Lösungen gefordert sind. In Kapitel 3 werden erste Lösungsansätze und Ergebnisse präsentiert. Kapitel 4 gibt einen Ausblick auf aktuelle und zukünftige Forschungsbereiche.

---

[5] Merlot Web-Site: *http://www.merlot.org*
[6] ARIADNE Web-Site: *http://www.ariadne-eu.org*

2. Probleme bei der Suche in aufgezeichneten Vorträgen

Ausgehend von dem Ziel, eine Suchmaschine für aufgezeichnete Vorträge zu realisieren, werden in diesem Kapitel die dabei entstehenden Probleme und Besonderheiten illustriert. Traditionelle Verfahren zur Indizierung und zur Suche von Textdokumenten beruhen in der Regel auf der Annahme, dass die Verteilung einzelner Terme in einem Dokument etwas über deren Relevanz für eben dieses Dokument aussagt [3]. Intuitiv nimmt man an, dass ein Term für ein Dokument relevanter als ein zweiter Term ist, wenn er häufiger darin vorkommt. Ferner besitzt ein Term der insgesamt in nur sehr wenig Dokumenten vorkommt eine wesentlich höhere Relevanz als ein Term, der in fast jedem Dokument der Kollektion erscheint. Mathematisch lässt sich die erste Aussage durch die sogenannte *Term Frequency* modellieren ($TF = (\log (t+1) / \log (dl))$, wobei t die Anzahl der Vorkommen des Terms im Dokument und dl die Dokumentlänge, d.h. die Anzahl aller Terme beschreibt), die zweite Aussage durch die sogenannte *Inverse Document Frequency* ($IDF = \log (N / n)$, wobei N die Gesamtzahl aller Dokumente der Kollektion und n die Anzahl der Dokumente, in denen der Term vorkommt, beschreibt).

Verfahren, die das Produkt dieser beiden Maßzahlen ($TF*IDF$) für die Suchbegriffe berechnen und als Gradmaß für die Relevanz verwenden, mit dem dann ein Ranking der gefundenen Dokumente bestimmt wird, haben sich in der Vergangenheit für die Suche in Texten bewährt und ihre Leistungsfähigkeit in zahlreichen empirischen Tests und Evaluationen bestätigt. Dies gilt jedoch im wesentlichen nur für „traditionelle“ Texte, wie z.B. Bücher, Artikel, Abstracts usw. sowie für homogene Datenbasen. Betrachtet man beispielsweise die Funktionsweise von Web-Suchmaschinen so wurden bei den ersten Versionen tatsächlich Verfahren zur Suche verwendet, die im wesentlichen auf einer Analyse der Termverteilung basierten. Einer der Hauptgründe, warum diese Verfahren in der Regel jedoch keine zufriedenstellenden Ergebnisse liefern, ist (neben der einfachen Manipulationsmöglichkeit durch einen Dokument-Anbieter) die Heterogenität der Datenbasis. Während manche Web-Seiten beispielsweise nur stichwortartigen Text und viele Bilder enthalten, ist auf anderen sehr viel kontinuierlicher Text dargestellt, so dass eine Vergleichbarkeit der beiden Seiten aufgrund der reinen Textinformation und Termverteilung kritisch ist. Unter anderem aus diesem Grund verwenden fast alle heutigen großen Suchmaschinen ein Ranking, das sich hauptsächlich an der Linkstruktur orientiert. Basierend auf der Annahme, dass ein Link von einer Seite auf eine andere eine Qualitätsaussage über diese Seite macht, wird hierbei mittels eines rekursiven Berechnungsschemas ein Qualitätsmaß für jede Seite berechnet, welches dann zur Bestimmung des Rankings herangezogen wird [4]. Dadurch wird von den eigentlichen, heterogenen Inhalten abstrahiert und statt dessen eine einheitliche, vergleichbare Struktur – die Verlinkung der Seiten miteinander – zur Relevanzbewertung herangezogen. (Man beachte, dass die Inhalte der Seiten bei der Auswahl einer Menge potentiell relevanter Dokumente nach wie vor berücksichtigt werden. Lediglich das Ranking, d.h. die Sortierung dieser Dokumente gemäß ihrer Relevanz für eine bestimmte Suchanfrage basiert auf einer Analyse der Verlinkung.)

Bei der Suche in aufgezeichneten Vorträgen steht man aufgrund der enthaltenen Datenströme vor einer ähnlichen Situation. Die wesentliche, für die Suche relevante Information bei aufgezeichneten Vorträgen ist zum einen im Folienstrom enthalten,

welcher die in einer Vorlesung verwendeten Folien enthält, sowie im Audiostrom, in welchem die Stimme des Vortragenden aufgezeichnet wurde. Der Folienstrom unterscheidet sich von „traditionellem“ Text im wesentlichen dadurch, dass häufig keine kontinuierlichen Sätze, sondern nur stichwortartige Umschreibungen verwendet werden. Häufige Wiederholungen einzelner Begriffe werden oft zugunsten einer knappen und kurzen Darstellung vermieden, selbst wenn die betreffenden Terme im jeweiligen Kontext besonders wichtig sind. Die „Wichtigkeit“ eines Terms wird also häufig nicht durch sein vermehrtes Verwenden, sondern durch andere Dinge, meist visueller Natur, ausgedrückt, wie z.B. Hervorhebungen durch eine andere Farbe, Fettdruck oder ähnliches. Ferner kommt hinzu, dass Information auf Folien, ähnlich wie bei Web-Seiten, auf sehr unterschiedliche Art und Weise präsentiert werden kann. Beispielsweise verwenden manche Vortragende relativ viel Text, während andere eher nur Stichworte gebrauchen und vermehrt auf den Einsatz von Bildern und Grafiken setzen, was zu einer starken Heterogenität der Datenbasis führt und eine Relevanzbeurteilung erschwert.

Zur Suche in Sprachsignalen wird meist folgendes Verfahren angewandt: Mit Hilfe von automatischer Spracherkennung wird aus dem Audiosignal eine textuelle Beschreibung, ein Audio-Transcript, erstellt, auf welches dann Standardverfahren zur Indizierung und Suche von Texten angewendet werden. Trotz der in den letzten Jahren erzielten immensen Fortschritte in der automatischen Spracherkennung enthalten die derart erstellten Transcripts immer noch sehr viele Fehler, die sich natürlich bei der Suche negativ auswirken können. Zahlreiche empirische Studien (siehe z.B. [5,6]) sowie realisierte Systeme (z.B. [7]) haben jedoch bestätigt, dass der Einfluss dieser Erkennungsfehler auf das Suchergebnis eher gering ist. Grund dafür ist zum einen die hohe Redundanz des Sprachsignals. Wie für geschriebenen Text (vgl. Argumentation zu Beginn dieses Kapitels) gilt auch für die gesprochene Sprache, dass Begriffe, die in einem bestimmten Kontext eine höhere Relevanz besitzen, häufiger vorkommen. Beispielsweise wird ein Term, der in einem Sprachsignal zehn mal auftaucht, bei einer Erkennungsrate von nur 50% im Schnitt fünf mal richtig erkannt, während ein anderer, welcher nur zwei mal vorkommt, durchschnittlich nur ein mal richtig erkannt (und damit bei der Suche gefunden) wird. Hinzu kommt, dass ein Großteil der Fehler bei der automatischen Spracherkennung im Zusammenhang mit kurzen Worten, z.B. Füll- oder Stopworte, vorkommt, während längere Begriffe, welche auch für die Suche eine wichtigere Rolle spielen, in der Regel viel zuverlässiger erkannt werden. Trotz einer hohen Fehlerrate bei der Spracherkennung ist man daher in der Lage, bei der Suche immer noch ein brauchbares Ergebnis zu erzielen. Auch hier ist es jedoch so, dass sich die Wichtigkeit eines Terms nicht nur durch seine Verteilung im Sprach- bzw. daraus produzierten Text-Signal ausdrücken kann, sondern dass auch andere Merkmale existieren, die auf eine höhere Relevanz schließen lassen, wie z.B. die Prosodie oder Satzmelodie, Pausen, die Gestik und Mimik des Sprechers, usw. Ferner kann auch die grundlegende Struktur der Sprache von Sprecher zu Sprecher sehr verschieden sein. Während die Sprache mancher Vortragenden einer nahezu perfekten Schriftsprache entspricht, existiert auch das andere Extrem: Spontansprache, die geprägt ist von grammatikalischen Fehlern (Sätze oder gar Worte werden in der Mitte abgebrochen und auf eine andere Art fortgesetzt, etc.), der Verwendung von Füllworten („Ähm“, „Hmm“, etc.), dem häufigen Auftreten von Versprechern usw. Selbst mit einer perfekten Spracherkennung entstünden somit Texte, die sich in Struktur und Aufbau mitunter recht

stark voneinander unterscheiden, weshalb sich derartige Phänomene nicht nur auf das Ergebnis der Spracherkennung auswirken können, sondern möglicherweise auch einen Einfluss auf die verwendeten Such- und Rankingverfahren haben.

3. Ansätze zur Suche in aufgezeichneten Vorträgen

Wie im vorangehenden Kapitel illustriert, hat man es in beiden Fällen, beim zugrundeliegenden Folienstrom wie auch bei den Sprachsignalen und den daraus resultierenden Transcripts, mit einer sehr heterogenen Datenbasis zu tun, eine für die Suche und Relevanzbewertung problematische Situation, die den Einsatz von Standard-Suchverfahren fraglich macht. Im Folgenden werden daher einige alternative Ansätze für die Indizierung und Suche in diesen beiden Datenströmen präsentiert.

3.1 Suche im Folienstrom

Bei unseren Untersuchungen im Zusammenhang mit einer Indizierung und Suche in aufgezeichneten Vorträgen, welche rein auf den Texten beruht, die auf den jeweiligen Folien vorkommen, hat sich wie erwartet herausgestellt, dass ein auf Termverteilung basierendes Rankingverfahren (wie z.B. *TF*IDF*, vgl. Kapitel 2) zwar häufig zu guten Ergebnissen führt, es aber auch zu Problemen und Fehlbeurteilungen kommen kann, insbesondere aufgrund der Heterogenität der Folien unterschiedlicher Vortragender und der damit verbundenen, oben beschriebenen Besonderheiten des Textes. Aus diesem Grund wurden von uns weitere Merkmale untersucht, die Rückschlüsse auf eine besondere Relevanz einzelner Terme zulassen und damit eine potentielle Möglichkeit darstellen, die Suchergebnisse zu verbessern. Besonders häufig werden von Vortragenden visuelle Mittel eingesetzt, um eine besondere Bedeutung hervorzuheben, insbesondere eine andere Font-Farbe, ein anderer Stil (Fett- oder Schrägdruck) sowie eine andere Font-Größe. Betrachtet man diese Merkmale eines einzelnen Terms relativ zu den entsprechenden Werten der anderen Terme (z.B. ein Term hat eine im Verhältnis zu allen anderen Termen auf einer Folie oder in einem Dokument relativ selten vorkommende Farbe), dann lassen sich diese Eigenschaften zur Schätzung der Relevanz eines Dokuments einsetzen. Ein weiteres von uns berücksichtigtes Merkmal ist die Auflagedauer einer Folie, wobei eine längere Auflagedauer einer höheren Relevanz entspricht. Intuitive Annahme hierbei ist, dass von zwei ansonsten völlig identischen Folien diejenige wichtiger ist, die den Zuhörern länger präsentiert wurde.

All diese beschriebenen Merkmale können dazu verwendet werden, eine besondere Relevanz einzelner Terme zu verdeutlichen. Jedoch ist zu beachten, dass es sich hier nur um intuitive Annahmen handelt, die in der Praxis zwar häufig jedoch nicht immer zutreffen. Aus diesem Grund sollte bei der Schätzung der Relevanz basierend auf diesen Kriterien ein Verfahren angewendet werden, welches sämtliche angeführten Merkmale berücksichtigt und robust gegen einzelne Abweichungen ist. In unserer derzeitigen Implementierung wird eine einfache lineare Kombination der erwähnten Merkmale durchgeführt, deren Gewichtung aufgrund von Beispieldaten bestimmt wurde. Erste Experimente mit ausgefeilteren, auf maschinellem Lernen basierenden Ansätzen zur

Relevanzabschätzung verliefen vielversprechend, bedürfen jedoch noch einer detaillierteren Untersuchung, um ihre Leistungsfähigkeit abschließend einschätzen zu können. Abbildung 1 zeigt einen Snapshot unserer derzeitigen Implementierung, welche auch online verfügbar ist[7]. Suche und Relevanzbewertung basieren hier jedoch ausschließlich auf den aus den Folien extrahierten Texten gemäß der oben beschriebenen Vorgehensweise. Die Suche im aufgezeichneten Audiostrom wird im nächsten Abschnitt behandelt.

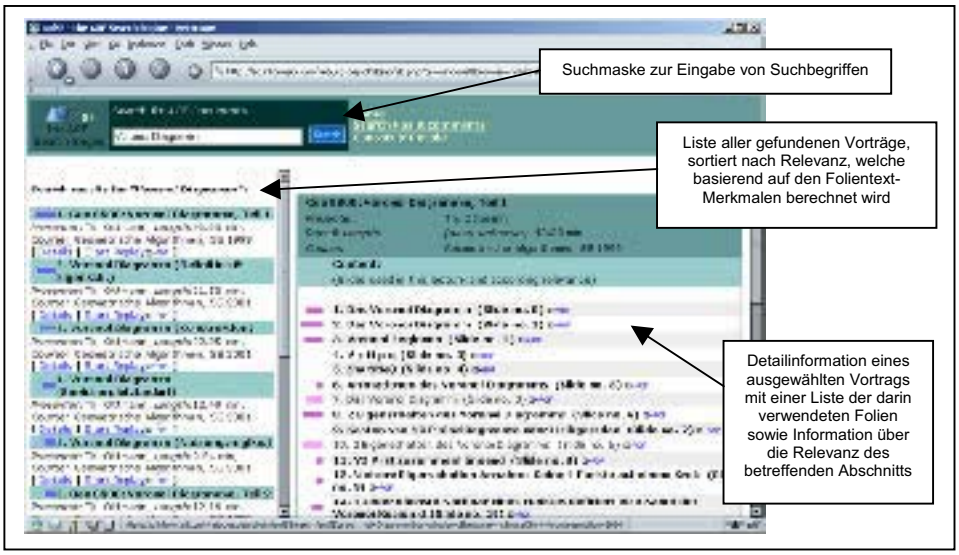

Abbildung 1: Snapshot eines Prototypen zur Folien-basierten Suche in aufgezeichneten Vorträgen.

3.2 Suche im Audiostrom

Wie bereits in Kapitel 2 erwähnt, wirken sich die durch die automatische Spracherkennung entstandenen Fehler im Audio-Transcript häufig nicht allzu sehr auf das Suchergebnis aus. Viel problematischer im Zusammenhang mit der Suche ist ein Problem, welches in der Spracherkennung unter dem Namen Out-of-Vocabulary-Problem (OOV) bekannt ist. Gängige Spracherkennungsverfahren beruhen in der Regel auf einem festen Lexikon oder Vokabular und können nur Worte dieses Vokabulars erkennen. Insbesondere bei fachspezifischen Dokumenten, wie z.B. aufgezeichneten Vorträgen, kann es jedoch vorkommen, dass einzelne (Fach-)Begriffe nicht im betreffenden Vokabular eines Spracherkenners existieren und daher nicht erkannt werden. Gerade Fachbegriffe haben jedoch in unserem Fall für die Suche aus naheliegenden Gründen eine besondere Bedeutung.

Bei aufgezeichneten Vorträgen bietet sich jedoch eine relativ naheliegende und einfache Vorgehensweise zur Abhilfe für dieses Problem an: Der Text aus dem zu jedem

[7] aofSE– The AOF Search Engine: *http://ad.informatik.uni-freiburg.de/aofSE*

Audiostrom in der Regel existierenden Foliensatz kann benutzt werden, um automatisch themenabhängige Vokabulare zu generieren und damit das OOV-Problem zumindest in einem gewissen Rahmen zu lösen. Zunächst können natürlich die Folientexte selbst dem Vokabular hinzugefügt werden. Darüber hinaus können sie auch dazu benutzt werden, das Thema eines Vortrags oder einer Menge aufgezeichneter Vorträge zu klassifizieren, um anschließend weitere Worte aus der gleichen Kategorie dem Vokabular hinzuzufügen. Im Rahmen einer empirischen Studie haben wir auf diese Art exemplarisch verschiedene themenabhängige Vokabulare erzeugt und die betreffende Spracherkennungsrate sowie die Qualität der Suchergebnisse untersucht. Das Standardvokabular des dazu benutzten Spracherkennungssystems wurde dafür einmal um die Folientexte der jeweiligen Vorträge erweitert, zum anderen wurden weitere themenabhängige Begriffe hinzugefügt, welche aus manuell ausgewählten Texten stammten. Die damit durchgeführten Studien bestätigten nicht nur eine signifikante Verbesserung gegenüber dem Gebrauch des nicht modifizierten Standardvokabulars, sondern auch, dass sich auf diese Weise ein brauchbares Suchergebnis erzielen lässt. Im Vergleich zu einem Ergebnis, welches mit einer Top-10-Suche auf einem perfekten, von Hand erstellten Transcript erzielt wurde, wurden mit einem automatisch erstellten Transcript im besten Fall über 80% der relevanten Ergebnisse gefunden. Wie erwartet zeigte sich, dass die relativ niedrige Erkennungsrate für die Suche nicht allzu kritisch ist, wenn sie auch nicht völlig vernachlässigt werden sollte, weshalb wir in [8] verschiedene Evaluationen präsentieren, die untersuchen, ob und wie man das Spracherkennungsergebnis im Falle von Vortragsaufzeichnungen verbessern kann.

Interessant an diesem Ansatz sind nicht nur die damit erzielbaren Ergebnisse, sondern vor allem auch die Tatsache, dass er sich weitestgehend automatisieren lässt, auch wenn für unsere Evaluationen noch rein manuell vorgegangen wurde. Durch den Einsatz bereits existierender, sehr zuverlässig arbeitender Text-Klassifikationsverfahren (siehe z.B. [9]) lassen sich die themenabhängigen Vokabulare vollautomatisch erstellen. [10] berichtet über ein Verfahren, bei dem dies bereits erfolgreich durchgeführt wurde. Auch das Hinzufügen einzelner Worte zum Vokabular des Spracherkenners geschieht automatisch. Bei dem von uns verwendeten Erkennungssystem war dies bereits implementiert und wurde über Schnittstellen als Funktionalität zur Verfügung gestellt, eine Beschreibung, wie eine derartige Vokabularerweiterung automatisch durchgeführt werden kann, findet man z.B. in [11]. Hauptaugenmerk unserer derzeitigen Arbeiten im Bereich Audio-Suche ist deshalb die automatische Erstellung und Klassifikation themenabhängiger Vokabulare mittels Standardverfahren zur Text-Klassifikation. Ein erstes prototypisches System zur Audio-basierten Suche wurde von uns entwickelt und ist online verfügbar[8]. Abbildung 2 enthält einen Snapshot der betreffenden Benutzerschnittstelle. Die Suche wird hierbei auf einem mittels Spracherkennung vollautomatisch erstellten Transcript durchgeführt. Die jeweiligen Folientexte der Vorlesungen wurden nur zur Modifikation des vom Spracherkenner verwendeten Vokabulars benutzt, bleiben bei der Suche jedoch unberücksichtigt. Die gleichzeitige Suche sowohl im Audio-Transcript als auch in den Folientexten ist Gegenstand unserer aktuellen Forschungsarbeiten und wird u.a. im nächsten Kapitel angesprochen.

---

[8] aofSE – The AOF Search Engine, Audio Demos: *http://ad.informatik.uni-freiburg.de/mmgroup/aofSEaudio*

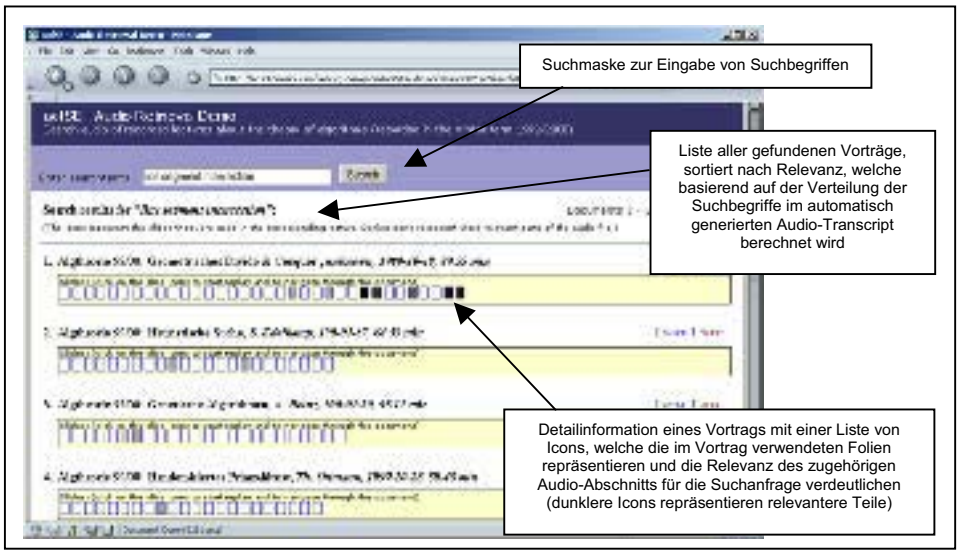

Abbildung 2: Snapshot eines Prototypen zur Audio-basierten Suche in aufgezeichneten Vorträgen.

4. Aktuelle und zukünftige Forschungsschwerpunkte

In den vorangehenden Kapiteln wurden diverse Probleme bei der Indizierung und Suche aufgezeichneter Vorträge erläutert sowie erste Ansätze zur Entwicklung einer Suchmaschine für Vorlesungsaufzeichnungen präsentiert. Neben den bereits erwähnten Schwerpunkten unserer gegenwärtigen Forschungsarbeiten – nämlich dem Einsatz alternativer, auf maschinellen Lernverfahren basierender Rankingverfahren (vgl. Kapitel 3.1) sowie der automatischen Generierung und Klassifikation themenabhängiger Vokabulare für die Spracherkennung (vgl. Kapitel 3.2) – gilt unser Hauptaugenmerk vor allem der Frage nach einer optimalen Integration der Folien-basierten und der Audio-basierten Suche. In einem ersten Ansatz wurde mit den im vorherigen Kapitel erwähnten Verfahren basierend auf dem Folien- bzw. Audiostrom jeweils ein Ergebnis getrennt berechnet und durch einfache Kombination der beiden Ergebnislisten ein neues Suchergebnis generiert. Die Evaluation des derart gebildeten Ergebnisses ergab eine signifikante Verbesserung der Recall-Leistung, welche sich definiert als *Anzahl relevanter Dokumente, die gefunden wurden / Anzahl aller relevanten Dokumente*, während die Precision, welche sich definiert als *Anzahl aller relevanten Dokumente, die gefunden wurden / Anzahl aller Dokumente, die gefunden wurden*, deutlich abnahm. Dies bedeutet, dass durch die gleichzeitige Berücksichtigung beider Ströme insgesamt zwar wesentlich mehr relevante Dokumente gefunden wurden, diese Leistungsverbesserung jedoch durch ein Ergebnis erkauft wurde, welches neben mehr relevanten auch mehr irrelevante Dokumente enthält. Gesucht sind daher intelligentere Kombinationsverfahren, die eine kombinierte Suche derart realisieren, dass sich beide Informationsquellen optimal ergänzen, zum Beispiel durch eine stärkere Berücksichtigung des Audiostroms, wenn sich die Folien aufgrund eines (zu) großen

Grafikanteils nur bedingt für die Suche eignen oder einer stärkeren Gewichtung des Folienstroms, wenn die Verwendung des aus dem Audiostrom erzeugten Transcripts aufgrund einer zu großen Anzahl von Spracherkennungsfehlern nicht erfolgversprechend erscheint. Auch hier bieten sich maschinelle Lernverfahren zur Schätzung der Relevanz an und werden derzeit von uns untersucht.

Neben einer gewünschten Verbesserung der Suchleistung, stellt sich auch die Frage, wie man die Ergebnisse dem jeweiligen Benutzer bestmöglich präsentiert. Selbst bei einer „perfekten" Suchmaschine wird ein Benutzer aufgrund von ungenauen, mehrdeutigen Suchanfragen immer wieder mit irrelevanten Dokumenten konfrontiert sein, so dass sich an jede Suchanfragenbearbeitung in der Regel ein Filterprozess durch den Benutzer anschließt, bei dem dieser die zurückgelieferte Ergebnisliste überfliegt, einzelne Einträge begutachtet und letztendlich die für ihn relevanten auswählt. Auch wenn diese finale Auswahl vom Benutzer selbst vorgenommen werden muss, kann und sollte das System ihn dabei unterstützen, indem die Ergebnisse so aufbereitet und dargestellt werden, dass dieser Filter- und Auswahlprozess auf die bestmögliche Art und Weise unterstützt wird. Wie bei der Indizierung und Suche selbst lässt sich auch hier argumentieren, dass hierfür die reine Berücksichtigung von Metainformation, wie z.B. Dokumenttitel und Verfasser, nicht ausreicht, um dem Benutzer einen schnellen und direkten Zugriff auf die einzelnen Stellen eines Dokuments zu ermöglichen, die für ihn relevant sind. Aus diesem Grund haben wir in [12] und [13] bereits verschiedene Schnittstellen-Varianten für Folien- bzw. Audio-basierte Suchmaschinen aufgezeichneter Vorträge separat voneinander untersucht und evaluiert. Die Antwort auf die Frage nach einer optimalen Integration der Darstellung beider Medientypen in ein einheitliches Interface ist jedoch auch hier, wie bei der Suche, noch völlig offen.

Während wir uns in vorliegendem Artikel auf die Suche in aufgezeichneten Vorträgen beschränkt haben, stellt sich für die Zukunft natürlich auch die Frage, ob und wie man weitere Dokumente, die für Lernzwecke zur Verfügung gestellt werden, in eine derartige Suchmaschine integrieren kann. Denkbar wäre z.B. eine direkte Integration, bei der auf eine Suchanfrage hin eine Ergebnisliste zurückgeliefert wird, die nicht nur aufgezeichnete Vorträge, sondern auch weitere Dokumente, wie z.B. Übungsaufgaben, weiterführende Literatur, veranschaulichende Animationen usw. enthält. Die Relevanzabschätzung und Berechnung eines Rankings dieser Dokumente ist jedoch aufgrund der völlig heterogenen Datenbasis äußerst kritisch. Alternativ bietet sich auch ein Ansatz an, bei dem die jeweiligen Dokumente vorab automatisch klassifiziert werden (z.B. in Vortragsaufzeichnung, Übungsblatt, Skriptum, etc.) und der Benutzer selbst entscheidet, nach welcher Art von Dokument gesucht werden soll, sei es durch vorherige Auswahl (z.B. „Suche nach aufgezeichneten Vorträgen" oder „Suche nach Übungsblättern", etc., ähnlich der bei manchen Web-Suchmaschinen möglichen Wahl, nach „normalen" Web-Seiten, Bildern oder Audio-Dokumenten suchen zu können) oder durch einen interaktiven Suchprozess (z.B. „Finde Übungsblätter, die thematisch zu diesem Vortrag passen", ähnlich der von vielen Web-Suchmaschinen bekannten „Find similar documents"-Funktion).

Acknowledgements

Die Arbeiten in vorliegendem Artikel wurden im Rahmen des Schwerpunktprogramms „Verteilte Verarbeitung und Vermittlung digitaler Dokumente“ (V3D2) von der Deutschen Forschungsgemeinschaft gefördert.

Literaturverzeichnis

[1] T. Ottmann, S. Trahasch, T. Lauer: Systems Support for Virtualizing Traditional Courses in Science and Engineering. Proceedings of Quality Education at a Distance, IFIP WG3.6, Geelong, Australien, Februar 2003.

[2] B. Zupancic, H. Horz : Lecture Recording and its Use in a Traditional University Course. Proceedings of ITICSE 2002, 7th Annual Conference on Innovation and Technology in Computer Science Education, Aarhus, Dänemark, 2002.

[3] G. Salton: A Blueprint for Automatic Indexing. SIGIR Forum 31(1): 23-36, 1997.

[4] S. Brin, L. Page: The Anatomy of a Large-Scale Hypertextual Web Search Engine. Proceedings of the 7th International World Wide Web Conference, Brisbane, Australien, April 1998.

[5] J. Garofolo, C. Auzanne, E. Voorhees: The TREC Spoken Document Retrieval Track: A Success Story. NIST Special Publication 500-246: The 8th Text REtrieval Conference (TREC-8), 2000.

[6] A. Hauptmann, H. Wactlar: Indexing and search of multimodal information. Proceedings of ICASSP ’97, International Conference on Acoustics, Speech and Signal Processing, München, Deutschland, 1997.

[7] J.-M. Van Thong, P.J. Moreno, B. Logan, B. Fidler, K. Maffey, M. Moores: Speechbot: An experimental speech-based search engine for multimedia content on the web. IEEE Transactions on Multimedia, Volume 4, Nr. 1, 2002.

[8] W. Hürst, T. Kreuzer, M. Wiesenhütter: A Qualitative Study Towards Using Large Vocabulary Automatic Speech Recognition to Index Recorded Presentations for Search and Access over the Web. Proceedings of WWW/Internet 2002, IADIS International Conference, Lissabon, Portugal, November 2002.

[9] Yang, Y. An Evaluation of statistical approach to text categorization. Technical Report CMU-CS-97-127, Computer Science Department, Carnegie Mellon University, 1997.

[10] I. Rogina, T. Schaaf: Lecture and Presentation Tracking in an Intelligent Meeting Room. Proceedings of the 2002 International Conference on Multimodal Interfaces (ICMI '02), Pittsburgh, PA, USA, Oktober 2002.

[11] W. Fischer: A Statistical Text-to-Phone Function Using N-Grams and Rules. Proceedings of ICASSP ’99, International Conference on Acoustics, Speech and Signal Processing, Phoenix, Arizona, März 1999.

[12] W. Hürst: Presenting Results of a Search Engine for Recorded Lectures in order to Support Relevance Decisions by the User. Proceedings of HCI International 2003 Conference, Kreta, Griechenland, Juni 2003.

[13] W. Hürst: User Interface Issues for Speech-based Retrieval of Recorded Lectures and Presentations. Eingereicht zur Veröffentlichung beim „5th International ACM SIGMM Workshop on Multimedia Information Retrieval, MIR 2003.