

Mobile verteilte Dokumentenrecherche in Bibliotheken und Archiven

D. Biella, W. Luther

Institut für Informatik und Interaktive Systeme (IIS)
Fakultät für Ingenieurwissenschaften
Universität Duisburg-Essen
Lotharstr. 65
47048 Duisburg
{biella,luther}@informatik.uni-duisburg.de

Abstract: Der Beitrag stellt ein Konzept zur mobilen verteilten Dokumentenrecherche in Bibliotheken und Archiven basierend auf WLAN-Notebooks, Kommunikationssoftware und einem Dokumentenserver vor. Anwendungskontext ist ein interdisziplinäres Retrodigitalisierungsprojekt, in dem Literatur zur Nachwirkung Friedrich Nietzsches erfasst und archiviert wird.

Projektumfeld und Anwendungskontext

An der Universität Duisburg-Essen sind am Standort Duisburg in den letzten Jahren Hörsäle und Seminarräume sowie die Servicebereiche Rechenzentrum, Bibliothek und das Audio-Visuelle Medienzentrum mit hochschulweit vernetzter multimedialer Präsentations- und Arbeitsplatztechnik ausgestattet worden, die neben Video-Conferencing auch interaktive Annotationen und Anreicherungen des Materials aus der Diskussion der Lehrenden mit den Studierenden heraus an einer elektronischen Tafel oder an Notebookarbeitsplätzen oder Plätzen mit Standrechnern zulassen.

Studierende können sich Arbeitsmaterialien herunterladen, gemeinsam bearbeiten und verteilen. Diese Räume spielen eine entscheidende Rolle, da von hier aus die Recherchen der Studierenden in Bibliotheken und Archiven begleitet werden können. In Ergänzung zu den festen Arbeitsplätzen besteht die Möglichkeit, sich in den Hörsaal- und Seminarraumbereichen, sowie an weiteren wichtigen studentischen Aufenthaltsbereichen des Campus in ein drahtloses Netz einzuloggen und die gleichen Angebote zu nutzen.

Diese technische Infrastruktur wird in einem von Bundesministerium für Bildung und Forschung unterstützten Teilprojekt *Mobile verteilte Dokumentenrecherche in Bibliotheken und Archiven* im Rahmen des Projekts *eCampus Duisburg* genutzt und weiterentwickelt [KKS03]. Projektteilnehmer digitalisieren Text und Bilder über eine am Notebook angeschlossene CCD-Kamera. Über das WLAN im Archiv werden die Daten zum Dokumentenserver geleitet und sind von dort über Internet einer Expertenrunde zugänglich. Diese koordiniert von einem vernetzten Seminarraum aus die Materialrecherchen, berät über die Vorgehensweise vor Ort und entscheidet über die Relevanz der Quellen. Dabei

kann eine OCR-Erfassung auch auf wichtige Textpassagen, die bei der Recherche an Hand von standardisierten Schlagwortsammlungen markiert werden, beschränkt oder durch manuelle Erfassung ersetzt werden. Diese Vorgehensweise eröffnet den Weg zu einer Teilerfassung von Seitenweise abgelegten Texten, was eine erhebliche Zeit- und Kostenersparnis gegenüber der klassischen Volltext-Erfassung zur Folge haben kann.

Anwendungskontext ist ein seit dem Frühjahr 2001 an der Gerhard-Mercator-Universität in Duisburg verfolgtes interdisziplinäres Retrodigitalisierungsprojekt, in dem Sekundärliteratur zur Nachwirkung Friedrich Nietzsches erfasst, digitalisiert und archiviert wird [Bc03, Bi02] und das im Kontext einer Vielzahl von internationalen Projekten zur Bewahrung des kulturellen Erbes zu sehen ist [Mi02]. Im Mittelpunkt des Projektes im Studiengang „Angewandte Kommunikations- und Medienwissenschaften“ steht Schrifttum, das zwischen 1865 und 1945 im deutschen Sprachraum verfasst worden ist. Ziel des Vorhabens ist es, erstmals bedeutende technische Grundlagen für eine effiziente Erforschung der Rezeptions- und Wirkungsgeschichte Friedrich Nietzsches zu schaffen, deren entscheidendes Quellenmaterial zu einem großen Teil nur in Frakturschrift vorliegt [Bi02, Mt01, Ju02]. Ein Erfolg des Projektes würde einen wichtigen Beitrag zur Kulturgeschichte Deutschlands im 19. und 20. Jahrhundert bedeuten und hätte wegweisenden Vorbildcharakter für ähnliche Untersuchungen.

Projektziele und Arbeitsmethodik

Ziel des Projektes ist es, Studierende, Wissenschaftlerinnen und Wissenschaftler bei Materialrecherchen in Archiven und Bibliotheken durch netzbasierte Beratung, Verteilung und Ablage der Dokumente zu unterstützen. Die folgenden Arbeitshypothesen werden anhand einer Fallstudie im Rahmen des langfristigen Retrodigitalisierungsprojekts untersucht.

- Über den Einsatz von Notebooks zur mobilen Bibliotheks- und Archivrecherche können Dokumente koordiniert erfasst, beurteilt, bearbeitet und verteilt werden. Dabei wird der zeitaufwändige Ausleihvorgang über Nah- oder Fernleihe verbunden mit Kopiervorgängen und Medienbrüchen durch neue Recherche- und Erfassungswege ergänzt.
- Durch parallele von Experten koordinierte Recherchen und Dokumentenerfassungen an verschiedenen Orten werden Arbeitsvorgänge zeit- und kosteneffizient gestaltet.
- Die Studierenden können individuell Materialrecherchen für Abschlussarbeiten vornehmen und dabei über das Netz von den Dozenten oder Tutoren angeleitet werden.
- Die Dokumente können mit standardisierten Werkzeugen weiterverarbeitet und auf einem zentralen Server mit transparenten Zugangsregelungen und digitalem Rechtemanagement (DRM) abgelegt werden.

Der Einsatz von mobilen WLAN-Notebooks verspricht weitere Vorteile in folgender Hinsicht:

- Ausgewählte Dokumente können bereits durch eine besondere OCR-Teilerfassung entsprechend ihrer Bedeutung und Beschaffenheit weiter bearbeitet werden.
- Durch Aufbau einer einheitlichen Hardware- und Softwareumgebung kann Gruppenarbeit an Dokumenten besser unterstützt werden.
- Die gesamten Projektaktivitäten können dezentral von den Studierenden auf ihren Notebooks dokumentiert werden. Abschlussberichte werden individuell erarbeitet, von den Lehrenden angeleitet und vom Notebook der Gruppe präsentiert.
- Da Projekte oft über mehrere Semester veranstaltet werden, erleichtert die Aufsetzung eines Dokumentenservers mit mobilen Frontends, so wie er vom Kompetenzzentrum der Hochschule vorgehalten wird, die Übergabe von Teilergebnissen von einem Projektteilnehmer zum nächsten. So können Vorgänge, die am Ende eines Semesters noch nicht abgeschlossen worden sind, weiter bearbeitet werden.

Die Arbeitshypothesen werden bis Ende des Sommersemesters 2003 in vier Phasen (Technische Vorarbeiten, mobile Recherche zur Erstellung eines digitalen Archivs, mobile Recherche zu Forschungsarbeiten in (Universitäts)bibliotheken und im Friedrich Nietzsche Kolleg Weimar, Evaluation und Nutzerschulung) untersucht.

Zur Literaturrecherche melden sich die Benutzer (mobiler Benutzer und Teilnehmer der Expertenrunde) je nach Arbeitsumgebung an den bevorzugten bzw. nutzbaren Systemen an. Hernach können die Quellen digital erfasst, übermittelt und diskutiert werden. Ein besonderer Synergieeffekt der mobilen Recherche ergibt sich nun in der anschließenden Speicherung des Dokuments auf einem zentralen Dokumentenserver und der kurz- oder mittelfristigen Textauszeichnung.

Server- und Archivierungskonzept

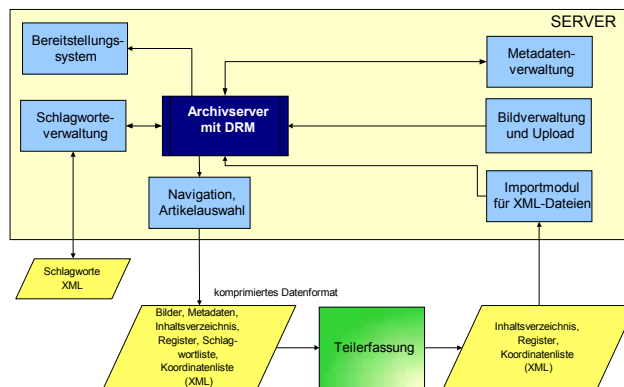


Abbildung 1: Server- und Archivierungskonzept

Um von der Digitalisierungsleistung des mobilen Benutzers zu profitieren, bietet es sich an, das erfasste Material in einem XML-konformen Datenformat mit den zugehörigen Metadaten auf dem zentralen Archivserver (BSCW) beim Kompetenzzentrum Digitale Medien (KDM) der Universität zu speichern. Von dort kann das Dokument sofort oder

zu einem späterem Zeitpunkt zur Ansicht oder zu einer Teilerfassung (Markierung) bezogen werden, wobei zur Wahrung von Autoren- und Verlagsrechten eine digitale Rechteverwaltung implementiert werden muss.

Technische Voraussetzungen und Umsetzung der Arbeiten

Clientseitig notwendige Hardware-Voraussetzungen sind ein Multimedia-PC bzw. ein Notebook mit WLAN-Karte, eine CCD-Kamera mit mindestens 4 Megapixel Auflösung (optional ein Scanner oder Buchscanner vor Ort) sowie eine funktionsfähige WLAN-Umgebung im Archiv. Zur Sprachübertragung sollte ein Headset verwendet werden. Die Softwareausstattung umfasst Chatsoftware, einen Web-Browser und ein Java-basiertes Teilerfassungsprogramm.

Bei der Implementierung ist zu beachten, dass restriktive Firewalls, welche die benötigten IP-Ports überwachen und filtern, die Router, welche die gestreamten Datenpakete verteilen, sowie die Maskierungsmechanismen die angestrebte Funktionalität der verwendeten Kommunikationssoftware erheblich einschränken können.

Aus Kostengründen und um trotz vielerorts unbekannter Sicherheitseinstellungen und Routerkonfigurationen möglichst viele Kommunikationsdienste bereit zu stellen, haben wir uns auf frei erhältliche und frei konfigurierbare Sprach- und Text-Chat-Lösungen in Client-/Server-Architektur konzentriert. Hierbei sollte zumindest der Bilddatenaustausch sowie der Text-Chat web-basiert möglich sein, um auch in sehr restriktiven Netzwerkumgebungen ein Minimum an Funktionalität zu gewährleisten.

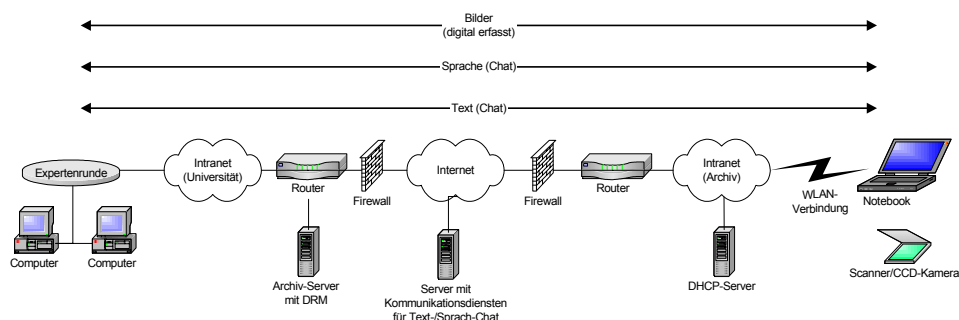


Abbildung 2: System (Netzwerksicht)

Nach einer umfangreichen Evaluation wurde das Programm TeamSpeak2 für die Sprachkommunikation gewählt, welches Mehrbenutzer-Kommunikation in guter Qualität ermöglicht. Ergänzend dazu arbeiten wir im laufenden Praxisprojekt an der Bereitstellung eines Text-basierten Chat-Systems sowie einer Bilddateiübertragungslösung.

Da sich das Projekt momentan noch in der Implementationsphase befindet, liegen zum jetzigen Zeitpunkt (Mai 2003) noch keine Erfahrungsberichte vor. Wir planen, eine Teilnehmerbefragung im späteren Verlauf des Sommers 2003 zur Evaluation zu nutzen, wenn Erfahrungen mit der Kommunikationssoftware bei Recherchen zu Abschlussarbeiten in Weimar vorliegen.

Ausblick

Wir haben in diesem Beitrag die Projektziele und erste Ergebnisse einer *Mobilen verteilten Dokumentenrecherche in Bibliotheken und Archiven* im Anwendungskontext des Duisburger Retrodigitalisierungsprojekts zur Nietzsche-Rezeption in Deutschland in den Jahren 1865 bis 1945 vorgestellt. Eine CD-basierte kleine Quellsammlung auf HTML-Basis sowie eine web-basierte Version sind fertiggestellt, erlaubten vielfältige Erfahrungen und sind Ausgangspunkt neuer Forschungsfelder. Mit der Erweiterung der Rezeptionsfelder gehen eine Standardisierung des Erfassungsvorgangs im Projekt zur mobilen Dokumentenrecherche in Archiven und die Verwendung standardisierter Dokumenttypen und XML-basierter Datenbanken einher. Fortschritte bei der Texterkennung im Bildmaterial lassen eine Teilerkennung von relevantem Text in Frakturschriftdokumenten als vielversprechenden neuen Zugang erscheinen. Durch Überwindung der Medienbrüche und Konzentration der Erfassungs-, Verschlagwortungs- und Auszeichnungsprozesse nach anerkannten Standards [BS95] kann die Dokumentenerfassung entscheidend verbessert werden. Wir erwarten von diesem Konzept deutliche Synergieeffekte, die einen Beitrag zur nachhaltigen Literaturrecherche und -archivierung versprechen.

Literaturverzeichnis

- [Bc03] Bick, A.: Ein Vorgehensmodell zur Retrodigitalisierung von Frakturtexten im Spiegel aktueller Retrodigitalisierungsprojekte, 2003, http://www.informatik.uni-duisburg.de/Studienprojekte/Nietzsche/Vorgehensmodell_Retrodigitalisierung_Nietzsche.pdf
- [Bi02] Biella, D.; Dyllong, E.; Kaiser, H.; Luther, W.; Mittmann, Th.: Wege zur digitalen Erfassung der Nachwirkung Nietzsches in Deutschland von 1865 - 1945. Ein Arbeitsbericht zum Duisburger Retrodigitalisierungsprojekt. Tagung „Vom Umgang Nietzsches mit Büchern zum Umgang mit Nietzsches Büchern“, Weimar, 23.9.-25.9.2002
- [BS95] Burnard, L.; Sperberg-McQueen, M.: TEI Lite: An Introduction to Text Encoding for Interchange, 1995, <http://www.tei-c.org/Lite/>
- [KKS03] Kalz, M.; Kerres, M.; Stratmann, J.: Notebooks in der Hochschullehre. Didaktische und strukturelle Implikationen, 2003, http://online-campus.net/edumedia/publications/ke-nuernberg_03.doc
- [Mi02] Miesenberger, K.: Bibliothèques électroniques et édition électronique pour les personnes privées d'accès à l'imprimé État de l'art et perspectives d'avenir, 2002, <http://www.snv.jussieu.fr/inova/villette2002/act4b.htm>
- [Mt01] Mittmann, Th.: Friedrich Nietzsche: Judegegner und Antisemitenfeind. Erfurt, 2001
- [Ju02] Weimarer Nietzsche-Bibliographie (WNB), bearb. von Susanne Jung, Frank Simon-Ritz, Clemens Wahle, Erdmann v. Wilamowitz-Moellendorff, Wolfram Wojtecki. Bd. 1: Primärliteratur 1867-1998. Stuttgart, Weimar, 2000 - Bd. 2-5: Sekundärliteratur 1867-1998, 2002.