

MyBib eRoom: Digitalisierte Literatur urheberrechtskonform bereitstellen

Vom Scan bis zur Bildschirmpräsentation durchgängig automatisiert

Helga Bergmann

Massendigitalisierung und webgestützte Präsentation des weltweit vorhandenen Wissens – diesen Aufgaben stellen sich Bibliotheken, Archive und Museen im Internet-Zeitalter. Die größten Herausforderungen dabei sind, den Produktionsprozess für Digitalisate effizient und bezahlbar zu machen und die Bestimmungen des Urheberrechtsgesetzes einzuhalten. Mit dem elektronischen Lesesaal MyBib eRoom ist es dem Bonner Software-Unternehmen ImageWare Components GmbH und dem Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme (IAIS) in Sankt Augustin gelungen, eine urheberrechtskonforme Präsentationsplattform für Digitalisate zu entwickeln. Ihre Entwicklungspartnerschaft haben sie Ende 2010 in einen langfristigen Kooperationsvertrag überführt. Im Rahmen ihrer Partnerschaft veranstalten ImageWare und Fraunhofer IAIS Seminare, in denen grundlegendes Fachwissen zu Rechts-, Organisations- und Technikfragen rund um den elektronischen Lesesaal vermittelt wird.

■ „Bibliotheken sind Orte der Wissensstrukturierung, der Wissensaufbereitung und der Wissensvermittlung“, erklärt Rolf Rasche, Geschäftsführer von ImageWare, und ergänzt: „Das werden sie auch bleiben.“ Heute allerdings sind sie vor die Aufgabe gestellt, das Wissen der Welt digital aufzubereiten: Gleichgültig ob Bücher vor dem Zerfall bewahrt, Unikate und schwer auffindbare Informationen oder neuestes Fachwissen zugänglich gemacht werden sollen. Die mit der Digitalisierung beschäftigten Bibliotheken müssen neben den technischen Herausforderungen auch juristische und organisatorische Anforderungen berücksichtigen. Für jüngere Publikationen gelten fast immer Urheberrechte. Aber auch ältere Werke, bei denen der Urheberrechtsschutz eigentlich abgelaufen ist, können in Sonderfällen wie-



In guter Tradition der Mönche wird am Fraunhofer IAIS das Kulturgut Wissen ins digitale Zeitalter übertragen, damit sich spätere Generationen noch genau so darin vertiefen können, wie der kleine Mönch am Aufgang des Schloss Birlinghoven, dem Sitz des Fraunhofer-Institutszentrum in Sankt Augustin.

der Copyright-behaftet sein – z. B. wenn das Vorwort neueren Datums ist und aus mehr als zwei Seiten besteht. Selbst Informationen, die nach dem neuen Prinzip des freien Zugangs zu Information Open Access (OA) publiziert werden, können bestimm-

Der elektronische Lesesaal „MyBib eRoom“ macht es möglich, die für Publikationen geltenden Urheberrechte auch bei der elektronischen Bereitstellung zu berücksichtigen. Die Anwendungsoberfläche bietet komfortable Funktionen, beispielsweise stufenloses Zoomen von Scans und Bildern.

ten Lizenzvorschriften unterliegen, wie beispielsweise in den „Creative Commons“ definiert.

Juristisch hieb- und stichfeste Lösungen gefragt

Die in großem Stil vorangetriebene Digitalisierung hat zu einer Reihe juristischer Probleme und zu zahlreichen Gerichtsprozessen geführt. Besonders beachtet war in diesem Zusammenhang die juristische Auseinandersetzung um das sogenannte Google Book Settlement. Ende März dieses Jahres hat das oberste Berufungsgericht der Vereinigten Staaten sein Urteil gefällt: Es untersagt Google das Kopieren von Büchern ohne explizite Genehmigung der Rechteinhaber. Obwohl das deutsche Urheberrecht derzeit nicht unmittelbar betroffen ist, rechnet Rolf Rasche mit weitreichenden Konsequenzen: „Das Urteil des amerikanischen Gerichts bedeutet, dass Google nun Einzelgenehmigungen einholen muss – ein Riesenaufwand. Es wird aber auch dazu führen, dass ein Mehrbedarf an juris-

tisch abgeprüften Systemen entsteht.“ Mit dem elektronischen Lesesaal MyBib eRoom haben ImageWare und die Abteilung Net-Media des Fraunhofer IAIS eine Präsentationsplattform entwickelt, die deutsches und europäisches Urheberrecht berücksichtigt.

ImageWare und Fraunhofer IAIS bündeln ihr Know-how

Bei der Realisierung der MyBib eRoom-Lösung brachte ImageWare ihre Kompetenz als Entwickler von Workflow-Lösungen zur Digitalisierung und Erhaltung von Kulturgütern ein. Die Bonner haben ein fast 15jähriges Know-how beim Erstellen, Verwalten und Liefern von Digitalisaten. Das Team des Fraunhofer IAIS steuerte seine Kompetenzen zur Analyse von multimedialen Dokumenten sowie neue Erkenntnisse und Entwicklungen bei, die im Rahmen des THESEUS-Programms im Anwendungsfall CONTENTUS erarbeitet wurden: In Summe automatisierte Prozesse zur Bereitstellung, Erschließung und Indexierung von multimedialen Wissen. In Projekten mit Zeitungsverlagen haben die Fraunhofer-Wissenschaftler unter anderem automatische Analyseverfahren entwickelt, die Layout-Strukturen in Dokumenten erkennen, diese inhaltlich erschließen und mit semantischen Metadaten anreichern. In dem gemeinsamen Projekt MyBib eRoom mit ImageWare wurden diese Arbeitsabläufe unter anderem an die Bedürfnisse von Bibliotheken angepasst und um neue Funktionen erweitert.

Das Resultat: Der elektronische Lesesaal

Mit MyBib eRoom wurde eine Präsentationsplattform entwickelt, die es Bibliotheken, Archiven und Museen erlaubt, unterschiedlichste Medien elektronisch in einem geschützten Bereich bereitzustellen und dabei die engen Grenzen des Urheberrechts zu berücksichtigen. Damit ist für Bibliotheken das Dilemma gelöst, ihre Kunden zeitgemäß zu bedienen, ohne dabei das Urheberrecht zu verletzen. Nutzer haben am Leseplatz in der Bibliothek – nach einer von der Bibliothek vergebenen Zugangsberechtigung, die von der MyBib eRoom-Software noch einmal in einem Authentifizierungsverfahren verifiziert wird – elektronischen Zugriff auf digitalisierte Lehrbücher sowie auf Kopien aus Campus-Lieferdiensten. Ein integrierter Abgleich zwischen MyBib eRoom und dem lokalen Bibliothekssystem gewährleistet die Übereinstimmung der Anzahl physisch vorhandener und elektronischer Exemplare („Exemplarzähler“), über die die Anzahl der gleichzeitig zu lesenden elektronischen Kopien gesteuert wird.



Verknüpfen ihr wissenschaftliches und wirtschaftliches Know-how für den elektronischen Lesesaal: Die Wissenschaftler von Fraunhofer IAIS, Dr. Stefan Eickeler (l.), Ulrich Wernecke (2.v.l.h.) und Dr. Stefan Paal (2.v.r.) und die Fachleute aus der Wirtschaft, Rolf Rasche, Geschäftsführer von ImageWare (r.) und Roland Keck, Leiter Support und Fachberatung

Produktionsprozess und Repository

Der durchgängige Produktionsprozess beginnt mit dem Scanvorgang, für den sich jeder Buch-, Flachbett- oder Durchlaufscanner eignet. Die erzeugten Scans (Images) werden in die Scan- und Capturing-Software BCS-2® von ImageWare importiert. Dort erfolgt die OCR-Erkennung und Indexierung, ggf. des kompletten Volltextes sowie die Übergabe zur Speicherung in einem Repository.

Der eigentliche Zugriff auf die Digitalisate erfolgt im Allgemeinen aus dem loka-

len OPAC. Dazu werden die Metadaten dieses Systems mit den Digitalisaten verlinkt. Zur Erhöhung der Datensicherheit werden die Digitalisate über eigene Verfahren komprimiert und verschlüsselt in den MyBib eRoom Viewer übertragen, der für die Anzeige der Images genutzt wird.

Stufenloser Zoom für Scans und Einzelbilder

MyBib eRoom verfügt über stufenlose Zoom-Funktionen. Volltextsuche unterstützt den Bibliothekskunden bei der



Intensive Kommunikation zwischen Wissenschaft und Wirtschaft: Die Kooperationspartner ImageWare und Fraunhofer IAIS treffen sich regelmäßig zum Informationsaustausch. Im Bild Dr. Stefan Paal (vorne), Ulrike Holtkamp, Marketingreferentin von ImageWare und Ulrich Wernecke, bei Fraunhofer zuständig für die Geschäftsfeldentwicklung Medialayse- und Archivsysteme.

Recherche. Begriffe aus der Suchformulierung (Anfrage) werden an den Fundstellen in den Dokumenten hervorgehoben (Search Term Highlighting). Der Benutzer kann durch die vom System angebotenen Funktionen auch in großen Datenbeständen intuitiv navigieren. Er kann aber weder das Dokument noch Dokumentteile kopieren, ausdrucken oder speichern. Das System ermöglicht auch eine zeitliche Begrenzung des Zugriffs, wie es das Urheberrecht für die Bereitstellung von Kopien fordert.



Effizienter durch Scan-Roboter und Spezialsoftware: Vergleichbar den Produktionsstraßen in der Industrie entwickelt sich die Herstellung von Digitalisaten zunehmend zu einem durchgängig automatisierten Prozess vom Scan bis zur Präsentation des Dokumentes am Bildschirm

Nach Ablauf der Frist wird das Dokument automatisch vom Server genommen. Durch systemimmanente, anonymisierte statistische Auswertungen unterstützt MyBib eRoom auch Verwaltungsarbeiten in der Bibliothek.

Auszeichnung für automatische Layoutanalyse

Die neuesten Entwicklungen im Digitalisierungs-Workflow, die in das Fraunhofer-/ImageWare-Verfahren integriert werden können, sind automatische Layoutanalysen und die semantische Texterschließung direkt vom gescannten Image. Das Verfahren setzt nach der automatischen Verbesserung des Scans (Entzerrung, Falzkorrektur etc.) an, erkennt nach zuvor trainierten Regeln Überschriften, Untertitel, Spalten, Bilder usw. und segmentiert die Seiten für eine strukturierte Ablage. Die Wissenschaftler vom Fraunhofer IAIS um Dr. Stefan Eickler wurden für dieses Verfahren zur automatischen Layoutanalyse von der größten internationalen Fachgesellschaft für technologische Innovationen, der IEEE, mit dem

ICDAR2009 Page Segmentation Competition-Preis ausgezeichnet.

Semantische Erschließung gescannter Daten

Im Rahmen des vom Bundesministerium für Wirtschaft und Technologie (BMWi) geförderten Forschungsprogramms THESEUS haben die Fraunhofer-Forscher im Anwendungsfall CONTENTUS Zeitungen, Zeitschriften und Bücher diesem Layoutanalyseverfahren unterzogen und auf die gewonnenen Daten eine automatische semantische Erschließung der Inhalte aufgesetzt. Die semantische Erschließung erkennt Begriffsbedeutungen, zum Beispiel Personen, Orte und Institutionen, markiert diese in unterschiedlichen Farben und verlinkt sie mit weiterführenden Informationsquellen, beispielsweise mit Wikipedia im Internet. Dr. Stefan Paal vom Fraunhofer IAIS erklärt, wie den Wissenschaftlern diese automatische Informationsklassifizierung gelungen ist: „Für die automatische Erschließung haben wir Entitäten extrahiert und mit dem Kontext, in dem sie vorkommen, verknüpft, um mehrdeutige Begriffe bei der Informationssuche später klar zuordnen zu können. Ein typisches Beispiel: Michael Jackson. Prominente mit diesem Namen waren neben der Pop-Ikone beispielsweise ein US-General, zwei Fußballspieler, ein Whisky-Tester (Autor) ...“ Die vollautomatische Erschließung sei natürlich nie perfekt, so Paal, aber es sei gelungen, einen sehr hohen Anteil der Ermittlung von Bedeutungsstrukturen anhand von Algorithmen (Data Mining / Knowledge Discovery) zu automatisieren. „Was mit der Massendigitalisierung an Datenmengen produziert wird, verlangt nach automatischen Verfahren.“ Mit diesen Worten hatte Rolf Rasche vor über zwei Jahren zum ersten Mal den Hintergrund der Forschungsarbeiten beschrieben. Inzwischen veranstalten die Experten von Fraunhofer IAIS Net Media und ImageWare mit Rechtsanwälten und Rechtswissenschaftlern gemeinsam Seminare, in denen sie grundlegendes Fachwissen zu Rechts-, Organisations- und Technikfragen rund um den elektronischen Lesesaal vermitteln.

AUTORIN

HELGA BERGMANN-OSTERMANN

Journalistin
Am Stockacker 9
94508 Schöllnach
h.bergmann-ostermann@t-online.de

Seminare zum „elektronischen Lesesaal“ MyBib eRoom“

Das Wissen, das notwendig ist, um eine fundierte Entscheidung zu treffen, ob eine bibliothekseigene Produktionsstraße für Digitalisate sinnvoll ist, wird von Experten von Fraunhofer IAIS Net Media und ImageWare in den Seminaren zusammen mit Rechtsanwälten und Rechtswissenschaftlern vermittelt. Dabei geht es um grundlegendes Fachwissen zu Rechts-, Organisations- und Technikfragen rund um den elektronischen Lesesaal. Die Veranstaltungen sind primär für Entscheider aus Bibliotheken, Archiven, Museen und Verlagen konzipiert. Auf der Agenda stehen Themen wie die „Bibliothekarische Motivation für die Einrichtung eines elektronischen Lesesaals“, Grundlagen des § 52b UrhG sowie die technische Umsetzung und Demonstration des elektronischen Lesesaals MyBib eRoom. Diese Lösung schützt sowohl Rechteinhaber als auch Nutzer vor Daten-Missbrauch und eröffnet die Möglichkeit, unter Einhaltung des Urheberrechts unterschiedlichste Medien sicher bereit zu stellen.

Die Veranstaltungen finden am Fraunhofer IAIS in Sankt Augustin, Schloss Birlinghoven, statt. Die Teilnehmerzahl ist auf maximal 15 begrenzt. Für Bibliotheken, Archive und Körperschaften des Öffentlichen Rechts kostet ein Platz 175,- € inkl. MWST.

Die nächsten Termine sind der 30.06.2011 und der 27.10.2011. Weitere Veranstaltungsinformationen gibt es auf der Webseite von ImageWare <http://www.imageware.de/de/termine/schulungen-und-workshops/>.

Anmeldungen unter
ImageWare Components GmbH
Am Hofgarten 20
53113 Bonn
info@imageware.de
www.imageware.de