



Einstieg in die Kataloganreicherung an der SUB Hamburg

Motivation und
Projekterfahrungen



„Wer sind wir?“

-1-

- Funktion : Zentrale Ausleihbibliothek in einem zweischichtigen Bibliothekssystem mit der Sonderrolle Landesbibliothek
- Bestand: 4,3 Mio. Einheiten mit jährlichem Zugang von ca. 76 Tsd. Einheiten
- Benutzung: Überwiegend klassische Magazinbibliothek mit großer Bedeutung des Katalogzugangs

„Wer sind wir?“

-2-

- SSGs der Bibliothek: Politik- und Friedensforschung, Verwaltungswissenschaften, Spanien/Portugal, Indianer- und Eskimosprachen und -kulturen, Küsten- und Hochseefischerei
- Projektschwerpunkt u.a.: Seit 2000 starke Beteiligung am Aufbau Virtueller Fachbibliotheken

„Warum setzen wir uns für Katalog- anreicherung ein?“

-1

Übertragene Verantwortung für die SSGs:

- Überregionale Sichtbarkeit der SSG-Bestände für ein deutschlandweites Fachpublikum verbessern
- Den Mangel der fehlenden Einsicht vor Ort ausgleichen
- Die hohe Nutzung der OLC-Ausschnitte motiviert zu vergleichbarem Servicelevel für Monos und Sammelwerke

Ziel: Optimale Steuerung der überregionalen Literaturversorgung in den von uns verantworteten Fächern (Fehlbestellungen verhindern/größere Treffermengen durch mehr Metadaten)

„Warum setzen wir uns für Katalog- anreicherung ein?“

-2-

- Lösung 1: Ausweitung der Erschließungsarbeit an den SSG-Bibliotheken => scheitert an zu hohen Kosten
- Lösung 2: Erweiterter Geschäftsgang mit Scan + OCR von Inhaltsverzeichnissen und Titelblättern
- Lösung 3: Scan + OCR der gesamten Werke bzw. Einbindung von E-Books => künftiges Ziel mit erkennbaren Ansätzen und aktuell noch zu vielen ungelösten Brocken (Authentifizierung, Autorisierung, DRM ...)

„Welche konkreten organisatorischen Ziele waren uns wichtig?“

-1-

- Anschubfinanzierung durch den GBV ermöglicht Start mit Dienstleister, um zügig eine kritische Masse zu produzieren
- Retro-Aktion für die letzten 10 Jahre der großen SSGs "Spanien/Portugal" und "Politikwissenschaft / Verwaltungswissenschaften" => 44.000 Titel ergänzt um 8000 Titel der LINGA-Spezialbibliothek
- Dienstleistungspaket schließt Ausheben der Bestände ein

„Welche konkreten organisatorischen Ziele waren uns wichtig?“

-2-

- Einkauf eines "Rundum-Sorglos"-Paketes mit automatisierter Datenübergabe an den Verbundkatalog und Übernahme von Hard- und Software vom Dienstleister nach Projektende
- Strikte Beachtung der Kostenseite für die Übernahme einer Lösung für den nachfolgenden Regelbetrieb

„Wie bewerten wir den Projektverlauf? Wie sind wir im Alltag angekommen?“ -1-

- Die äußere Organisation hat sich schnell zurechtgerückt und das Projekt konnte vorzeitig abgeschlossen werden.
- Keine Kostensteigerung, Gesamtkalkulation hatte Bestand
- Anbindung an den GBV musste noch optimiert werden - läuft jetzt reibungslos und auch seitens des GBVs zeitnah

„Wie bewerten wir den Projektverlauf? Wie sind wir im Alltag angekommen?“ -2-

- Übernahme von Scanstation und die Aufnahme des Regelbetriebs erfolgreich/ Support gut organisiert
- Die Entscheidung für die Einbindung des Scannens am Ende des Geschäftsgangs hat sich bewährt.
- Weitere Lizenzkosten belasten den Bibliotheksetat

„Was brauchen wir für die nächste Qualitätsstufe?“

-1-

*Prämisse: Nutzer erwartet für alle Titel
Inhaltsverzeichnisse und deren
Recherchierbarkeit im Katalog*

Unsere Erwartungen:

- Lokale Katalogsoftware mit Recherchemöglichkeit für die indexierten Inhaltsverzeichnisse
- Alle SSG-Bibliotheken beteiligen sich am Zusammentragen der Inhaltsverzeichnisse

„Was brauchen wir für die nächste Qualitätsstufe?“

-2-

- Die DNB deckt - wie angekündigt - die deutsche Verlagsproduktion ab
- Voranschreiten der DNB und der Verbände bei der Etablierung eines sinnvollen und schnellen Austauschkonzeptes für die Referenzen und der Indexbegriffe zu den Scans

Digitalisierungsprojekt an der SUB Hamburg: Erfahrungsbericht aus der Perspektive des Scanteams

Joel Gecelter, Hamburg

Von Ende Februar bis Mai 2007 nahm ich den Digitalisierungsprojekten der SUB Hamburg teil und berichte aus der Perspektive des Projektteams, dessen Aufgabe in der Digitalisierung von am Ende 52.000 Inhaltsverzeichnissen war. Begonnen hatte das Projekt mit einem Volumen von ursprünglich 42.000 Titeln.

Das von AGI ausgewählte Team bestand aus 4 weiblichen und 4 männlichen Bibliothekaren, Dokumentaren, Archivaren und mir, ausgebildeter Tontechniker und Lebensgefährtin einer Bibliothekarin. Allen war somit der Raum „Bibliothek“ wohl vertraut, nicht nur als Nutzer. Dennoch war die Größe der SUB Bibliothek und ihres riesigen Magazins eine neue, interessante Erfahrung. Anfangs wurde das Magazin auch mal zur „düsteren Kluft“, wenn die Stromsparschaltung mal wieder plötzlich das Licht ausdrehte. Ein freundlicher Hausmeister löste das Problem. Oder wenn man sich in den ersten Tagen in den vielen Gängen und Gebäudeteilen verlor.

Das Team arbeitete von 6:30 morgens bis 21 Uhr am abend - fast pausenlos, freundlich kooperativ und sehr konzentriert - was wohl machen Bibliothekar der SUB angesichts der Dauer zum Staunen brachte. Die Stimmung im Team kann als recht gut bewertet werden.

Wesentliche Voraussetzung dieser erfreulichen Arbeitssituation waren zwei Faktoren. Erstens die Einschulung, für fast alle war die gesamte Software-Umgebung bis auf MS-Windows neu. Die Schulung endete auch während des Projektes nicht, denn durch einen Realtime-Online-Support konnten Fragen zum Umgang mit besonderen Büchern, Ergänzungen von Workflows oder am Anfang nach noch nicht ganz verständliche Meldungen der Software während der gesamten Arbeitszeit dem Backoffice von AGI direkt gestellt werden. Die Antworten kamen jeweils sofort entweder aus München von Herrn Herder oder aus Neustadt an der Weinstraße von Herrn Hauer. Doch auch das Team konnte sich über diese Realtime-Kommunikation jederzeit intern unterhalten und half sich so auch wechselseitig.

Jeweils drei von uns waren gleichzeitig aktiv - verteilt auf verschiedene Etagen oder ohne Blickkontakt direkt zwischen den Magazin-Regalen. Es gab einen gemeinsam vereinbarten Zeitplan, den das Team aber nach Bedarf intern modifizieren konnte. Zu Ausfällen kam es nie, irgendjemand schloß sofort die Lücke, sofern ein Teammitglied verhindert war oder tauschen wollte.

Der zweite Grund für die gute Stimmung und das wohl recht erfreuliche Projektergebnis - sowohl hinsichtlich Menge und Qualität - war die Gesamt-Lösung „intelligentCAPTURE mobile“. Das Software-Programm besteht aus mehreren Teilprogrammen, die ineinander verzahnt sind und jeweils gleichzeitig und parallel laufen. Bevor wir jedoch intelligentCAPTURE kennen lernten, mussten alle die Realtime-Kommunikation kennen und anwenden lernen. Denn Probleme treten mit Computern immer auf - und dann muss man sich zu helfen wissen. Dies ist tief unten in Magazinen ohne Telefon in der Nähe und ohne Handy-Verbindung nicht gerade leicht - es sei denn man hat intelligentCAPTURE.

intelligentCAPTURE basiert auf IBM Lotus Notes & Domino zu dem ein Chatprogramm namens Sametime gehört. Damit kann man nicht nur Mitteilungen schreiben und sofort eine Antwort erhalten, sondern es bietet die Möglichkeit kinderleicht ausgeschnittene Bildschirmausschnitte dem Partner zu zeigen, z.B. eine Fehlermeldung, einen Scan oder eine Konfiguration. Chat mit der Ausschneidefunktion genügt für die meisten Problemfälle. Manche hatten auch Spaß an den Emotions - um Stimmungen wie Freude, Überraschung oder Ärger zu zeigen.

Doch für eine Online-Nachschulung, Software-Überprüfung oder Diskussion waren zusätzlich Sprachkommunikation und Application Sharing integriert. So war es z.B. dem IT-Support aus München möglich, die Tastatur aus der Ferne zu übernehmen und Problemfälle selbst nachzuspielen und dabei den Debugger - ein Programm zur Kontrolle des Programmcodes - zu nutzen.

Die guten Stereoheadsets nutzen einige von uns auch schlicht für die Unterhaltung während der Arbeit, sei es für Small Talk im Team, um sich über Hörbücher nebenbei weiterzubilden oder einfach um Musik oder Radio zu hören. Denn die eigentliche Arbeit hatten alle schnell verstanden und beherrscht.

intelligentCAPTURE arbeitet sehr konsequent wissensbasiert und versucht Fehler von Menschen bei der Bedienung möglichst erst gar nicht entstehen zu lassen.

Automatisch wurde das jeweils nächste zu bearbeitende Buch am Bildschirm angezeigt. Die von der SUB gelieferten Listen waren nach der Ordnung im Regal sortiert worden, so dass man meist nicht lange Suchen musste - weder beim Herausnehmen noch bei der Rückstellung. Dazu trug auch wesentlich das „mobile“ von intelligentCAPTURE mobile bei. AGI führte im Verlauf der ersten Projektwochen die ersten drei mobilen Scanstationen ein. Sie können zwischen den Kompaktmagazinen eingesetzt werden - das ist zwar eng, ja oft schon etwas klaustrophobisch eng, aber es funktionierte gut. Ein 50 langes Stromkabel und eine WLAN-Anbindung erlaubte eine hohe Flexibilität. Die WLAN-Anbindung war inmitten dieser Stahlregale nicht immer optimal, doch mit Geschick konnte man die Verbindung halten.

Das Scanning mit den Flachbettscannern ging recht flott. Jeweils die erste Seite eines Buches - das interne Titelblatt - diente zur Fixierung eines Scanausschnitts und mit etwas Geschick, passte dieser Rahmen ohne Nachkorrektur oder Verschieben auch für die Folgeseiten. Die individuelle Leistung der Teammitglieder war unterschiedlich.

Zwischen uns männlichen Teammitgliedern gab es einen Wettstreit, wer wieoft mehr als 400 Bücher pro Tag scannen konnte. Drei lagen Kopf an Kopf. Mittlerweile liegt der Rekord bei 624 Inhaltsverzeichnissen pro Tag - ich bin der Rekordhalter, dieser Sieg gelang mir aber erst in einem Folgeprojekt an der ZBW in Kiel in 9 Stunden.

Die wesentliche Aufgabe bestand also in der Auffindung der Bücher und des präzisen Scannings. Die Imaging-Software bietet Optionen zur Verbesserung, die man schnell versteht und vor allem auch schnell anwenden lernt. Ich meine, die Bedienung ist denkbar einfach und der Mitarbeiter am Scanner wird durch die verschiedenen Arbeitsschritte gut geführt. Im Scanfenster werden die gesamten Seiten in der Vorschau angezeigt und können, je nach Beschaffenheit des Mediums, bearbeitet und korrigiert werden.

Doch man musste trotz aller Software-Unterstützung aufpassen, denn nicht alle vom PICA-Server abgerufenen bibliografischen Angaben stimmten mit dem tatsächlichen Buch überein. Vor allem gab es bei den Sprachen Unstimmigkeiten - insgesamt wurden Titel aus 23 Sprachen verarbeitet. Sehr wichtig für gute OCR und maschinelle Indexierung ist die korrekte Spracheinstellung. Deshalb war unser Team auch bewußt nach Fremdsprachenkenntnissen ausgewählt worden. Da Spanisch und Portugiesisch einen hohen Anteil hatte, mussten die meisten diese Sprachen können - dies war übrigens meine Chance, in das Projekt zu kommen.

Die Qualitätskontrolle war über statistische Maßangaben sehr schnell überblickbar und nach ersten Anfängerfehlern oder Schwächen, waren wir schnell alle so trainiert, dass die Qualität der Images und damit auch des OCR-Ergebnisses wohl geschätzt wurde. Angangs und zwischendrin stichprobenhaft konnte der Support von AGI „mitlesen“, denn alle 10 Minuten wurden die Ergebnisse an einen zentralen Server in Göttingen beim GBV übermittelt. AGI nutzt den GBV als Hosting-Center.

Die eigentliche OCR-Verarbeitung, Erstellung der drei Dateitypen TIF, TXT und PDF und auch die Erstellung der maschinellen Indexierung lief vollautomatisch parallel zum Scanning. Sie musste nur morgens gestartet werden und war kurz nach dem letzten Scan am Abend auch schon fix und fertig.

Die Datenexporte zum Server des GBV lief als einziger Job nicht parallel - wir nutzen dafür unsere Pausen. Direkt danach erfolgte der Export zu dandelon.com. D.h., die produzierten Ergebnisse waren jeweils am gleichen Tag in dandelon.com schon recherchierbar. Die Einspielungen in den PICA-Katalog durch den GBV dauerten länger. Auch die sofortige Sichtbarkeit und Suchbarkeit der persönlichen Arbeitsergebnisse im Internet bei dandelon.com trug zu Motivation bei.

Noch wichtiger für die Motivation war aber das Resultat der Arbeit, der hohe Durchsatz pro Stunde führte bei den Schnellen im Team zu erfreulichen Honoraren, denn die Bibliothek bezahlte AGI pro Stück und genauso wurden auch alle Teammitglieder bezahlt. Schon bald lag die tägliche Produktionsmenge der drei Stationen zusammen bei durchschnittlich 1.000 Titeln. Schon nach 6 Wochen war das Projektziel erreicht - deutlich schneller als geplant. Mit einer einfachen

Statistikfunktion konnte man jederzeit sehen, wieviel man pro Tag oder Monat bereits verdient hat. Mit jedem Export wurde die Statistik automatisch aktualisiert.

Das Projekt lief gut, manchmal für die Bibliothek zu schnell, denn sie hatte den Anteil der ausgeliehenen Titel nicht ausreichend bei der Produktion der ursprünglichen Listen berücksichtigt und musste mehrfach Folgelisten produzieren, um die gewünschte Anzahl von 42.000 Titeln zu erreichen - nur am Rande haben wir dies mitgekomen. Nennwerte Lücken gab es in unserer Produktion jedoch nie. Im Hintergrund importierte die AGI-Projektleitung die Folgelisten und sie erschienen bei uns automatisch. Zusätzlich nutzten wir aber auch ausgedruckte Listen, insbesondere dort, wo wir mit Bücherwagen ausheben mussten oder wo die Bücher sehr weit auseinander standen.

Unsere Koordination mit der Bibliothek war eher spärlich - oder besser unkompliziert. Wir arbeiteten komplett selbstständig und mussten nur gelegentlich wegen einer Ausleihe den Magazin-Mitarbeitern etwas Platz machen oder eine Buch weiterreichen. Jeder Mitarbeiter erhielt vom Wachdienst für die jeweilige Schicht einen Schlüssel und konnte sich dann frei in den geschlossenen Bibliotheksbereichen bewegen. Nur am Anfang erhielten wir eine einmalige Bibliothekseinweisung.

Nach dem ersten Projekte folgte gleich noch ein Anschlußprojekt in der Linga-Bibliothek, einer im gleichen Gebäudekomplex assoziierten Bibliothek und schließlich spendierte die SUB noch einen Nachschlag, um die mittlerweile eingetroffenen Neuerscheinungen und die Rückläufer aus den automatisch genierten Vormerkungen nachzuarbeiten.

Wir alle waren mit dem Projekt sehr zufrieden und einige von uns, nahmen dann an weiteren Projekten in Braunschweig, Kiel, Göttingen, Mainz und Berlin teil. Ein digitalisierender Wanderzirkus ;-)

Heute stehe ich übrigens hier, weil ich mittlerweile dauerhaft für AGI arbeite als First Level-Support und für die Schulung neuer Kunden. Vielleicht sehen wir uns ja bald wieder.