

Anlage 4 Kosten für die Erfassung eines Standardbuches (Ecker)

In dieser Ausarbeitung wird eine Aufstellung der Kosten für das Verfilmen, Scannen, Speichern, Indexieren und teilweise inhaltliche Erschließen (Texterfassen) eines typischen Buches gegeben. Außerdem wird gezeigt, welche Aufgaben sich für eine Vergabe an einen professionellen Dienstleister eignen und welche Aufgaben von der Bibliothek selbst durchgeführt werden sollten.

Die aufgeführten Zeitangaben und Kosten sind aus der Praxis abgeleitete Werte für durchschnittliches Material. In Sonderfällen können erhebliche Abweichungen auftreten.

1 Spezifikationen

1.1 Durchzuführende Arbeiten

Die durchzuführenden Arbeiten werden wie folgt festgelegt:

Scannen:	400 dpi, Schwarz/Weiß-Modus ohne Graustufen
Speichern:	TIFF G4-Format auf CD-R (incl. Index)
Filmen:	35 mm Silberhalogenid-Archivrollfilm unter Berücksichtigung der einschlägigen DIN Normen
Indexieren:	Bibliographische Kerndaten
Inhaltsererschließen:	Texterfassung von Inhaltsverzeichnis, Kapitelüberschriften, Register, ggf. auch Volltext

Die Bücher müssen pfleglich behandelt werden; sie dürfen nicht geschnitten oder beschädigt werden. Es werden daher spezielle Buchscanner und Kameras mit Buchwippe eingesetzt.

Die Scanimages sollen den Zustand der jeweiligen Seite möglichst originalgetreu wiedergeben. Daher werden bei Bedarf die Imageseiten in einer Nachbearbeitung geradegerückt und von erfassungsbedingt enthaltenen dunklen von Schatten oder Strichen (Kanten, Falz) gesäubert. Bei doppelseitiger Aufnahme werden die Seiten getrennt und sequentiell sortiert.

Eine interaktive Entfernung von Grauschleier, der sich aus der Aufnahme von vergilbten Seiten ergeben kann, wird nicht vorgesehen. Wohl jedoch eine teilweise Säuberung durch automatisierte Prozeduren, soweit diese vorhanden sind.

Die Indexierung erfolgt durch maschinelle Übernahme von elektronischen Katalogdaten.

Die Texterfassung von Inhaltsverzeichnis, Kapitelüberschriften, Register erfolgt mit einer Genauigkeit von >99,985% (0,015% Fehlertoleranz); der Volltext wird mit 99,7% Genauigkeit erfaßt.

Zur Sicherung der erforderlichen Qualität werden alle Arbeitsergebnisse einzeln überprüft. Nach Abschluß der Erfassung erfolgt eine Endkontrolle.

Der durchschnittliche Speicherbedarf je gescannte Seite wird auf 150 KByte geschätzt. Bei stark vergilbten Buchseiten kann sich ein wesentlich höherer Speicherbedarf ergeben. In diesem Fall wären unter Umständen zusätzliche Reinigungsschritte vorzusehen.

1.2 Spezifikation des Standardbuches

Dieses Standardbuch wird folgendermaßen spezifiziert:

Spezifikation des Standardbuches	
Format:	Quart, mindestens A5, höchsten A4
Umfang:	300 Seiten Text zu jeweils 4000 Zeichen
Inhaltsverzeichnis:	3 Seiten zu jeweils 2000 Zeichen
Register:	6 Seiten zu jeweils 2000 Zeichen

2 Alternativen für die Vorgehensweise

Für das Scannen und Verfilmen bieten sich folgende Vorgehensweisen an:

- a) Getrenntes Scannen und Verfilmen vom Original
- b) Verfilmen des Originals und Scannen des Films
- c) Scannen des Originals und Erzeugung einer Filmkopie nach dem COM-Verfahren

Die diesen Vorgehensweisen entsprechenden Arbeitsschritte werden im folgenden Abschnitt beschrieben.

3 Arbeitsschritte

3.1 Auswahl der zu erfassenden Bücher

Arbeitsinhalt	Die zu erfassenden Bücher werden (z.B. nach Inhalt, Erhaltungszustand, urheberrechtlichen Gesichtspunkten, Bedarf etc.) ausgewählt und bereitgestellt.
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Bibliothek ausgeführt werden
Qualifikation	Diplom-Bibliothekar
Arbeitsaufwand	unterschiedlich (z.B. 5 Minuten)

3.2 Bibliothekarische Vorbereitung

Arbeitsinhalt	Das zu erfassende Buch wird unter bibliothekarischen Gesichtspunkten für das Scannen/Verfilmen vorbereitet
Details	<ul style="list-style-type: none"> • Identitätsprüfung (Titel, Autor, ...) und Abgleichung mit den Katalogdaten • Prüfen, ob das Buch aus inhaltlichen und technischen Gründen gescannt werden kann • Festlegung der Indexierung • Festlegung der zu erfassenden Seiten. • Festlegung von Sonderpaginierungen • Festlegung der Bearbeitung von Überformaten oder verkleinerten Seiten (Miniprints) • Festlegung der Erfassung von Seiten mit Abbildungen in Farbe , SW-Halbton oder Rasterung • Bearbeitung von Errata/Corrigenda • Hinweis auf beigelegte Sonderseiten • Bibliotheksseitige Dokumentation
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Bibliothek ausgeführt werden
Qualifikation	Studentische Aushilfe unter fachlicher Anleitung durch einen Diplom-Bibliothekar
Arbeitsaufwand	15-30 Minuten (realistischer Durchschnittswert: 20 Minuten)

3.3 Technische Vorbereitung

Arbeitsinhalt	Das zu erfassende Buch wird unter technischen Gesichtspunkten für das Scannen/Verfilmen vorbereitet
Details	<ul style="list-style-type: none"> • Erfassung der Buchstruktur • Ggf. Anfertigung von einzelnen Zwischenkopien • Festlegen der technischen Erfassungsparameter (z.B. Schwellwert, Maskenhintergrund, ..)
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Erfassungsstelle ausgeführt werden
Qualifikation	Operator
Arbeitsaufwand	Kann sehr stark differieren, besonders wenn Zwischenkopien erstellt werden müssen. Realistischer Erfahrungswert bei durchschnittlich 8 Minuten.

3.4 Indexierung

Arbeitsinhalt	Eingabe von Begriffen (z.B. Minimalsatz bibliographischer Daten) nach Vorgabe von Abschnitt 3.2 um das Buch suchbar zu machen
Details	<ul style="list-style-type: none">• Maschinelle Verknüpfung mit Katalogdaten• Manuelle Eingabe
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Erfassungsstelle ausgeführt werden
Qualifikation	technischer Mitarbeiter
Arbeitsaufwand	5 Minuten

3.5 Erfassung

a) Getrenntes Scannen und Verfilmen vom Original

Arbeitsinhalt	Das Buch wird gescannt und getrennt verfilmt
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Erfassungsstelle ausgeführt werden
Qualifikation	technischer Mitarbeiter
Arbeitsaufwand	60 Minuten

b) Verfilmen des Originals und Scannen des Films

Arbeitsinhalt	Das Buch wird verfilmt; der resultierende Film wird in einem zweiten Schritt eingescannt
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Erfassungsstelle ausgeführt werden
Qualifikation	technischer Mitarbeiter
Arbeitsaufwand	30 Minuten

c) Scannen des Originals und Erzeugung einer Filmkopie nach dem COM-Verfahren

Bei dem COM-Verfahren (Computer Output on Microfilm) werden digitale Daten per Laserbelichtung auf einen Mikrofilm „geschrieben“. Das Ergebnis entspricht einem aus der Originalvorlage photographisch erstellten Film.

Dieses billige und qualitativ sehr hochwertige Verfahren hat sich in den letzten Jahren in vielen Anwendungsbereichen auf Basis von 16-mm-Film durchgesetzt. COM-Belichter für den 35-mm-Film sind noch neu und wegen der daher noch hohen Gerätekosten nur dort wirtschaftlich, wo es um die Erzielung höchster Qualitäten geht. Es wird jedoch erwartet, daß dieses Verfahren in wenigen Jahren gerade im Bibliotheksbereich verstärkt angewandt werden wird.

Arbeitsinhalt	Das Buch wird gescannt; die resultierende digitale Datei wird in einem zweiten Schritt zusätzlich auf 35-mm-Film ausgegeben
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Erfassungsstelle ausgeführt werden
Qualifikation	technischer Mitarbeiter
Arbeitsaufwand	50 Minuten

3.6 Nachbearbeitung

a) Getrenntes Scannen und Verfilmen vom Original

Im Normalfall entbehrlich.

b) Verfilmen des Originals und Scannen des Films

Der Nachteil dieser Vorgehensweise liegt vor allem darin, daß beim Verfilmen von gebundenen Vorlagen dunkle Streifen am Rand und (bei doppelseitiger Aufnahme) am Falz entstehen. Außerdem läßt sich eine Verkantung der Seiten kaum vermeiden.

Die beim Einscannen des Films erhaltenen digitalen Dateien entsprechen damit jedoch nicht mehr den üblichen Anforderungen. Es sind folgende Arbeiten erforderlich: Trennen von Doppelseiten; sequentielles Sortieren der resultierenden Einzelseiten; Entfernen von Schatten und Falzrändern; lotrechtes Ausrichten der Satzspiegel. Diese Arbeiten können interaktiv oder automatisiert erfolgen.

Automatisierte Verfahren erkennen Randstreifen sowie die Lage des Satzspiegels mittels regressionsanalytischer Verfahren.. Bei manuellen Verfahren werden vom Bearbeiter die zu bearbeitenden oder zu löschenden Teile der Imageseite mittels Rahmenmarkierung am Bildschirm festgelegt.

Arbeitsinhalt	Trennen von Doppelseiten, Ausrichten, Entfernen von Schatten und Strichen
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Erfassungsstelle ausgeführt werden
Qualifikation	technischer Mitarbeiter
Arbeitsaufwand	Automatisiert: 10-20 Minuten Manuell: 60 Minuten

c) Scannen des Originals und Erzeugung einer Filmkopie nach dem COM-Verfahren

Im Normalfall entbehrlich.

3.7 Text-Erfassung

Die Texterfassung kann maschinell (Optical Character Recognition; OCR) oder manuell erfolgen. Für Textteile (Inhaltsverzeichnisse, Kapitelüberschriften, Register), die zur späteren Recherche genutzt werden sollen, ist eine besonders sorgfältige manuelle Doppelerfassung mit interaktiver Verifikation erforderlich. Häufig sind auch Kombinationen von OCR und manueller Eingabe zweckmäßig. Die Erfassung erfolgt üblicherweise auf Grundlage der Imagedatei. Die Kosten für die OCR-Erfassung sind zumeist nur geringfügig niedriger als die manuelle Eingabe in Ländern Südostasiens.

Historische Schrifttypen bilden oftmals ein besonderes Problem, weil maschinelles Lesen wenig Erfolg verspricht und immer weniger Menschen diese Schrift hundertprozentig beherrschen.

Arbeitsinhalt	Eingabe von Inhalten des Buches		
Details	<ul style="list-style-type: none"> • Inhaltsverzeichnis • Kapitelüberschriften • Register • ggf. Fußnoten • ggf. Volltext 		
Ausführende Stelle	Die Arbeiten können von Mitarbeitern der Erfassungsstelle oder auch der Bibliothek ausgeführt werden		
Qualifikation	Datenerfasser (manuelle Eingabe) Technische Mitarbeiter (OCR)		
Arbeitsaufwand	Inhaltsverzeichnis:	60 Minuten	manuell
	Inhaltsverzeichnis:	10 Minuten	OCR incl. Kontrolle
	(Kapitelüberschriften:	60 Minuten)	manuell
	(Kapitelüberschriften:	10 Minuten)	OCR incl Kontrolle
	Register:	120 Minuten	manuell
	Register:	20 Minuten	OCR incl Kontrolle
	Volltext:	100 Stunden	manuell
	Volltext:	20 Stunden	OCR

3.8 Permanente Speicherung

Arbeitsinhalt	Speicherung auf CD-R (Kapazität z.Zt. 650 MByte = ca. 15 Bücher)		
Details	<ul style="list-style-type: none"> • Images • Index-Daten • Inhaltsverzeichnisse, Kapitelüberschriften • Register • Volltext 		
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Erfassungsstelle ausgeführt werden		
Qualifikation	Technische Mitarbeiter		
Arbeitsaufwand	4 Minuten (wenn die CD in einem Arbeitsgang komplett beschrieben wird = ca. 15 Bücher = 1 Stunde)		

3.9 Endkontrolle, Abnahme

Arbeitsinhalt	Abschließende stichprobenweise Überprüfung aller Arbeiten
Details	<ul style="list-style-type: none">• Vollständigkeit• Erfassungsqualität (Filmen, Scannen, Texterfassung)• Richtigkeit der Seitennummerierung und Reihenfolge der Seiten• Strukturierung• Gesamteindruck
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Bibliothek ausgeführt werden
Qualifikation	Studentische Aushilfe unter fachlicher Anleitung durch einen Diplom-Bibliothekar
Arbeitsaufwand	10 Minuten

3.10 Schlußdokumentation

Arbeitsinhalt	Abschließende Dokumentation über die durchgeführten Arbeiten
Details	Strukturierte Eingabe in Datenbank, Katalog etc.
Ausführende Stelle	Die Arbeiten sollten von Mitarbeitern der Bibliothek ausgeführt werden
Qualifikation	Studentische Aushilfe unter fachlicher Anleitung durch einen Diplom-Bibliothekar
Arbeitsaufwand	5 Minuten

4 Zusammenfassung: Durchschnittlicher Aufwand für das Erfassen des Standardbuches

Arbeitsschritt	Bibliothek	Dienstleister	Total
Auswahl der zu erfassenden Bücher	5 Minuten		5 Minuten
Bibliothekarische Vorbereitung	20 Minuten		20 Minuten
Technische Vorbereitung		8 Minuten	8 Minuten
Indexierung		5 Minuten	5 Minuten
Erfassung		30-60 Minuten	30-60 Minuten
Nachbearbeitung		10-60 Minuten	10-60 Minuten
Text-Erfassung Inhaltsverz./Kapitelüberschr.	10-60 Minuten	10-60 Minuten	10-60 Minuten
Text-Erfassung Register	20-120 Minuten	20-120 Minuten	20-120 Minuten
Permanente Speicherung		4 Minuten	4 Minuten
Endkontrolle, Abnahme	10 Minuten		10 Minuten
Schlußdokumentation	5 Minuten		5 Minuten
Total			127-357 Minuten

Zusätzliche Volltexterfassung	20-100 Stunden	20-100 Stunden	20-100 Stunden
-------------------------------	----------------	----------------	----------------

Ohne die Erfassung des gesamten Volltextes erfordert das Verfilmen, Scannen, Indexieren, Texterfassung von Inhaltsverzeichnis/Kapitelüberschriften und Registern und Abspeichern auf CD-R einen Arbeitsaufwand von - je nach gewählter Vorgehensweise - 2 bis 6 Stunden. Die Bandbreite ergibt sich vor allem aus der unterschiedlichen Art der Texterfassung.

Für ein Buch aus dem 19. Jahrhundert, dessen Schriftzeichen sich nicht für OCR-Erkennung eignen, resultiert ein Zeitaufwand von rund 5 bis 6 Stunden. Diese Arbeiten können folgendermaßen zwischen der Bibliothek und dem Dienstleister (der natürlich auch eine Bibliothek mit entsprechender Ausstattung und Erfahrung sein kann) verteilen:

Arbeitsstelle	Zeit (Minuten)
Bibliothek	40
Dienstleister	57 - 137
Bibliothek oder Dienstleister	180

5 Zusammenfassung: Durchschnittliche Fremdkosten für das Erfassen des Standardbuches

Die nachstehenden Kostenangaben beruhen auf Angeboten der letzten Monate.

Arbeitsschritt	Einheit	Preis je Einheit		Total je Band	
		von	bis	von	bis
Technische Vorbereitung, Indexierung	1 Band	0	8,00	0	8,00
Verfilmen	100 Seiten	10,50	20,00	31,50	60,00
Scannen vom Film, einfache Nachbearbeitung	100 Seiten	15,00	20,00	45,00	60,00
Nachbearbeitung	100 Seiten	10,00	10,00	30,00	30,00
Texterfassung Inhaltsverz/Kapitelüberschr.	1000 Zeichen	3,50	18,00	21,00	108,00
Texterfassung Register	1000 Zeichen	3,50	18,00	42,00	216,00
Speicherung: Erste CD-R (4000 Seiten)	1 CD	35,00	130,00	2,70	10,00
Zusätzliche Kopie (4000 Seiten)	1 CD	28,00	80,00	2,20	6,20
Total		105,50	304,00	174,40	498,20
Scannen vom Buch	100 Seiten	30,00	92,00	90,00	276,00
Zusätzliche Volltexterfassung	1000 Zeichen	1,50	9,00	2700,00	10800,00

Nicht enthalten sind: Transport, Versicherung (ca. 2% der Kosten).

Einmalige Kosten: Software zur lokalen Kontrolle (einfache Dokumentenverwaltung, Viewing, Print): Zwischen 1000 und 5000 DM.

Anpassung der Erfassungssoftware: nach Aufwand

Rechnerkonfiguration: PC 133 MHz, 32 MB RAM, 2 GB SCSI-HD, hoch auflösender Bildschirm ca. 20“, Arbeitsplatz-Laserdrucker (z.B. HP Laserjet 5L), zusammen ca. 5000- 9000 DM.

Kosten des Hosts: Umwandlung in Internet-Format
Eintrag in das Datenbank-Managementsystem
Speicherung

In diesem Modell entfallen auf die Bibliothek selbst pro Buch weitere ca. 40 Minuten Arbeitsaufwand, der überwiegend von studentischen Aushilfen unter qualifizierter Anleitung erbracht werden kann.

Interessant ist der folgende Vergleich der Komplettangebote:

Billigster Gesamtanbieter

Arbeitsschritt	Einheit	Preis je Einheit	Total je Band
Technische Vorbereitung, Indexierung (im Verfilmen+Scannen enthalten)	1 Band	0	0
Verfilmen	100 Seiten	20,00	60,00
Scannen vom Film, einfache Nachbearbeitung	100 Seiten	20,00	60,00
Nachbearbeitung (entfällt)	100 Seiten	0	0
Texterfassung Inhaltsverz/Kapitelüberschr.	1000 Zeichen	5,00	30,00
Texterfassung Register	1000 Zeichen	5,00	60,00
Speicherung: Erste CD-R (4000 Seiten)	1 CD	35,00	2,70
Zusätzliche Kopie (4000 Seiten)	1 CD	28,00	2,20
Total		113,00	214,90
Scannen vom Buch	100 Seiten	38,00	114,00

Teuerster Gesamtanbieter

Arbeitsschritt	Einheit	Preis je Einheit	Total je Band
Technische Vorbereitung, Indexierung (im Verfilmen+Scannen enthalten)	1 Band	0	0
Verfilmen	100 Seiten	10,50	31,50
Scannen vom Film, sehr einfache Nachbearbeitung	100 Seiten	15,00	45,00
Nachbearbeitung (entfällt)	100 Seiten	0	0
Texterfassung Inhaltsverz/Kapitelüberschr.	1000 Zeichen	18,00	108,00
Texterfassung Register	1000 Zeichen	18,00	216,00
Speicherung: Erste CD-R (4000 Seiten)	1 CD	130,00	10,00
Zusätzliche Kopie (4000 Seiten)	1 CD	80,00	6,20
Total		271,50	416,70
Scannen vom Buch	100 Seiten	k.A.	k.A.

Dabei ist anzumerken, daß der billigere Anbieter die bessere Qualität (auf Basis seiner Muster) zusichert.

Auffällig sind die großen Abweichungen bei den Kosten einzelner Positionen. Es sollte daher geprüft werden, inwiefern ein Splitten des Auftrages auf mehrere Anbieter möglich ist.