

Das Staatsarchiv Ludwigsburg und Susanne Belovari vom Universitätsarchiv der University of Illinois, entwickelten in einem dreiwöchigen Kooperationsprojekt einen ersten einfachen Zugang, um digitale Sammlungen rasch und mit wenig und einfacher Software bewerten und bearbeiten zu können.

08 - Rasche und einfache Bearbeitung digitaler Sammlungen

Software Tests

Da Bearbeitungsschritte auch mittels Software erfolgen, müssen wir oder unsere Fachorganisationen anfangen, Softwarekapazitäten zu testen und zu bewerten. Nur dann kann sichergestellt werden, dass Softwareprogramme die Bearbeitungsschritte zuverlässig und dokumentierbar durchführen.

Tests - Entfernung von Doppelungen sowie leerer Dateien/Ordner

Effizient können doppelte Dateien und leere Dateien/Ordner nur mit Software entfernt werden. Wir testeten daher zehn Programme zur Deduplizierung mehrfach anhand einer Stichprobe, die 10 % der Dateien (+ lange Dateinamen), 20% der Ordner, 1/3 der Duplikate und eine Vielfalt an Formaten der Projektsammlung beinhaltete. Acht der zehn Softwareprodukte versagten beim Identifizieren des Inhaltes (Dateien, Ordner, Größe) und der Anzahl doppelter Dateien sowie bei der korrekten Deduplizierung. Außerdem waren sie bezüglich Softwareeinstellungen, graphischen Darstellungen und Berichts-/ Dokumentationsfunktionen unzulänglich. Bei zusätzlichen Tests der zwei geeigneten Programme, Duplicate File Detective (v.6.0.76) und Tree Size Pro (TSP, freeware), konnte nur TSP alle Duplikate regelmäßig identifizieren und entfernen und bestach durch einfache Handhabung, Einstellungsmöglichkeiten, Geschwindigkeit (<1 Min.), großartige Graphikeinstellungen und Berichtsfunktionen, welche die Verifizierung und Dokumentation ermöglichen.

In weiteren Tests von Remove Empty Directories und TSP, überzeugte TSP auch hier durch die bessere und korrekte Identifizierung leerer Dateien/Ordner und durch seine Berichtsfunktionen

Bearbeitungsschritte – ein Beispiel

1. Rasche Durchsicht: erste Bewertungsfragen + Kriterien, potentielle Risiken

Nachdem wir uns über die abgebende Stelle (oder Person) und Übergabe informiert hatten, sichteten wir die Sammlung visuell und dann mit Hilfe von TSP.

Die Dateisammlung (1960-2015) der Johannes-Wagner-Schule Nürtingen für Hörgeschädigte und Sprachbehinderte beinhaltete 677 GB, 65057 Dateien, 3638 Ordner und eine komplexe (10+) Ordnerhierarchie. Formate waren hauptsächlich Fotos und AV, digitalisierte Schmalfilme und Zeitungsartikel sowie Office-Dateien. Es gab eine inadäquate Inventarliste und eine visuell geschätzte Doppelung von ca. 70% der Ordner und Dateien.

2. Grobe Bewertung

Wenn gewollt, können mit guter Software und Segmentanalyse bestimmte Datei- und Ordnerarten gelöscht, riesige Sammlungen reduziert und so die weitere qualitative Bewertung vereinfacht werden.

a. Entfernung von Doppelungen

- Manuell (60 Min.): Löschung der visuell ersichtlichen, doppelten Hauptordner (minus 2/3 Speicherplatz).
- Software (<1 Min.): mit TSP dokumentierte Löschung der doppelten Dateien (minus 8% Dateien)

b. Entfernung leerer Ordner und leerer, temporärer + rein technischer Dateien

- Software (wenige Sek.): mit TSP Filter- und graphische Ansichtsoptionen rasche Identifikation + dokumentierte Löschung leerer Ordner und leerer, temporärer oder rein technischer Dateien.

3. Qualitative Bewertung

a. Entscheidung bezüglich Bewerten / Nichtbewerten

Unsere Projektsammlung hatte nun zirka 30% der ursprünglichen Größe und Dateienanzahl sowie 40% der ursprünglichen Ordner. Jetzt kann man entscheiden, ob sich eine qualitative Bewertung mit Hinblick auf z.B. Verdichtung der Akten, Speicherplatz, Arbeitszeit, Erhaltungsformat, Risiken, Kosten, Software usw. auszahlt.

- Software (wenige Sek.): mit TSP, besonders der graphischen Ansichten, konnten wir nach Format, Alter, Größe, Ordner usw. filtern, Resultate verifizieren, und Material oder Ordner leicht vorreihen, bei denen sich nach selbst gewählten Kriterien die Bewertung auszahlt.

Zum Beispiel entschieden wir uns, Video- und Bilddateien zu bewerten (7,3% an Videodateien belegten $\frac{3}{4}$ des Speicherplatzes); Office Dateien aber wegen der geringen Speicher- und Migrationsprobleme ad acta zu legen. Wir übernahmen ältere Video- und Fotoordner bis 2004 (< 2% der Dateien) komplett, da jeweils wenige Dateien das historische Geschehen dokumentierten. Jüngere Ordner von 2009-2014 (> 85% der Dateien) unterzogen wir einer qualitativen Bewertung.

b. Qualitative Bewertungskriterien (die auch während der Bewertung entwickelt werden)

Besonderheiten einer Sammlung, eines Archivs, der ArchivarInnen, das Hintergrundwissen und andere kontext-bezogene Faktoren bedingen immer (auch) die Auswahl qualitativer Bewertungskriterien.

Die folgenden inhaltlichen und visuell-spezifische Kriterien waren für uns ausschlaggebend.

Inhaltliche Kriterien (ein Beispiel)

- geschichtliche Dokumentation (auch langfristige Konsequenzen des Nationalsozialismus)
- Gruppen: z.B. Studenten, Eltern, Lehrer, Köche
- Infra/strukturen: z.B. Internat, Labors, Klassenzimmer, Garten
- Schulprogramm + dessen Vielfältigkeit: z.B. Gebärdensprache, handwerkliches Lernen, offsite Schaubesuche
- Teilabschnitte einer Aktivität: z.B. E-bike + Bienenhotel bauen
- Zeitgenössische Themen: z.B. Theater gegen Mobbing, Kochen + Diät
- Schuljahr + Höhepunkte
- Vielfalt innerhalb einer Veranstaltung: z.B. Spiele auf Landschulwoche, Weihnachtsmarkthandwerk
- Aufführung ≠ Proben: z.B. Theateraufführung, Jahresabschlussfeier
- bestechende Gendergleichheit bei Aktivitäten
- Neuheiten: z.B. Hörgerätemoden, Interessen / Stärken von Hörgeschädigten, frühe Integrationsbeispiele (z.B. Schifahren)
- pädagogisches Training und Lehrhilfsmittel

Visuell-spezifische Kriterien (ein Beispiel)

- (1) Art der Abbildung: z.B. Profil/Hinterkopffoto statt Frontalfoto, um getragene Hörgeräte zu dokumentieren
- (2) Übersicht: z.B. zusammenfassendes Video mit Text ≠ kurze Clips von Langläufern
- (3) Qualität der Aufnahme/Format, auch bezüglich zukünftiger Reproduzierbarkeit: z.B. Teamfoto ≠ Fußballvideoclip

c. Workflow

Zur Bearbeitung der AV Dateien war es am effizientesten, TSP und IconView/Preview zu verwenden. Mit TSP Segmentanalyse konnten wir z.B. rasch nach Größe oder Format priorisieren; TSP war auch sehr stark im Öffnen/-Abspielen der Video- und Audiodateien. Hingegen erlaubte die Icon View/Previewfunktion (eines jeden Computers) die rasche Durch- und Übersicht sowie Bewertung der Fotos innerhalb eines Ordners.

4. Resultat

Die Verdichtung der AV Dateien und die kleine Sammlung an Office-Dateien, schafften nun geringe Aufbewahrungsaufgaben für das Archiv und geringe Suchaufgaben an zukünftige Benutzer, die Office-Dateien ja mit Hilfe der normalen Suchfunktionen leicht durchforsten können. In vier Tagen reduzierten wir die Sammlung auf 74.2 GB (-89%), 1314 Ordner (-64%), und 8467 Dateien (-87%).

Susanne Belovari
Assistant Professor - Archivist for Faculty Papers
University of Illinois
belovari@illinois.edu

Corinna Knobloch
Staatsarchiv Ludwigsburg
corinna.knobloch@la-bw.de

Weitere Kurzartikel aus der Reihe „nestor Thema“ finden Sie auf www.langzeitarchivierung.de - der Webseite von **nestor – Kompetenznetzwerk Langzeitarchivierung**.