

Evaluierung

CARL RAUCH / ANDREAS RAUBER

Anwendung der Nutzwertanalyse zur Bewertung von Strategien zur langfristigen Erhaltung digitaler Objekte

Foto privat



Carl Rauch

Foto privat



Andreas Rauber

Durch die steigende Zahl (nur mehr) elektronisch vorhandener Daten, durch die wachsende Vielfalt an Datenformaten, deren zunehmende Komplexität und Funktionsvielfalt sowie die extreme Kurzlebigkeit dieser Formate wird die Herausforderung der langfristigen elektronischen Datenspeicherung immer dringlicher. Obwohl schon mehrere Ansätze, wie Migration, Emulation oder Computer-Museen vorgeschlagen wurden, ist doch keine Lösung unter allen Umständen besser als die anderen, die Wahl der optimalen Speicherlösung wird dadurch zu einer sehr komplexen Fragestellung. In diesem Artikel werden im ersten Teil die bekanntesten Strategien zur langfristigen Datenspeicherung vorgestellt. Im zweiten Teil wird auf eine adaptierte Version der Nutzwertanalyse näher eingegangen, die bewusste und dokumentierte Entscheidungen für oder gegen bestimmte Strategien erlaubt.

Due to the increasing number of (solely) electronically available data, the growing diversity of data formats and their increasing complexity and functional diversity, as well as the extreme short life-span of these formats, there is a growing urgency to meet the challenge of long-term preservation of electronic data. The first part of this article presents the most familiar strategies for long-term storage. A modified version of cost-benefit analysis which allows conscious and documentable decisions for or against certain strategies is discussed in detail in the second section.

EINFÜHRUNG

In den letzten Jahren haben Bibliotheken, Museen und die öffentliche Verwaltung immer stärker das Problem der langfristigen digitalen Datenspeicherung erkannt. Erst wenn digitale Objekte auch nach zwanzig Jahren verlässlich wieder geöffnet werden können, ohne an Aussagekraft zu verlieren, wird es möglich sein, auf Papierarchive zu verzichten und eine Organisation auf die funktional mächtigere digitale Datenspeicherung umzustellen.

Die Speicherung des rein binären Codes stellt keine große Herausforderung mehr dar – CDs mit einer angeblichen Haltbarkeit von 100 Jahren sind auf dem Markt,¹ Archivsysteme, die ihre Daten automatisch von einem Medium auf ein anderes überspielen, werden verkauft.² Ein zweites Problem spielt eine viel wichtigere Rolle: das Wieder-Öffnen dieser Daten, das richtige Interpretieren des Bit-Stromes ohne wesentliche Eigenschaften der Daten zu verlieren. Wie kann verhindert werden, dass neue Software alte Daten nicht anders interpretiert und Überschriften versetzt, das Seitenlayout ändert, eingebettete Makros anders interpretiert, Teile des ursprünglichen Files nicht mehr anzeigt, wie Animationen oder interaktive Programmteile, oder überhaupt nur noch eine lange wirre Zeichenkette wiedergibt?

Um nicht große Datenmengen in den nächsten Jahren zu verlieren, was in vielen Bereichen, wie Versi-

cherungen oder Bibliothekssammlungen eine große Rolle spielt, wurden eine Reihe von Bewahrungsstrategien entwickelt, die sich in die drei Gruppen Migration, Emulation und Computer-Museen gliedern. Diese Gruppen lassen sich in eine große Zahl von Untergruppen und Implementierungen unterteilen und zu einer noch größeren Zahl von Lösungen kombinieren. Abhängig von den zu bewahrenden Eigenschaften eignet sich die eine oder andere Strategie besser, um ein digitales Objekt zu bewahren. Um eine möglichst optimale Speicherstrategie zu finden, wird in diesem Artikel ein Werkzeug vorgestellt, mit dem zuerst alle entscheidungsrelevanten Kriterien erhoben, dann gemessen und schließlich so aggregiert werden können, dass sich daraus eine Empfehlung für eine Speicherstrategie ableiten lässt.

In diesem Beitrag werden im ersten Teil verschiedene Methoden zur langfristigen Datenspeicherung beschrieben, diese näher erläutert und ihre Vor- und Nachteile gezeigt. Im zweiten Teil wird die Nutzwertanalyse zur Entscheidungsunterstützung vorgestellt.

LANGFRISTIGE SPEICHERSTRATEGIEN

In den letzten Jahren wurden eine Reihe von Strategien entwickelt, um Daten langfristig zu bewahren.³ Diese lassen sich in kurz- und mittel- bis langfristige Ansätze teilen. In dieser Ausarbeitung werden Computer-Museen als Beispiel eines kurz- bis mittelfristigen Ansatzes und Migration und Emulation als wichtigste Vertreter der langfristigen Datenarchivierung vorgestellt.

Neben diesen großen Vertretern werden laufend alternative Speicherstrategien vorgeschlagen. Ein Beispiel wäre ein völlig unabhängiger Computer, aus dem man Daten nur mit Hilfe einer Leseschnittstelle auslesen kann.⁴ Diese Lösung scheitert vor allem an der Stabilität der Hardware, die nicht über lange Zeiträume hinweg sicher gestellt werden kann. Ein anderes Beispiel wäre das Ausdrucken der Daten, sei es auf Papier oder stark verkleinert auf Nickelplatten,⁵ wobei hier nur das Aussehen eines Objekts erhalten bleibt und alle weiteren Vorteile eines digitalen Dokuments, wie schnelles Suchen oder die hohe Mobilität, verloren gehen. Im Folgenden sollen kurz die wichtigsten Strategien vorgestellt werden.

Wieder-Öffnen von
Altdateien

Migration

Die am häufigsten verwendete Strategie ist wohl Migration. Sie wird von jedem Computernutzer oft unbemerkt angewandt, wenn die neue Version einer Software installiert wird und dann nach und nach alle Daten nur mehr mit dieser neueren Version einer Software geöffnet und gespeichert werden. Dieses Ändern und Anpassen der Daten an die aktuellen Formate und Softwareumgebungen, das heißt die Konvertierung auf ein anderes, neueres Format, wird Migration genannt. Vorteile dieses Ansatzes sind:

- **Aktuelles Format:** Die digitalen Daten liegen immer in einem aktuellen Format vor und können in jeder Standardumgebung gelesen werden.
- **Homogene Datenmengen:** Durch die oben genannte Standardisierung und ständige Migration der Daten ist es möglich, diese in einem einheitlichen Format zu speichern. Das österreichische Phonogrammarchiv zum Beispiel erhält Daten in rund 20 verschiedenen Eingangsformaten, migriert aber alle Daten sofort auf das Einheitsformat DPS.
- **Standardisierung:** Migration muss nicht nur innerhalb eines Datenformats, sondern kann auch zwischen mehreren Datenformaten erfolgen. So können Daten von seltenen Formaten oder von solchen, die sich schlecht für die langfristige Datenspeicherung eignen, auf passendere migriert werden. Ein Beispiel wäre die Konvertierung von Word Perfect auf PDF.

Nachteile dieses Ansatzes sind:

- **Regelmäßiger Arbeitsaufwand:** Zurzeit ändert sich ungefähr alle fünf Jahre⁶ ein Speicherformat so sehr, dass ein Migrationsschritt notwendig wird. Dabei müssen alle Daten einer Sammlung migriert werden, was bei größeren Mengen nur noch durch eigene Migrationssoftware möglich ist.
- **Verlust von Informationen:** Bei jeder Migrationswelle besteht die Gefahr, dass Teile der Information eines digitalen Objekts verloren gehen. Oft merkt man es gar nicht sofort, wenn die Auflösung eines Bildes geringer wird oder die Überschrift verrutscht, aber nach ein paar Jahren entsprechen die Daten nicht mehr ihrem Original. Ein Beispiel hierfür wäre die Migration von MS Word zu PDF, bei der zwar das Aussehen eines Dokuments erhalten bleibt, nicht aber die Funktionalität, Text ändern, Makros, integrierte Berechnungen oder Animationen nutzen zu können.⁷
- **Wahl der optimalen Alternative:** Selbst wenn Migration als Strategie gewählt ist und ein Zielformat für eine Sammlung gefunden wird, gibt es meist noch immer relativ viele verschiedene Softwaretools, um Daten vom Ursprungs- ins Zielformat zu bringen. Hier stellt sich zum ersten Mal das Problem der Wahl einer geeigneten Transformation.

Emulation

Auf einer völlig anderen Ebene setzt Emulation an. Hier ist das Ziel, das zu speichernde digitale Objekt unverändert zu lassen und nur die Umgebung, in der die Datei geöffnet werden soll, anzupassen. Dafür werden eigene Softwareprogramme geschrieben, so genannte Emulatoren, die in einer neuen Computerumgebung die Funktionalität und Umgebung des ursprünglichen Systems simulieren. In diesem simulierten System kann das originale File wieder geöffnet werden. Emulatoren kann es auf mehreren Ebenen geben: Entweder auf der Ebene der Hardware, auf der Ebene des Betriebssystems, wobei die originale Software innerhalb der emulierten Umgebung geöffnet wird, oder es kann die öffnende Software selbst der Emulator sein und nur das digitale Objekt selbst im Original erhalten bleiben.

Emulation ist in der Informatik schon lange bekannt und wird unter anderem bei neuen Computerplattformen eingesetzt, um auf neuer Hardware Teile eines älteren Betriebssystems verwenden zu können.⁸ Ein zweites Einsatzgebiet sind Computerspiele, die mit Hilfe von Emulatoren wieder zum Leben erweckt werden.⁹

Als Vorbereitung für Emulation ist es sinnvoll, die zu speichernden Dateien in einer Art Kapsel zu verpacken, in der alle Informationen gespeichert werden, die es in der Zukunft ermöglichen werden, dieses File wieder richtig zu interpretieren. Dazu gehören unter anderem die originale Software, das originale Betriebssystem, eine Hardwarespezifikation, eine Dokumentation, in der das originale Erscheinungsbild beschrieben wird, Metadaten, die zum Wiederaufmachen der Datei notwendig sind, und schließlich eine allgemeine Beschreibung der Datei und weiterer wichtiger Merkmale.¹⁰ Eine Spezialform von Emulation ist der Universal Virtual Computer (UVC), der zurzeit von IBM und von der Königlichen Bibliothek (Koninklijke Bibliotheek) in Den Haag entwickelt wird. Dieser relativ neue Ansatz hat das Ziel, eine geräteunabhängige Plattform zu bauen, die auf einer kleinen Anzahl von Basisbefehlen beruht und leicht in jeder neuen Computergeneration verwendet werden kann. Zusätzlich werden für bestehende Computerumgebungen Emulatoren geschrieben, um die zu speichernden Daten auf dem UVC abzubilden. Um digitale Objekte auf einer zukünftigen Plattform zu öffnen, muss nur noch ein einziger Emulator geschrieben werden, der die Darstellung innerhalb des UVC auf diesem neuen System ermöglicht. Für diese Strategie wird gerade ein Prototyp entwickelt, der innerhalb der nächsten Jahre evaluiert werden wird.¹¹ Der UVC scheint für einige spezielle Anforderungen, wie zum Beispiel Software

homogene Datenmengen

Universal Virtual Computer (UVC)

zur Steuerung von Auto-Motoren, die einzige sinnvolle Speichermöglichkeit zu sein.

Die Vorteile von Emulation sind:

- Unverändertes Original: Das ursprüngliche digitale Objekt bleibt unverändert und wird nur von anderen Emulatoren interpretiert.
- Niedrige Anfangskosten: Zu Beginn entstehen relativ geringe Kosten, es müssen die zu speichernden Daten nur in den Kapseln verpackt werden.

Die Nachteile sind:

- Technisch relativ aufwändig: Das Erstellen eines Emulators und vor allem der Emulator-Schnittstellen ist relativ aufwändig.
- Steigender Aufwand: Mit einer steigenden Anzahl an unterschiedlichen Dateiformaten in einer Sammlung steigt der Aufwand für das Erzeugen einzelner Emulatoren stetig an, eine mögliche Lösung wären Hierarchien von Emulatoren.

Computer-Museen

Die letzte hier vorgestellte Alternative zur langfristigen Datenspeicherung sind Computer-Museen. Dieser Ansatz wird vorgestellt, weil er eine oft verwendete und logische Lösung darstellt, er ist aber langfristig auf keinen Fall zu empfehlen. Computer-Museen sind Sammlungen von nicht mehr aktuellen Computern und Umgebungen, mit denen ältere digitale Objekte in ihrer originalen Umgebung wieder aufgemacht werden können. Für Zeiträume von fünf Jahren ist diese Lösung einsetzbar, bei längeren Zeiträumen werden die Kosten für die Wartung der Geräte ständig höher, bis überhaupt keine Ersatzteile mehr erhältlich sind.

WAHL DER AM BESTEN GEEIGNETEN STRATEGIE

Übersicht, Notwendigkeit der Metrik

Neben diesen drei allgemein anerkannten Strategien existieren noch eine ganze Reihe weniger verbreiteter Ansätze, und auch diese drei lassen sich noch in eine große Zahl von möglichen Alternativen teilen. ›Migriere ich meine Daten auf PDF, XML oder Word?‹, ›Verwende ich dazu den PDF Distiller, den ADOBE Printer oder PDF Factory?‹ sind nur einige von vielen Fragen, die man sich vor der Implementierung einer langfristigen Datensicherung stellen könnte. Keine dieser vielen

Alternativen ist in allen Szenarien besser als eine andere, die Wahl der optimalen Speicherstrategie hängt sehr stark von den vom Anwender definierten Anforderungen ab.

Das Ziel der Arbeit am Institut für Softwaretechnik und Interaktive Systeme ist, alle möglichen Anforderungen in einem Rahmenwerk zu integrieren, zu aggregieren, damit diese verschiedenen Alternativen zu evaluieren und schließlich eine Kenngröße zu schaffen, die die klare Wahl für eine Lösung auf einer dokumentierbaren und nachvollziehbaren Basis erlaubt. In der Infrastrukturökonomie, wo solche komplexen Fragestellungen zum Beispiel für Autobahntrassen oder bei der Planung von Stadtvierteln schon länger eine Rolle spielt, wurde in den siebziger Jahren die Nutzwertanalyse entwickelt.¹² In diesem Kapitel wird eine adaptierte Version der Nutzwertanalyse vorgestellt, die sich für die Anforderungen der langfristigen Datenspeicherung eignet.¹³

Ablauf der Metrik

In Abbildung 1 sind die acht Schritte der adaptierten Nutzwertanalyse zu sehen. Diese acht Schritte lassen sich entsprechend ihrer thematischen Zugehörigkeit in drei Teilbereiche gliedern: Während der ersten beiden Schritte werden die Ziele, die eine langfristige Speicherstrategie erfüllen soll, bestimmt. Bei den Schritten drei, vier und fünf werden die verschiedenen Alternativen definiert und evaluiert. Die Aggregation und Entscheidungsfindung erfolgt während der letzten drei Schritte. In der Folge werden die einzelnen Schritte zuerst theoretisch vorgestellt und dann anhand eines simplifizierten Beispiels verdeutlicht.

Bestimmung der Ziele

Dieser erste Block besteht aus zwei Schritten: Der Definition des Zielbaums und der Zuordnung von Maßeinheiten zu den Blättern dieses Baums. Bei der Implementierung der angepassten Nutzwertanalyse hat sich der erste Schritt als der aufwändigste und wichtigste für den Erfolg eines Langzeitarchivierungsprojekts herausgestellt. Dabei werden alle Kriterien, die bei der Entscheidung für die eine oder andere langfristige Methode berücksichtigt werden sollen, aufgelistet und in einer Baumstruktur geordnet. In der Praxis

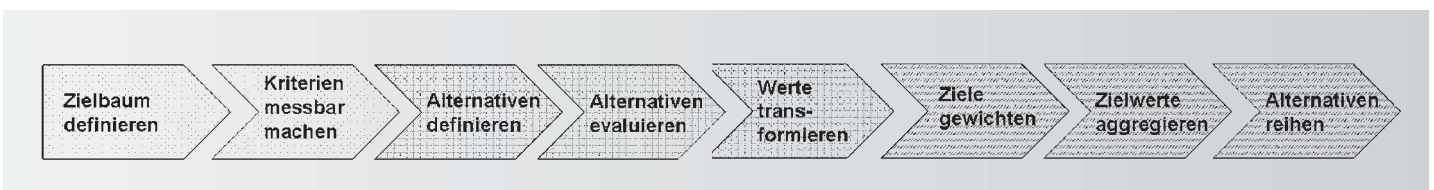


Abb. 1: Die acht Schritte der adaptierten Nutzwertanalyse

werden dafür zwei Methoden verwendet: Der Bottom up- und der Top down-Ansatz. Beim erstgenannten werden bei einer Besprechung mit Anwendern und Technikern alle Kriterien aufgelistet, die den Teilnehmern ad hoc einfallen, und diese in weiterer Folge geordnet und strukturiert.

Beim Top down-Ansatz hingegen wird das allgemeine Ziel, die digitalen Objekte möglichst langfristig und optimal zu speichern, immer weiter untergliedert. Dafür wurde ein Referenzbaum erstellt, der die Kriterien der ersten beiden Ebenen beinhaltet. Die detaillierten Punkte werden den Anforderungen der Anwender entsprechend ergänzt. Dieser Referenzbaum, der in Tabelle 1 zu sehen ist, teilt die Kriterien in der ersten Ebene in Charakteristika des zu speichernden digitalen Objekts, in die technischen Attribute des darunter liegenden Files, in Charakteristika des Speicher- bzw. Archivierungsprozesses und in die dabei entstehenden Kosten.

Die Eigenschaften des digitalen Objekts beschreiben alle Aspekte, die vom Benutzer direkt wahrgenommen werden und die dieser benötigt, um das Objekt in seinem Zusammenhang zu verstehen. Daher ist die zweite Ebene in die von Jeff Rothenberg vorgeschlagenen Charakteristika ›Aussehen‹, ›Inhalt‹, ›Struktur‹, ›Kontext‹ und ›Verhalten‹ unterteilt.¹⁴

Bei den technischen Eigenschaften unterscheiden wir zwischen jenen, die das Speicherformat, jenen, die das Speichermedium und jenen, die die Visualisierungsumgebung beschreiben. Diese Kriterien werden vom Anwender nicht unmittelbar wahrgenommen, wenn er mit dem digitalen Objekt zu tun hat, trotzdem sind sie notwendig, um das Objekt wieder so herzustellen, wie es im Original gewirkt hat, bzw. um eine integrierte Funktionalität zu erhalten.

Die Prozess-Charakteristika werden anhand der Gliederung des Referenzmodells für ein ›Open Archival Information System‹ gegliedert, die den Archivierungsprozess in den Import des Objekts in eine Speicherumgebung, in die Bewahrung innerhalb dieser Umgebung und in das Wiederverfügbarmachen gliedert.¹⁵ Daneben werden alle Managementprozesse in einem Punkt zusammengefasst.

Die Kosten werden in dem generischen Entscheidungsbaum schließlich in Personal- und Sachkosten geteilt. Der hier vorgestellte Baum soll als Unterstützung für die Identifikation von Kriterien dienen und kann an einzelnen Stellen erweitert, an anderen reduziert werden, falls das eher den Vorstellungen einer Organisation entspricht.

Um weitere Kriterien zu identifizieren ist eine ausführliche Recherche von relevanter Literatur für das jeweilige Anwendungsgebiet und eine detaillierte Ana-

lyse der zu bewahrenden Dateien und Anwendungsanforderungen erforderlich.

Bei verschiedenen Implementierungen zusammen mit dem österreichischen Phonogrammarchiv und dem Nationalarchiv (Nationaal Archief) der Niederlande wurden rund 320, 350 und 500 Kriterien in sechs verschiedenen Hierarchien identifiziert, die bei der Entscheidung für eine optimale Strategie berücksichtigt werden sollen. Zurzeit wird daran gearbeitet, einen detaillierten generischen Baum zu entwickeln, bei dem, je nach Anwendungsgebiet, bestimmte Unterpunkte übernommen werden können.

Als Beispiel wird ein sehr simpler Entscheidungsbaum verwendet, der nur dazu dient, den Ablauf der adaptierten Nutzwertanalyse besser verständlich zu machen. Folgende Annahmen werden getroffen: Die zu bewahrenden digitalen Objekte liegen im Word95-Format vor, es soll die langfristige Haltbarkeit des Formats evaluiert werden. Alle anderen Eigenschaften der Umgebung, wie das Speichermedium oder die Metadatenbank, sollen dabei nicht verändert werden. Eine zweite Annahme ist, dass der Anwender nur die vier im Beispielbaum gezeigten Kriterien bei der Wahl des nächsten Speicherformats berücksichtigt. Der entsprechende Entscheidungsbaum ist in Tabelle 2 dargestellt. Prozesseigenschaften werden nicht angeführt, wodurch dieser Ast des Entscheidungsbaums verschwindet. In realen Anwendungen liegt die Anzahl

Erstellung des Referenzbaums

Tabelle 1:
Der generische Entscheidungsbaum

| <u>Ebene 1</u> | <u>Ebene 2</u> | <u>Beispiel für Ebene 3</u> |
|--------------------------|-------------------------------|--|
| Objekt-Eigenschaften | Aussehen | Farben, Schriftarten, ... |
| | Struktur | Logos, Seitenaufbau, ... |
| | Inhalt | Text, Buchstaben, ... |
| | Kontext | Erzeugung, Prozesse, ... |
| | Verhalten | Makros, Interaktion, ... |
| Technische Eigenschaften | File Format | Haltbarkeit, Besitzer, ... |
| | Speichermedium | Garantie, Lesbarkeit, ... |
| | Visualisierung | Bildschirm, Lautsprecher, ... |
| Prozess-Eigenschaften | Import | Kontext, Sicherungskopien |
| | Bewahrung | Überprüfungen, Speicherumgebung, ... |
| | Wiederaufmachen Verwaltung | Arbeitsaufwand, Zugriffsrechte, ... Wahl der Speicherstrategie, ... |
| Kosten | Personal | Pro File, pro Stunde, ... |
| | Sachkosten | Einkauf, Adaption |

der Kriterien, also der Blätter des Baums, üblicherweise im Bereich von 100 bis 500.

| Top level | Level 2 |
|--------------------------|------------------------------------|
| Objekt- Eigenschaften | Farbqualität Entstehungskontext |
| Technische Eigenschaften | Offene Spezifikation |
| Kosten | Importkosten pro File |

Tabelle 2: Der für das Beispiel verwendete Entscheidungsbaum

Kriterien messbar machen

Beim Aufbau des Entscheidungsbaums spielt die Art der Messung der Kriterien noch keine Rolle. Erst in diesem zweiten Schritt werden Maßeinheiten definiert, um die Kriterien auch quantitativ bewerten zu können. Allen messbaren Kriterien werden Maßeinheiten zugeordnet, also Euro zu den Kostenkriterien, Millimeter oder Bildauflösung zu anderen. Daneben gibt es zahlreiche Kriterien, denen keine Bewertungsgröße direkt zugeordnet werden kann, wie zum Beispiel Komplexität des Prozesses oder der Unterstützungsgrad der automatischen Metadatenextraktion. Diese Fälle sollen so weit wie möglich in weitere messbare Kriterien untergliedert werden. Wenn so eine Messung zu aufwändig wäre oder es keine sinnvolle Maßeinheit gibt, besteht die Möglichkeit, die Ausprägung subjektiv zu bewerten. In diesem Fall wird nur eine Skala definiert, auf der bei der Bewertung eine Note, entsprechend z. B. dem Schulnotensystem, verwendet wird. In den praktischen Implementierungen wird üblicherweise eine Skala von 0 bis 5 verwendet. Fünf steht für sehr gut, null zeigt an, dass eine Alternative bei diesem Kriterium nicht akzeptierbar schlecht abschneidet. Wird bei einer Strategie nur einmal ein Aspekt so schlecht bewertet, wird die Strategie in der endgültigen Reihung nicht mehr berücksichtigt.

Bei dem als Beispiel verwendeten Entscheidungsbaum werden den vier behandelten Kriterien in diesem Schritt messbare Einheiten zugeordnet. Dabei werden den Importkosten pro File Euro zugeteilt. Die Frage, ob eine Spezifikation veröffentlicht ist oder nicht, wird mit ja oder nein beantwortet. Für das Kriterium Farbqualität eignet sich keine messbare Einheit, darum wird eine Bewertung durch den Anwender verwendet. Alternativ dazu könnte man versuchen, das Kriterium in weitere Einheiten zu gliedern, die dann eventuell messbar sind. Das gleiche gilt für den Ent-

stehungskontext, der auch vom Anwender bewertet werden soll.

Alternativen definieren

Die Festlegung von alternativen Methoden zur langfristigen Bewahrung, die für das vorliegende Projekt in Frage kommen, ist der erste Schritt im zweiten Teil der adaptierten Nutzwertanalyse.

Wie in Abschnitt zwei gezeigt, gibt es mehrere langfristige Strategien, Emulation oder Migration sind die beiden wichtigsten Gruppen. Von diesen grundsätzlichen Strategien abgeleitet, gibt es eine ganze Reihe von technischen Implementierungen, oft auch in anderen Anwendungsbereichen als der langfristigen Datenspeicherung. Beispiele wären Migrationstools von einem Format in ein anderes oder Emulatoren, wie sie für Computerspiele verwendet werden. In diesem Schritt werden alle möglichen Alternativen aufgelistet, ein Prozess definiert, wie sie umgesetzt und die Durchführung von Experimenten mit diesen Speicherlösungen vorbereitet werden könnten.

Um das Beispiel nicht zu kompliziert werden zu lassen, werden drei Alternativen definiert: Migration der Word 95-Daten in Word 2002, in PDF 1.4 mit Hilfe des Acrobat Distiller-Konverters 3.0 oder die Aufbewahrung der digitalen Objekte in einem Computer-Museum. In diesem Fall müsste die derzeit verwendete Hardwareumgebung konserviert werden.

Alternativen evaluieren

Die im vorigen Schritt definierten alternativen Strategien werden in diesem Schritt angewandt und evaluiert. Als Basis für diese Bewertung dienen Daten, die entweder aus der zu speichernden Sammlung selbst stammen oder die typische Eigenschaften dieser Daten aufweisen. Sie können entweder speziell für diese Evaluierungen erzeugt oder von Referenzdatenbanken bezogen werden. Ein Beispiel einer solchen Datensammlung, in der Daten mit vielen verschiedenen Ausprägungen in unterschiedlichen Dateiformaten gespeichert werden, ist das Berkeley Digital Library Project.¹⁶ Für Kriterien, die nur subjektiv bewertet werden, ist es sinnvoll, eine kurze Begründung zu notieren, um in späteren Evaluierungen Referenzdaten zu haben und ähnlich entscheiden zu können.

Mit Hilfe dieser Daten, die bei allen Testfällen die gleichen sein sollen, werden die Prozesse der einzelnen Alternativen durchgeführt. Für die Evaluierung werden die im Entscheidungsbaum definierten Kriterien in der im zweiten Schritt bestimmten Maßeinheit bewertet. Das Endergebnis ist eine Tabelle, bestehend aus den Blättern des Entscheidungsbaums und den Al-

Schulnoten für Kriterien ohne Bewertungsgröße

ternativen, die Felder sind jeweils mit den Bewertungen gefüllt.

Die Bewertung der Alternativen ergibt folgendes Bild: Die Farbqualität ist bei allen drei Alternativen gut erhalten geblieben, leichte Änderungen wurden nur bei PDF festgestellt. Darum wird PDF 1.4 mit 4 Punkten bewertet, die anderen beiden Alternativen mit 5 Punkten.

Der Kontext der Entstehung ist in der Metadatenbank gespeichert und ist daher von der Migration der Datei nicht betroffen. Daher bleibt sie unverändert erhalten und wird für alle Alternativen mit 5 bewertet.

Bei den technischen Eigenschaften wird evaluiert, ob die Datei in einem offenen oder proprietären Dateiformat gespeichert wird. PDF 1.4 ist ein offenes Format, die anderen beiden Alternativen mit Word 95 und Word 2002 basieren auf einem proprietären Format und werden daher mit ›nein‹ bewertet.

Das letzte Kriterium sind schließlich die Importkosten pro File. Um diesen Wert korrekt zu erheben, müssen die Migrationen mehrmals durchgeführt und ein Mittelwert berechnet werden. Dieser Mittelwert wird dann mit dem Stundenlohn eines typischerweise mit solchen Aufgaben beschäftigten Mitarbeiters multipliziert. Bei den Computer-Museen ist hier kein Aufwand notwendig, da die bestehende Lösung übernommen wird. Laut Kalkulationen des niederländischen Nationalarchivs¹⁷ dauert die Migration eines Texts unabhängig vom Zielformat rund eine halbe Minute. Das dafür notwendige technische Verständnis ist relativ gering, als Stundenlohn werden von uns 30 Euro angenommen. Die Kosten für den Import eines Files betragen damit rund 0,50 Euro, sowohl für die Migration auf MS Word 2002 als auch auf PDF 1.4. Diese Zahlen beruhen auf der Annahme, dass der Migrationsprozess, wie in vielen Archiven, manuell und nicht automatisiert durchgeführt wird.

Werte transformieren

Nachdem die Evaluierung durchgeführt worden ist und alle Bewertungen in den unterschiedlichen Maßeinheiten vorliegen, ist der nächste Schritt, diese Werte zu vereinheitlichen. Dafür werden alle Werte in vergleichbare Zahlen transformiert. Deren Skala sollte jener der subjektiv bewerteten Kriterien entsprechen, also null bis fünf in den bis jetzt durchgeführten Testfällen. In diesem Fall ist auch die Transformation der subjektiv bewerteten Kriterien leicht, sie können einfach übernommen werden.

Die Transformation der anderen Werte ist schwieriger, wobei hier grundsätzlich zwei Möglichkeiten an-

gewandt werden können: Bei der ersten wird die beste Bewertung eines Kriteriums unter allen Alternativen heraus gesucht und diesem Wert die höchste Bewertung zugewiesen. Die restlichen Bewertungsintervalle können in Prozentschritten definiert werden (z.B. Bewertung 4 zwischen 60 % und 80 % des besten Werts). Die zweite Möglichkeit besteht darin, eigene Intervalle in einer Transformations-Tabelle festzulegen. Dies bietet den Vorteil, dass die Grenzen genauer den Umständen bei jedem Kriterium angepasst werden können.

Die gemessenen Werte werden entsprechend den Umrechnungstabellen in vergleichbare Werte umgewandelt. Damit ist die Evaluierung der Alternativen abgeschlossen.

Die Umwandlung stellt für die vom Anwender beurteilten Kriterien kein Problem dar. Sowohl die Bewertung des Entstehungskontexts als auch der Farbqualität können eins zu eins übernommen werden. Schwieriger ist es für die beiden anderen Kriterien, für die eine Transformationstabelle, wie sie in Tabelle 3 zu sehen ist, festgelegt werden muss. Bei der offenen Spezifikation kann hier festgelegt werden, ob ein proprietäres Format entweder relativ schlecht oder nicht akzeptierbar schlecht ist. Im ersten Fall wird einem ›nein‹ der Wert 1 zugewiesen, im zweiten der Wert 0. In diesem Beispiel ist ein proprietäres Format akzeptierbar, darum wird dem ›nein‹ auch der Wert 1 zuge-
teilt.

Bewertung der Farbqualität

| Kriterium | Wert 5 | Wert 4 | Wert 3 | Wert 2 | Wert 1 | N.A. |
|----------------------------|--------|---------|--------|---------|---------|-------|
| Offene Spezifikation | Ja | - | - | - | Nein | > |
| Importkosten / File [Euro] | 0 | 0.1-0.5 | 0.6-1 | 1.1-1.5 | 1.6-2.6 | > 2.6 |

Tabelle 3: Transformationstabelle für das Beispiel

Ziele gewichten

Der letzte Teil der adaptierten Nutzwertanalyse ist die Reihung der Alternativen und damit die Abgabe einer Empfehlung für eine der möglichen Lösungen zur langfristigen Datenspeicherung. Der erste Schritt dabei ist, die einzelnen Knoten und Blätter des Entscheidungsbaums zu gewichten, um so die Prioritäten der Anwender festzulegen. Für diese Gewichtung wird jede Ebene jedes Teils des Entscheidungsbaums einzeln betrachtet und mit Prozentwerten gewichtet. Im Falle des generischen Top down-Baums wäre die erste Aufgabe, die Elemente der ersten Ebene zu gewichten, also ›Objekt-Eigenschaften‹, ›Technische Eigenschaften‹, ›Prozess-Eigenschaften‹ und ›Kosten‹.

Transformierung in vergleichbare Zahlen

**unterschiedliche
Sichtweisen der Ressorts**

Zwischen diesen drei Knoten werden die ersten 100 % aufgeteilt. Im nächsten Schritt beginnt man, die Ebenen des ›Objekt- Eigenschaften‹-Baums zu bewerten, also die 100 % auf den Inhalt, die Struktur, das Verhalten usw. der Datei zu verteilen. In dieser Form werden alle Knoten und Blätter des Baums bewertet. Die Bewertung sollte von mehreren Mitarbeitern aus unterschiedlichen Bereichen der anwendenden Organisation durchgeführt werden, da hier oft unterschiedliche Sichtweisen bestehen. Während die Finanzabteilung wohl Kosten als die wichtigste Größe sieht, spielen diese für einen Archivar unter Umständen eine untergeordnete Rolle. Bei größeren Differenzen könnten die Ergebnisse der Alternativen in mehreren Durchläufen mit unterschiedlichen Gewichtungen bewertet werden.

Noch ein zweiter Schritt ist notwendig, bevor die Zielwerte aggregiert werden können: Nachdem nur die Blätter des Baums evaluiert worden sind, ist für die nächsten Schritte eine Berechnung der individuellen Gewichte der Blätter notwendig. Um diese Gewichte zu erhalten, wird die erste Gewichtung des Blatts mit der prozentuellen Gewichtung aller übergeordneten Knoten multipliziert. Das Produkt ist das endgültige Gewicht eines Blattes. (Die Summe der Gewichtung aller Blätter sollte dann eins ergeben.)

**Gewichtung des
Beispielbaums**

Bei der Gewichtung des Beispielbaums haben sich die Techniker, Archivare und Finanzmitarbeiter auf folgende Gliederung geeinigt: Objekt-Eigenschaften 60 %, Technische Charakteristika und Kosten mit jeweils 20 %. Eine Gewichtung auf einer tieferen Ebene ergibt sich nur bei den Eigenschaften des digitalen Objekts, hier wird der Entstehungskontext mit 40 % und die Farbqualität mit 60 % gewichtet.

Nach der Multiplikation der Gewichtungen der Blätter mit ihren übergeordneten Knoten ergeben sich folgende endgültige Gewichte: Farbqualität mit 0.36, Entstehungskontext mit 0.24, die offene Spezifikation mit 0.2 und die Importkosten mit 0.2.

einzigste Gesamtkennzahl pro Alternative berechnet werden. Diese Gesamtkennzahl bildet die Grundlage für die abschließende Reihung der Alternativen. Zur Berechnung dieser Kennzahl werden zuerst die Gewichtung und der vergleichbare Wert miteinander multipliziert, um einen einzelnen Wert pro Kriterium und Alternative zu erhalten. Durch die Addition all dieser Werte pro Alternative kann der Zielwert aggregiert werden.

Bei der Verwendung einer Skala von null bis fünf für die vergleichbaren Zahlen liegt auch der Zielwert zwischen 0 und 5. Die Relation zu fünf zeigt an, in wie weit eine – entsprechend den vorher definierten Kriterien und Transformationen – optimale langfristige Datenspeicherungsmethode gefunden wurde oder nicht.

Auch für Alternativen, von denen ein oder mehrere Kriterien mit null, also ›nicht akzeptierbar‹ bewertet sind, sollte der Zielwert berechnet werden, um zum einen Referenzwerte für die Güte der anderen Ergebnisse zu erhalten und zum zweiten zu wissen, wie gut diese Alternativen ohne die ›nicht akzeptierbaren‹ Defizite abgeschnitten hätten.

Die Aggregation ergibt Tabelle 4, in der sich die vorige Evaluierung deutlich widerspiegelt.

Alternativen reihen

Der letzte nicht mehr aufwändige Schritt ist die Reihung der Alternativen nach der Höhe ihres Zielwerts. Bevor eine endgültige Empfehlung abgegeben wird, sollten noch zusätzliche Faktoren, die die Entscheidung beeinflussen, aber nicht in den Entscheidungsbaum aufgenommen wurden, berücksichtigt werden. Solche Kriterien wären zum Beispiel eine besonders gute Beziehung zu einem Anbieter oder die gute Kenntnis des Umgangs mit einer bestimmten Speicherlösung. Das Endergebnis dieses Schritts ist eine Liste von langfristigen Speicherstrategien mit einer Kennzahl und einer kurzen Begründung der Reihung, falls diese von den Zahlen abweicht.

Die Reihung der drei Alternativen ergibt, dass sich die Migration auf PDF 1.4 am besten für den nächsten Schritt zur langfristigen Datenspeicherung eignen würde. Den zweiten Platz belegt das Computer-Museum, den letzten die Migration innerhalb von Word. (Diese Reihung gilt selbstverständlich nur unter den gegebenen Umständen, mit den hier festgelegten vier Kriterien, der hier definierten Gewichtung und den Werten, die in dem Beispielarchiv gelten.) Wie man sieht, wirkt sich das proprietäre Format relativ stark auf die Zielwerte aus, während die höheren Importkosten durch die relativ schwach sinkende Transfor-

| Kriterium / Strategie | MS Word 2002 | PDF 1.4 | Computer-Museum |
|-----------------------|--------------|-------------|-----------------|
| Farbqualität | 1.8 | 1.44 | 1.8 |
| Entstehungskontext | 1.2 | 1.2 | 1.2 |
| Offene Spezifikation | 0.2 | 1 | 0.2 |
| Importkosten / File | 0.8 | 0.8 | 1 |
| Zielwert | 4 | 4.44 | 4.2 |

Tabelle 4: Die aggregierten Werte pro Alternative

Zielwerte aggregieren

Auf Basis der vergleichbaren Werte und der Gewichtungen der einzelnen Blätter kann als nächstes eine

mationstabelle nicht so sehr ins Gewicht fallen. Selbst wenn das proprietäre Format als ›nicht akzeptierbar‹ eingestuft worden wäre, hätte sich in der Reihenfolge nichts geändert. Welche Lösung tatsächlich umgesetzt wird, hängt von den begleitenden Umständen ab. Aus der Nutzwertanalyse lässt sich aber eine klare Empfehlung ableiten, die mit direkt sichtbaren Argumenten hinterlegt, sorgfältig abgewogen ist und sich daher optimal als Entscheidungsgrundlage eignet.

ZUSAMMENFASSUNG

Die große Anzahl an langfristigen Speicherstrategien wie Migration, Emulation oder Computer-Museen und deren Implementierungen, von denen keine unter allen Umständen besser als eine andere ist, machen die Entscheidung für eine von ihnen schwer. Um diese Entscheidung auf eine argumentierbare Basis zu stellen und um die langfristige Bewahrung der Daten nicht zu einer Gefühlssache werden zu lassen, wurde die hier vorgestellte adaptierte Nutzwertanalyse entwickelt. Sie bietet die Möglichkeit, anhand eines klar definierten Prozesses und einer ausführlich beschriebenen Entscheidungsgrundlage in Form des Entscheidungsbaums die optimale Wahl für die jeweiligen Anforderungen und Möglichkeiten zu treffen. Trotz des relativ hohen Aufwands, der für die Entwicklung des Entscheidungsbaums verwendet werden muss, hat sich die adaptierte Nutzwertanalyse in mehreren praktischen Implementierungen für die Video- und Audio-Sammlung des österreichischen Phonogrammarchivs und für die elektronischen Akten des niederländischen Nationalarchivs bewährt und die Basis für eine fundierte, argumentierbare und nachverfolgbare Entscheidung über die Art der langfristigen Datenspeicherung geschaffen.

DANKSAGUNG

Teile der hier präsentierten Arbeit wurden durch die Europäische Union im 6. Rahmenprogramm des Information Society Technologies Programme (IST) durch das DELOS Network of Excellence, Vertrag 507618, gefördert.

LITERATUR

[1] Bechmann, Arnim: Nutzwertanalyse, Bewertungstheorie und Planung. In: Beiträge zur Wirtschaftspolitik, Vol. 29 (1978)
[2] Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System (OAIS). First Edition. May 1999. – www.ccds.org
[3] Apple Computer, Inc.: G5 Processor Architecture. 2004. – www.apple.com/g5processor/architecture.html

[4] GRAU Data Storage AG: Infinistore Tape Library – Model XL. 2003. – www.graadatastorage.de

[5] Hanusch, Horst; Biene, Peter; Schlumberger, Manfred: Nutzen-Kosten-Analyse. Verlag Franz Vahlen, 1987

[6] Kranch, Douglas A.: Beyond Migration: Preserving Electronic Documents with Digital Tablets. In: Information Technologies Libraries 17 (1998), Nr. 3. – ISSN 0730-9295

[7] Lawrence, Gregory W.; Kehoe, William R.; Rieger, Oya Y.; Walters, William H.; Kenney, Anne R.: Risk Management of Digital Information: A File Format Investigation. Washington D.C.: Council on Library and Information Resources, 2000. – ISBN 1-887334-78-5

[8] Lorie, Raymond: The UVC: a Method for Preserving Digital Documents – Proof of Concept / IBM Netherlands, Amsterdam. 2002. – Forschungsbericht. – ISBN 90-6259-157-4

[9] Norsam Technologies: HD-Rosetta Data Preservation. January 2004. – www.norsam.com/hdrosetta.htm

[10] Ogle, Virginia; Wilensky, Robert: Testbed Development for the Berkeley Digital Library Project. In: D-LIB Magazine (1996), July/August. – www.dlib.org. – ISSN 10829873

[11] Oltmans, Erik; Diessen, Raymond J.; Wijngaarden, Hilde van: Preservation Functionality in a Digital Archive. In: Proceedings of the Joint Conference on Digital Libraries, June 7–11, Tuscon, Arizona, USA, ACM, 2004, S. 279–286

[12] National Library of Australia: Preserving Access to Digital Information (PADI) Migration. Website, December 2003. – www.nla.gov.au/padi/topics/21.html

[13] Rauber, Andreas; Aschenbrenner, Andreas: Part of Our Culture is Born Digital – On Efforts to Preserve it for Future Generations. In: TRANS. On-line Journal for Cultural Studies (Internet-Zeitschrift für Kulturwissenschaften) 10 (2001), July. – www.inst.at/trans/10Nr/inhalt10.htm

[14] Rauch, Carl; Rauber, Andreas: Preserving digital media: Towards a preservation solution evaluation metric. In: Proceedings of the 7th International Conference on Asian Digital Libraries, ICADL 2004, Springer, December 2004, S. 203–212

[15] retroK; Azrail: AEP Emulation page. 2004. – www.aep-emu.de

[16] Rothenberg, Jeff: Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Council on Library and Information Resources Washington D.C., 1999. – www.clir.org/pubs/reports/rothenberg/contents.html. – ISBN 1887334637

[17] Rothenberg, Jeff: Carrying Authentic, Understandable and Usable Digital Records Through Time. Leiden:

Nutzwertanalyse als optimale Entscheidungsgrundlage

erfolgreicher Einsatz in Österreich und den Niederlanden

RAND Europe, 1999. – Report to the Dutch National Archives and Ministry of the Interior

[18] Slats, Jacqueline; Verdegem, Remco: Digital Preservation Costmodel / Nationaal Archief. 2005. – Forschungsbericht. www.digitaleduurzaamheid.nl/home.cfm

[19] Stinson, Douglas; Ameli, Fred; Zaino, Nick: Lifetime of KODAK Writable CD and Photo CD Media. Rochester, New York State: KODAK, Digital & Applied Imaging, 1995. – www.cd-info.com/CDIC/Technology/CDR/Media/Kodak.html

[20] UNESCO, Information Society Division: Guidelines for the preservation of digital heritage. October 2003. – www.unesco.org/webworld/mdm

¹ Stinson, Douglas; Ameli, Fred; Zaino, Nick: Lifetime of KODAK Writable CD and Photo CD Media. Rochester, New York State: KODAK, Digital & Applied Imaging, 1995. – www.cd-info.com/CDIC/Technology/CDR/Media/Kodak.html

² GRAU Data Storage AG: Infinistore Tape Library – Model XL. 2003. – www.graadatastorage.de

³ Rauber, Andreas; Aschenbrenner, Andreas: Part of Our Culture is Born Digital – On Efforts to Preserve it for Future Generations. In: TRANS. On-line Journal for Cultural Studies (Internet-Zeitschrift für Kulturwissenschaften) 10 (2001), July. – www.inst.at/trans/10Nr/inhalt10.htm

UNESCO, Information Society Division: Guidelines for the preservation of digital heritage. October 2003. – www.unesco.org/webworld/mdm

⁴ Kranch, Douglas A.: Beyond Migration: Preserving Electronic Documents with Digital Tablets. In: Information Technologies Libraries 17 (1998), Nr. 3. – ISSN 0730-9295

⁵ Norsam Technologies: HD-Rosetta Data Preservation. January 2004. – www.norsam.com/hdrosetta.htm

⁶ National Library of Australia: Preserving Access to Digital Information (PADI) Migration. Website, December 2003. – www.nla.gov.au/padi/topics/21.html

⁷ Lawrence, Gregory W.; Kehoe, William R.; Rieger, Oya Y.; Walters, William H.; Kenney, Anne R.: Risk Management of Digital Information: A File Format Investigation. Washington D.C.: Council on Library and Information Resources, 2000. – ISBN 1-887334-78-5

⁸ Apple Computer, Inc.: G5 Processor Architecture. 2004. – www.apple.com/g5processor/architecture.html

⁹ retroK; Azrail: AEP Emulation page. 2004. – www.aep-emu.de

¹⁰ Rothenberg, Jeff: Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Council on Library and Information Resources Washington D.C., 1999. – www.clir.org/pubs/reports/rothenberg/contents.html. – ISBN 1887334637

¹¹ Lorie, Raymond: The UVC: a Method for Preserving Digital Documents – Proof of Concept / IBM Netherlands, Amsterdam. 2002. – Forschungsbericht. – ISBN 90-6259-157-4; Oltmans, Erik; Diessen, Raymond J.; Wijngaarden, Hilde van: Preservation Functionality in a Digital Archive. In: Proceedings of the Joint Conference on Digital Libraries, June 7–11, Tuscon, Arizona, USA, ACM, 2004, S. 279–286

¹² Bechmann, Arnim: Nutzwertanalyse, Bewertungstheorie und Planung. In: Beiträge zur Wirtschaftspolitik, Vol. 29 (1978) Hanusch, Horst; Biene, Peter; Schlumberger, Manfred: Nutzen-Kosten-Analyse. Verlag Franz Vahlen, 1987

¹³ Rauch, Carl; Rauber, Andreas: Preserving digital media: Towards a preservation solution evaluation metric. In: Proceedings of the 7th International Conference on Asian Digital Libraries, ICADL 2004, Springer, December 2004, S. 203–212

¹⁴ Rothenberg, Jeff: Carrying Authentic, Understandable and Usable Digital Records Through Time. Leiden: RAND Europe, 1999. – Report to the Dutch National Archives and Ministry of the Interior

¹⁵ Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System (OAIS). First Edition. May 1999. – www.ccds.org

¹⁶ Ogle, Virginia; Wilensky, Robert: Testbed Development for the Berkeley Digital Library Project. In: D-LIB Magazine (1996), July/August. – www.dlib.org. – ISSN 10829873

¹⁷ Slats, Jacqueline; Verdegem, Remco: Digital Preservation Costmodel / Nationaal Archief. 2005. – Forschungsbericht. www.digitaleduurzaamheid.nl/home.cfm

DIE VERFASSER

Ao. Univ. Prof. Dr. Andreas Rauber ist Assistent am Institut für Software Technologie und Interaktive Systeme der Technischen Universität Wien, Favoritenstraße 9–11, A–1040 Wien, E-Mail: rauber@ifs.tuwien.ac.at

Mag. Carl Rauch ist Projektassistent am Institut für Software Technologie und Interaktive Systeme der Technischen Universität Wien, Favoritenstraße 9–11, A–1040 Wien, E-Mail: rauch@ifs.tuwien.ac.at