

Jan Frederik Maas

Inhaltserschließung für Discovery-Systeme gestalten

1 Einleitung

Bibliotheken und Informationseinrichtungen erfahren in den letzten Jahren einen Wandel in ihren Aufgaben. Die zunehmende digitale Vernetzung ermöglicht es den Agierenden des Publikations- und Informationswesens, verstärkt die eigene Rolle zu verlassen und in Bereichen tätig zu werden, die bisher nicht dem eigenen Kerngebiet entsprochen haben. So können z. B. Verlage und Autor:innen in Umgehung des Einzelhandels und der Bibliotheken ihre Werke direkt an die Kund:innen verkaufen und ausliefern und Bibliotheken können über Open-Access-Repositorien selbst Werke publizieren.

Dessen ungeachtet ist das Kerngeschäft der Bibliotheken weiterhin das Sammeln, Erschließen und Vermitteln von Werken¹ für ihr jeweiliges Publikum, ergänzt um zusätzliche Tätigkeiten. Diese klassische Auflistung der drei Aufgabengebiete suggeriert, dass die drei Tätigkeitsfelder unabhängig voneinander zu betrachten sind. Das ist aber nicht der Fall – das Vermitteln von Werken kann nur dann stattfinden, wenn das Werk in zugänglicher Form vorliegt (gesammelt wurde) und es durch Erschließung auffindbar gemacht wurde. Die Trennung der drei Aufgabenbereiche ist arbeitsorganisatorisch zunächst hilfreich, da Spezialist:innen für jeden der Bereiche separat ausgebildet und beschäftigt werden können. Auf der anderen Seite entsteht aber die Gefahr, dass durch mangelnde Abstimmung der Bereiche ineffizient gearbeitet wird – zum Beispiel, wenn in der Erschließung formalen Systemen entsprochen wird, deren Entwurfskriterien aber nicht zielführend für die Recherchesysteme sind, die den Vermittlungsprozess ermöglichen. Oder umgekehrt, wenn für ein spezielles Recherchesystem erschlossen wird, aber nach einem Wechsel der Software die Erschließung für das neue Produkt nicht mehr geeignet ist.

Ich möchte in diesem Beitrag dafür argumentieren, Tätigkeiten in Bibliotheken und Informationseinrichtungen – speziell Erschließung und das Entwickeln von Discovery-Systemen – konsequent in Bezug auf *Anwendungsfälle* zu denken und zu gestalten. Nur so lässt sich effektiv und effizient sowohl bei der Erschließung als auch bei der Gestaltung der die Erschließung nutzenden Systeme und Prozesse arbeiten. Zunächst möchte ich dafür in den folgenden Abschnitten Be-

1 Vgl. Hacker, Rupert: Bibliothekarisches Grundwissen. 7. Aufl. München: Saur 2000, S. 12–13.

rührungspunkte aufzeigen, um auf die vielfältigen Probleme aber auch Möglichkeiten im Zusammenspiel von (Inhalts-)Erschließung und Discovery-Systemen hinzuweisen.

2 Resource-Discovery-Systeme (RDS)

Resource-Discovery-Systeme (kurz: *Discovery-Systeme*) ersetzen in den letzten Jahren zunehmend klassische Online-Kataloge² in der Funktion als Recherchesysteme für bibliografische Metadaten, zum Beispiel als Bibliothekskataloge, Recherchesysteme für Forschungsinformationssysteme oder Forschungsdaten. Der Begriff *Discovery-System* wurde in den vergangenen Jahren unterschiedlich verwendet, daher möchte ich in diesem Beitrag folgende Definition verwenden:

Discovery-Systeme sind konkrete bibliothekarische Informationssysteme, die unter anderem aus einem webbasierten Nutzungsinterface und einem auf Suchmaschinentechnologie basierenden Index für bibliografische Metadaten und ggf. für Volltexte bestehen.

Diese an die Wikipedia³ angelehnte Definition ignoriert bewusst, dass Discovery-Systeme ursprünglich das Ziel hatten, speziell entdeckendes Suchen zu fördern,⁴ da unklar ist, ob dies bei den aktuellen Discovery-Systemen wirklich der Fall ist.⁵

² Als klassische Online-Kataloge oder OPACs (Online Public Access Catalogue) bezeichne ich im Rahmen dieses Beitrags, wie im Folgenden geschildert, Bibliothekskataloge, die auf einem relationalen Datenmodell basieren, Boolesche Algebra bei der Anfragesprache unterstützen und nach dem *exact-match*-Paradigma arbeiten (vgl. Steilen, Gerald: *Discovery-Systeme – die OPACs der Zukunft?* Hamburg 2012; 101. Deutscher Bibliothekartag). Einer weiten, hier nicht vertretenen Auslegung des Begriffes nach können auch Discovery-Systeme als OPACs bezeichnet werden.

³ Siehe <https://de.wikipedia.org/wiki/Discovery-System> (12.6.2020).

⁴ Vgl. Blenkle, Martin: *Inhaltserschließung als Navigationspunkte für den Suchprozess. eine Bibliothek gestaltet ihr Discovery System selbst...* Bremen 2013 (Fachreferententagung Wirtschaftswissenschaft), S. 12.

⁵ Viele Discovery-Systeme bieten Funktionen für entdeckendes Suchen (z. B. *more-like-this*-Funktionen oder thematisches Browsing) nicht oder nur als nicht zentrale Funktion an. Auch die für Suchmaschinen typische unscharfe Suche fördert entdeckendes Suchen nicht automatisch besser als klassische OPACs. Insgesamt wird der Begriff *Discovery* in Bezug auf Discovery-Systeme seit einiger Zeit kontrovers diskutiert, vgl. u. a. Koster, Lukas: *Discovery tools: a rearguard action?* Palma de Mallorca 2012 (ELAG2012).

Der wichtigste technologische Unterschied von Discovery-Systemen und OPACs liegt in der Nutzung von Websuchmaschinentechnologie.⁶ Während OPACs die exakte Suche mit Booleschen Operatoren in den Vordergrund stellen (*exact match*), sollen Discovery-Systeme auf Suchanfragen größere Ergebnismengen als OPACs liefern und diese nach einer möglichst guten Relevanzsortierung ausgeben (*best match*). Die Nutzung von Suchmaschinentechnologie ermöglicht so z. B. die Recherche in Volltexten, deren automatische Indexierung⁷ oft zu extrem großen Treffermengen führt. Die Darstellung derselben ohne Relevanzsortierung – wie in OPACs üblich – erzeugt Trefferlisten, in denen die gesuchten Werke ggf. nur mühsam oder mit nicht vertretbarem Aufwand gefunden werden können.

3 Erschließung und Discovery-Systeme

Das Zusammenspiel von bibliothekarischer Erschließung und Discovery-Systemen ist komplex. Suchmaschinen sind im Kern darauf angelegt mit unstrukturierten oder schwach strukturierten Volltexten (in der Regel Webseiten) zu arbeiten. Bibliothekskataloge arbeiten dagegen seit langer Zeit mit Metadaten, die im Gegensatz zu Volltexten hochstrukturiert und deutlich kürzer sind. Um ein Beispiel zu nennen: Elementare Metriken der für Discovery-Systeme typischen Relevanzsortierung wie *term frequency* (tf) und *inverse document frequency* (idf) wurden ursprünglich für die Verarbeitung von nicht strukturierten Volltexten entwickelt. Auf reinen Metadaten ist zumindest die tf nur eingeschränkt aussagekräftig, was die Konfiguration der Relevanzsortierung von Discovery-Systemen erschwert. So kann zum Beispiel ein zentraler Begriff eines Werkes in den Metadaten nur einmal vorkommen (z. B. als Vorzugsbezeichnung eines Sachschlagworts⁸) was zu einer geringen tf führt, während der Begriff im Text des Werkes sehr häufig vorkommt (hohe tf).

Eine seit geraumer Zeit diskutierte Frage ist, ob Erschließung durch die automatische Indexierung von Volltexten grundsätzlich hinfällig ist, da die

⁶ Steilen, Gerald: Discovery-Systeme – die OPACs der Zukunft? Hamburg 2012 (101. Deutscher Bibliothekartag).

⁷ Der Begriff *Indexierung* wird im Rahmen dieses Beitrags für den technischen Prozess verwendet, der Daten in einen Suchindex speichert. Er wird nicht für Prozesse der intellektuellen oder automatischen Erschließung genutzt.

⁸ *Sachschlagwort* bezeichnet in diesem Beitrag i. d. R. einen Sachbegriff im Sinne der *Regeln für die Schlagwortkatalogisierung* (RSWK). Ggf. lassen sich die Aussagen des Beitrags aber fallbezogen auch auf Geografika, Personen, etc. ausweiten.

Suchmaschinentechnologie das Auffinden von Textinhalten anhand von Stichwörtern, *snippets* und – z. B. bei Verwendung von automatischer Textstrukturerkennung – Facetten bei vorhandenen Volltexten auch ohne intellektuelle Erschließung ermöglicht.

Im Rahmen dieses Beitrags kann diese Frage nicht geklärt werden. Aus verschiedenen Gründen ist es aber naheliegend, dass intellektuelle Erschließung auch in Zeiten von automatischer Indexierung wichtig ist und somit auch in den kommenden Jahren eine Kernaufgabe von Bibliotheken sein wird:

1. Erschließung stellt auch in Zeiten der Suchmaschinentechnologie einen Mehrwert für die Informationssysteme dar.⁹
2. Es gibt einen signifikanten Anteil von Bibliotheksbeständen, die den Bibliotheken aus technischen, finanziellen oder rechtlichen Gründen nicht als maschinenlesbarer und indexierbarer Volltext vorliegen. Somit werden durch Erschließungsprozesse generierte Metadaten weiterhin benötigt.

Trotzdem ist es notwendig, dass sich beide Prozesse (Erschließung und die Entwicklung von Discovery-Systemen) aneinander annähern, um gut zusammenarbeiten zu können.

4 Technische und inhaltliche Anforderungen

Im Wesentlichen kommen Inhaltserschließungsmerkmale in den folgenden Bereichen von Discovery-Systemen zur Anwendung:

4.1 Suche

In der Regel werden in Discovery-Systemen nicht-hierarchisch geordnete Metadatenfelder definiert, die bei einer Suche auf das Vorkommen der eingegebenen Suchbegriffe überprüft werden. Kommen – je nach Konfiguration des Discovery-Systems – mindestens ein, einige oder alle Suchbegriffe in den spezifizierten Feldern eines Datensatzes vor, wird der Datensatz zu der Ergebnismenge hinzugefügt. In Abhängigkeit von der Einstellung des Datenindex und des Dis-

⁹ Stumpf, Gerhard: „Kerngeschäft“ Sacherschließung in neuer Sicht. Was gezielte intellektuelle Arbeit und maschinelle Verfahren gemeinsam bewirken können. Augsburg: Universitätsbibliothek 2015. S. 2; Wiesenmüller, Heidrun: Resource Discovery Systeme – Chance oder Verhängnis für die bibliothekarische Erschließung? Hildesheim 2012 (Jahrestagung der Deutschen Gesellschaft für Klassifikation).

covery-Systeme sind auf bestimmten oder allen Feldern Trunkierungen und das Recherchieren ohne Beachtung von Groß- und Kleinschreibung möglich.

Gerade Inhaltserschließungsmerkmale können bei der Suche nach Werken sehr hilfreich sein, indem sie die Menge an gefundenen relevanten Katalogeinträgen erhöhen und somit den durchschnittlichen *recall* des Discovery-Systems verbessern. Dies geschieht z. B. durch die ergänzende Indexierung von (bevorzugten) Benennungen von Sachschlagwörtern, die noch nicht in den Datenfeldern für die Formalerschließung oder in den Volltextfeldern zu finden sind. Durch die Verwendung von kontrollierten Vokabularen mit entsprechenden Ergänzungen können so auch Synonyme oder Oberbegriffe gesucht und gefunden werden.

Discovery-Systeme bieten als wichtigsten Sucheinstieg meist einen einfachen Suchschlitz für eine reine Stichwortsuche an. Aus Sicht der Discovery-Systeme ergibt sich daher folgende Anforderung an die Katalogisierung von Sachschlagwörtern: Die indexierten Vorzugsbezeichnungen (Synonyme, etc.) sollten den Erwartungen bzw. der Wortwahl der Nutzer:innen für diese Form der Suche entgegenkommen. Ein Beispiel: Wenn in der Teildisziplin der Informatik, die sich mit maschinellem Lernen beschäftigt, meistens englische Begriffe verwendet werden (*machine learning*, *artificial intelligence*) sollten diese in den Inhaltserschließungsmerkmalen zu finden sein, damit die Nutzer:innen die relevanten Dokumente erfolgreich recherchieren können.

Die Erweiterung des Suchraums um Gattungsbegriffe, Oberbegriffe, Synonyme etc. kann je nach Anwendungsprofil ebenfalls Nutzen bringen – vor allem bei Synonymen und alternativen Schreibweisen. Suchmaschinen verfügen zwar über Funktionen wie Stemming, mit dem durch die Reduktion der Schlagwörter und Suchbegriffe auf den Wortstamm grammatische Eigenschaften entfernt werden, so dass auch z. B. flektierte Suchbegriffe gefunden werden können. Allerdings werden auf diese Weise sich im Wortstamm unterscheidende alternative Formen – wie z. B. *Apfelsine* und *Orange* – nicht gefunden. Techniken, die mit Hilfe von *machine learning* und/oder computerlinguistischen Verfahren solche Varianten identifizieren können, benötigen entweder umfangreiche Trainingsdatenmengen, die nicht immer vorhanden sind, oder Lexika, die intellektuell gepflegt werden müssen. Auch aus diesem Grund stellen gut inhaltsererschlossene Katalogisate und gut gepflegte kontrollierte Vokabulare nach wie vor eine sinnvolle bis notwendige Ergänzung für Discovery-Systeme dar.

Probleme bei der Suche können insbesondere bei heterogener Erschließung entstehen.¹⁰ Discovery-Systeme sind für diese besonders anfällig, da sie üblicherweise Metadaten aus mehreren unterschiedlich erschlossenen Quellen

¹⁰ Vgl. Wiesenmüller 2012.

zusammenführen. So kann die (absichtliche oder unabsichtliche) Suche nach Schlagwörtern aus einem Vokabular dazu führen, dass mit anderen Vokabularen oder gar nicht inhaltserschlossene Katalogisate nicht gefunden werden. Aus der Sicht der Discovery-Systeme ist daher eine möglichst homogene Erschließung mit hoher Abdeckung wünschenswert. In der Realität ist dieses Qualitätskriterium aufgrund von riesigen, heterogenen Indices und dem Wandel der Zeit unterlegenen Datenbeständen wohl nie vollständig zu erfüllen. Allerdings führt hier bereits eine schrittweise Verbesserung schon zu besseren Retrieval-Ergebnissen und die Anwendung von Cross-Konkordanzen und ähnlichen Techniken kann die Situation bestehender Datenmengen bereits deutlich verbessern.

4.2 Relevanzsortierung

Mittels Relevanzsortierungen sollen Datensätze innerhalb einer Trefferliste so sortiert werden, dass die für den Nutzenden relevantesten Datensätze am Anfang stehen.¹¹ In Technologien wie Apache Solr¹² wird dazu in Abhängigkeit von einer Suchanfrage pro Dokument ein numerischer Wert der Relevanz errechnet. Die Reihenfolge dieser Werte entspricht der Relevanzsortierung der Treffermenge.

Zur Ermittlung des Relevanzwerts werden komplexe Berechnungen angestellt, die von der verwendeten Suchmaschinentechnologie und der Konfiguration des Discovery-Systems abhängen. Es gibt verschiedene Parameter, die zur Einstellung der Relevanzberechnung herangezogen werden können. Im Folgenden eine kurze, keinesfalls vollständige Liste:

- *tf (term frequency)*: Die Häufigkeit des Vorkommens eines Suchbegriffs in den Datensätzen erhöht den Relevanzwert des Datensatzes. Dies ist teilweise kritisch zu sehen, da tiefer erschlossene Werke auf diese Weise oft einen höheren Relevanzwert erhalten als andere, die aber ggf. relevanter sind.
- *idf (inverse document frequency)*: Der Anteil der Dokumente, in denen ein Suchbegriff vorkommt, reduziert die positiven Auswirkungen des Suchbe-

11 Vgl. Lewandowski, Dirk: Web Information Retrieval. Technologien zur Informationssuche im Internet. Frankfurt a. M.: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e. V. 2005 (DGI-Schrift (Informationswissenschaft) 7).

12 Apache Solr ist eine Open-Source-Suchmaschinentechnologie, die sehr performant mit sich selten ändernden Datenmengen arbeitet und die daher gut für bibliografische Metadaten geeignet ist. Sie kommt zum Beispiel in dem Open-Source-Discovery-Framework Vufind zum Einsatz. Das wichtigste quelloffene Konkurrenzprodukt ist aktuell Elasticsearch, das genau wie Apache Solr auf Basis der Programmbibliothek Lucene arbeitet. Vgl. <https://lucene.apache.org/solr/> (24.7.2020).

- griffs auf den Relevanzwert. Füllwörter (*und, oder, die, etc.*) werden meist zusätzlich über Negativlisten aus der Relevanzberechnung ausgeschlossen.
- Phrasenfelder: Treffer in spezifizierten Feldern (z. B. Titel, Körperschaftsbezeichnungen), in denen Suchbegriffe nahe beieinanderstehen, können zu einer Erhöhung des Relevanzwertes führen. Dies ist z. B. hilfreich, wenn im Rahmen einer *known item search* exakte Titel in der Suche eingegeben werden.
 - Feld-Boosting: Einzelne Datenfelder können mit Boosting-Faktoren belegt werden, die auf Suchbegriffe angewandt werden, die in ihnen vorkommen. So kann z. B. bewirkt werden, dass Dokumente mit Suchbegriffen im Titel früher in der Ergebnisliste angezeigt werden als Dokumente, bei denen dieselben Suchbegriffe nur im Abstract vorkommen.
 - Boosting-Funktionen: Es können Formeln spezifiziert werden, deren Ergebnisse zum Relevanzwert addiert werden. In Discovery-Systemen kann dies beispielsweise genutzt werden, um aktuellere Werke anhand ihres Erscheinungsjahres oder ihrer Auflage in den Trefferlisten weiter oben einzusortieren. Für diese und ähnliche Anwendungsfälle ist es jedoch nötig, dass die Werte maschineninterpretierbar sind – dass also z. B. die Auflagenkennzeichnung durch einen numerischen Wert angegeben ist.

Somit ist die konkrete Verwendung von (Inhalts-)Erschließungsmerkmalen für die Berechnung des Relevanzwertes abhängig von der individuellen Konfiguration des Discovery-Systems. Es ist naheliegend, mittels Feld-Boosting mindestens Felder wie Haupttitel, Autor:in und Felder mit bevorzugten Benennungen von Schlagwörtern in der Annahme zu begünstigen, dass Nutzer:innen bei Recherchen oft nach diesen Merkmalen recherchieren. Als Ergebnis wird dann z. B. bei einer Suche nach *Faust Goethe* Goethes *Faust* in der Trefferliste am Anfang stehen, während Werke, bei denen die beiden Begriffe nur im Volltext vorkommen, weiter unten angezeigt werden.

Inhaltserschließungsmerkmale spielen bei *known-item*-Suchen – also bei Suchen nach konkreten Werken, bei denen i. d. R. Titel und Autor:in bereits bekannt sind – üblicherweise keine Rolle. Sie werden aber umso wichtiger, wenn thematisch recherchiert wird.

Wenn über Erschließungsmerkmalen zusätzlich inhaltliche Boosting-Funktionen definiert werden, ist es nötig, dass die Merkmale nicht nur maschinenlesbar, sondern auch maschineninterpretierbar vorliegen (vgl. Abschnitt *Maschinenlesbarkeit und -interpretierbarkeit* weiter unten). Klassischerweise muss diese Anforderung vor allem von Formalerschließungsmerkmalen erfüllt werden, z. B. dem Erscheinungsjahr oder der Auflagenkennzeichnung.

Probleme bei der Verwendung speziell von Inhaltserschließungsmerkmalen für die Relevanzberechnung entstehen z. B. dann, wenn nur ein Teil der Katalogisate inhaltlich erschlossen ist und ggf. noch zusätzlich mehrere Vokabulare zum Einsatz kommen. Die nicht oder nicht mit allen Vokabularen erschlossenen Katalogisate können dann allein aufgrund der Erschließungssituation und nicht aufgrund inhaltlicher Kriterien als weniger oder gar nicht relevant für eine Suchanfrage eingestuft werden. Als Beispiel: Drei Aufsätze zum Thema Systemtheorie wurden unterschiedlich erschlossen. Bei Aufsatz A kommt das Wort *Systemtheorie* in den Metadaten überhaupt nicht vor, da dieser nicht inhaltserschlossen wurde und das Wort auch nicht in den Metadaten der Formalerschließung enthalten ist. Bei Aufsatz B kommt *Systemtheorie* nur in den Metadaten der Inhaltserschließung einmal als bevorzugte Bezeichnung vor. Bei Aufsatz C ist es über verschiedene Inhaltserschließungssysteme mehrfach in die Metadaten aufgenommen worden. Ein klassisch konfiguriertes Discovery-System würde bei einer Suche nach *Systemtheorie* Aufsatz C der Ergebnisliste oben anzeigen (aufgrund der tf), Aufsatz B an einer späteren Stelle (geringere tf) und Aufsatz A überhaupt nicht, obwohl die Aufsätze für den Recherchierenden eventuell gleich relevant sind. Selbst ein Verzicht auf die tf als Rankingkriterium würde das Problem nur teilweise lösen. Außerdem kann die Information, dass ein Schlagwort bei einer bestimmten Inhaltserschließung *nicht* vergeben wurde, für ein Ranking zumindest theoretisch ebenfalls wichtig sein. Discovery-Systeme können aber nicht von alleine zwischen einer absichtlichen Nichtvergabe und einer Nichtvergabe aufgrund von Personalknappheit, veränderter Erschließungspraxis, reduzierter Erschließungstiefe, o. ä. differenzieren.

Die planhafte Nichtanwendung von Regelwerken zur Inhaltserschließung auf bestimmte Literaturgattungen (z. B. die RSWK auf Belletristik) kann dieses Problem theoretisch auch verursachen, in der Praxis sind aber Fälle wie der oben geschilderte am problematischsten.

Ein oben schon angedeuteter Spezialfall dieses Problems entsteht durch heterogene Erschließung im zeitlichen Verlauf: Wenn z. B. neue Katalogisate aufgrund von veränderten Erschließungspraktiken sparsamer erschlossen werden, werden ggf. ältere, besser erschlossene Datensätze als relevanter einsortiert oder gar nicht gefunden. Dies ist besonders ärgerlich, wenn auf diese Weise ältere Auflagen desselben Werkes in der Ergebnisliste den neueren Auflagen vorangestellt werden.

Aus Sicht des Anwendungsfalls *Relevanzsortierung* ergibt sich also eine Qualitätsanforderung an die Inhaltserschließung: Eine möglichst homogene Inhaltserschließung mit möglichst hoher Abdeckung ist erstrebenswert.

4.3 Facettierung

Facettierung ist eine Standardfunktion in vielen Informationssystemen. Für die Facettierung wird i. d. R. ein einzelnes Metadatenfeld herangezogen, dessen häufigste Werte in der umgekehrten Reihenfolge ihrer Häufigkeit angezeigt werden. Die Werte werden dann oft zusammen mit der Anzahl der passenden Dokumente links oder rechts von der Ergebnisliste angezeigt. Ein Klick auf diese Werte (Facetten) führt dazu, dass eine neue Suchanfrage durchgeführt wird, bei der die schon vorhandene Treffermenge auf die passenden Dokumente eingeschränkt wird (*Drill-Down*). Das Auswählen mehrerer Facetten desselben Datenfeldes führt in manchen Informationssystemen zur Erweiterung der Treffermenge um die passenden Dokumente, in anderen zu einer weiteren Beschränkung der Treffermenge.

Die Vorzugsbezeichnungen von Schlagwörtern eignen sich oft nur eingeschränkt für die Facettierung, da Sachschlagwort-Datenfelder meistens eine sehr hohe Zahl von möglichen Merkmalen mit oft nur wenigen dazugehörigen Dokumenten besitzen. Aus Gründen der Performance und der Usability sollten aber initial nur wenige Facetten eines Datenfeldes angezeigt werden (z. B. 3–10), so dass bei Sachbegriffen die Gefahr groß ist, dass relevante Facetten nicht angezeigt werden. Dies gilt sowohl für freie Deskriptoren als auch für kontrollierte Vokabulare. Trotzdem können je nach Vokabular und Anwendungsfall auch Sachbegriff-Facetten hilfreich sein.

Klassifikatorische Inhaltserschließungsmerkmale – also zum Beispiel Klassen der DDC vor dem Dezimalpunkt oder Stellen der Basisklassifikation – können dagegen für die Facettierung nach Themengebieten von sehr großem Nutzen sein. So kann z. B. eine Suchanfrage nach *Java* mit einem Klick auf eine thematische Facette entweder auf die Insel oder die Programmiersprache eingeschränkt werden, indem das jeweils zugehörte Sachgebiet (z. B. *Angewandte Informatik*) ausgewählt wird. Leider gibt es verschiedene Einschränkungen, aus denen der Nutzen auch dieser Merkmale für die thematische Facettierung verringert werden kann:

- Geringe Abdeckung: Nutzer:innen erwarten, nach einem Klick auf z. B. die thematische Facette *Geschichte Griechenlands* alle Werke zu finden, die zu der bisherigen Suchanfrage und diesem Thema gehören. Wenn aber ein Teil der Katalogeinträge nicht thematisch erschlossen ist, können die dazugehörigen Werke über diese Facette auch nicht gefunden werden. Trotzdem suggeriert die Facette fälschlicherweise Vollständigkeit. Dies ist aus Sicht der Discovery-Systeme ein starkes Argument dafür, für die zur Facettierung

herangezogene klassifikatorische Inhaltserschließung Abdeckungen von ~100 % anzustreben.

- Hierarchieabhängige Bedeutungen: In manchen Klassifikationen erschließt sich die Bedeutung einer Stelle nur aus der übergeordneten Hierarchie. Ein Beispiel aus der Regensburger Verbundklassifikation (RVK): Die Bedeutung der Notation *AN 59100 Altertum* kann nur verstanden werden, wenn man weiß, dass sie *AN 59000 – AN 59600 Bibliotheksgeschichte* untergeordnet ist – es handelt sich also um die Bibliotheksgeschichte des Altertums, nicht um das Thema *Altertum* insgesamt. Streng genommen müssen auch alle weiteren darüber liegenden Knoten für die Interpretation herangezogen werden. Eine Facettenbezeichnung in einem Discovery-System müsste also alle Bezeichner aus einer Hierarchie anzeigen, damit Nutzer:innen verstehen, um welches Thema es geht. Leider sind Facettenbezeichnungen wie *A – Allgemeines; AN – Buch- und Bibliothekswesen, Informationswissenschaft; AN 50000 – AN 89900 Bibliothekswesen; AN 58000 – AN 64950 Biographie, Geschichte; AN 59000 – AN 59600 Bibliotheksgeschichte; AN 59100 Altertum* aber aus Gründen der Lesbarkeit völlig ungeeignet für eine Facette.
- Zu detailliert: Bei Sachklassifikationen, die in Bezug auf den jeweiligen Datensatz zu feingranular sind, kann das weiter oben bei Sachbegriffen erwähnte Problem auftreten.
- Mangelnde Aktualität: Auch bei Facetten aus Sachklassifikationen gilt, dass sie den Sucherwartungen der Nutzer:innen entsprechen sollten. Nicht oder schlecht gepflegte Klassifikationen verlieren daher mit der Zeit ihren Nutzen für den Einsatz als Facetten.

Trotz dieser möglichen Einschränkungen sind Sachklassifikationen eine sehr hilfreiche Facettierung in Discovery-Systemen. Wenn ein Discovery-System eine Facettierung ohne Angabe von Suchbegriffen zulässt und der Abdeckungsgrad durch die Sachklassifikation 100 % beträgt, lassen sich sogar alle Werke eines Themas im Discovery-System finden, was für wissenschaftliche Recherchen sehr nützlich sein kann.

Als Fazit ist für den Anwendungsfall *Thematische Facettierung* also eine klassifikatorische thematische Erschließung wünschenswert, die für die Nutzer:innen verständliche, kompakte Klassenbezeichnungen hat und die einen sehr hohen Abdeckungsgrad besitzt.

4.4 Darstellung und Vernetzung

Inhaltserschließungsmerkmale eines Datensatzes sind eine wichtige Information für Recherchierende. Aus diesem Grund sollten sie trivialerweise bei der Anzeige der Metadaten der jeweiligen Datensätze mit aufgeführt werden.

Über die reine Anzeige hinaus ist es sinnvoll, dass die Nutzer:innen verstehen, was sie angezeigt bekommen. Dies ist bei Vorzugsbezeichnungen und Textbeschreibungen (Inhalten) von Notationen von Sachklassifikationen normalerweise unproblematisch, bei Schlagwortfolgen oder reinen Angaben der Klassen dagegen schwieriger. Somit lässt sich aus Sicht der Discovery-Systeme die – zugegeben triviale – Anforderung ableiten, dass Bezeichnungen von Inhaltserschließungsmerkmalen für den jeweiligen Nutzendenkreis verständlich sind und dass weniger verständliche Informationen weggelassen oder mit verständlichen ergänzenden Informationen versehen werden.

Die Vernetzung anhand von Inhaltserschließungsmerkmalen geht über die reine Darstellung hinaus und kann sich individuell in Bezug auf das Discovery-System unterscheiden. Eine einfache Form der Vernetzung besteht z. B. darin, dass durch einen Klick auf ein angezeigtes Schlagwort alle Katalogeinträge präsentiert werden, die dieses Schlagwort besitzen. Komplexere Formen der Vernetzung bieten z. B. eine Navigation anhand von Ober- und Unterbegriffen an, verweisen auf eine detaillierte Beschreibungsseite des Merkmals oder verlinken in andere webbasierte Systeme.

Die Herausforderung bei der Vernetzung liegt vor allem darin, sowohl Aspekten der Nutzungsfreundlichkeit als auch den Eigenschaften des Inhaltserschließungssystems zu entsprechen. So kann ein kontrolliertes Vokabular paradigmatische Beziehungen in Form von Verknüpfungen zu Unter- und Oberbegriffen enthalten – aber ist es für die Nutzer:innen des Discovery-Systems hilfreich, diese zu sehen? Vielleicht sind solche Beziehungen auch nur für einen Teil der Nutzenden interessant und können hinter Ausklappmenüs verborgen oder ganz weggelassen werden. Eine Kennzeichnung der semantischen Beziehungen durch Semantic-Web-Ausdrücke (geeignete RDF-Triples/Vokabulare) ermöglicht an dieser Stelle, allgemeingültige Festlegungen zu treffen, die dann alle verwendeten Vokabulare betreffen – also z. B. alle gefundenen Oberbegriffe aller verwendeten Erschließungssysteme anzuzeigen oder auszublenden.

Ein enger Austausch von Spezialist:innen aus IT-Entwicklung und Inhaltserschließung ist notwendig, um die vernetzte Darstellung von Inhaltserschließungsmerkmalen in Discovery-Systemen nutzungsfreundlich zu gestalten. Darin unterscheiden sich Discovery-Systeme nicht von klassischen OPACs. Der

laufende Paradigmenwechsel vom OPAC zum Discovery-System macht es jedoch nötig, diesen Prozess der Darstellungsoptimierung erneut zu vollziehen, was ermüdend scheinen mag. Leider ist dieser Prozess aber notwendig. Strukturiert notierte Empfehlungen zur Präsentation von Inhaltserschließungsmerkmalen in Informationssystemen können dabei unterstützen, Teile dieses Vorgangs zu beschleunigen. Diese müssten aber in einem intellektuellen Prozess angefertigt werden, der wie oben beschrieben stark durch die Verwendung von Semantic-Web-Technologien unterstützt werden kann.

4.5 Ähnliche Treffer-Funktion

Ein Spezialfall der Vernetzung ist die *more-like-this*- oder *ähnliche-Treffer*-Funktion.¹³ Hierfür werden bei der Detailanzeige eines Katalogeintrags andere *ähnliche* Einträge angezeigt, um Nutzer:innen zum Stöbern zu motivieren. Für diese Funktion werden im Discovery-System Indexfelder definiert, die sich gut für die Ermittlung von inhaltlicher Ähnlichkeit von Datensätzen eignen. Bei der Anzeige eines Katalogeintrags werden dann anhand von mathematischen Distanzmaßen andere Einträge gefunden, die mit dem angezeigten Eintrag in Bezug auf die spezifizierten Felder möglichst gut übereinstimmen.

Es liegt auf der Hand, für diese Ähnlichkeitsberechnung Inhaltserschließungsfelder hinzuzuziehen, ggf. ergänzt durch Titelfelder etc. Dies setzt allerdings wiederum voraus, dass diese Felder im Informationssystem einen hohen Abdeckungsgrad besitzen und hinreichend spezifisch sind, um die Ähnlichkeit gut ermitteln zu können. Unterschiedliche Vokabulare können den Nutzen von Inhaltserschließungsmerkmalen ebenfalls mindern, sofern keine Cross-Konkordanz zum Einsatz kommt.

4.6 Maschinenlesbarkeit und -interpretierbarkeit

Diese beiden Kriterien wurden weiter oben zwar schon erwähnt, sind aber wichtig genug, um sie noch einmal separat anzuführen: Für die Verarbeitung von Erschließungsmerkmalen ist es zum einen trivialerweise notwendig, dass die Daten maschinenlesbar vorliegen – also z. B. als Textfelder oder Zahlen in einer Datenbank. Zum anderen ist es aber für bestimmte Anwendungszwecke auch

¹³ Vgl. Apache Software Foundation: Apache Solr Reference Guide. MoreLikeThis. Version 8.4. https://lucene.apache.org/solr/guide/8_4/morelikethis.html (29.5.2020).

notwendig, dass die Daten maschineninterpretierbar sind, also dass auch ihre Bedeutung softwareseitig verarbeitet werden kann.

Dazu ist es zunächst sinnvoll, die Daten in so viele individuelle Datenfelder aufzutrennen, wie nötig – also möglichst auf Trennungen innerhalb von Feldern durch feldspezifische Trennzeichen wie Doppelpunkte zu verzichten. Die resultierenden Feldinhalte sollten idealerweise so gut wie möglich automatisch interpretierbar sein. Ein Beispiel aus der Formalerschließung wurde oben schon erwähnt: Auflagenkennzeichnungen, die Auswirkungen auf die Relevanzsortierung haben sollen, sollten nicht als (nur) Freitext notiert sein, sondern in einer numerischen Schreibweise vorliegen, oder so, dass die direkte Überführung in einen numerischen Wert möglich ist. Also z. B. statt *2. erg. Aufl. 2* oder *2,5*. Wenn die umgangssprachliche Kennzeichnung notwendig ist, kann der numerische Wert in einem separaten Feld ergänzend notiert werden.

Semantische Bedeutungen von Daten und inhaltliche Beziehungen zwischen Daten oder Datenfeldern müssen als solche gekennzeichnet sein – z. B. mittels Semantic-Web-Technologien, also über RDF-Tripel und unter Einbeziehung geeigneter Vokabulare. Ein Beispiel hierfür ist die standardisierte Kennzeichnung von Sprachen anhand einer der Tabellen von ISO 639,¹⁴ die Kennzeichnung, welche Tabelle verwendet wurde und die semantische Relation, in der der Sprachcode zu dem gekennzeichneten Dokument steht (Sprache des Titels, Sprache des Textes, etc.). Auf diese Weise lässt sich die Nachnutzbarkeit der Inhaltserschließung nicht nur für Discovery-Systeme, sondern auch für mögliche Nachfolgetechnologien sicherstellen.

14 International Organization for Standardization: Codes for the representation of names of languages. Part 1: Alpha-2 code (ISO 639-1:2002) 2002 (2002); International Organization for Standardization: Codes for the representation of names of languages. Part 2: Alpha-3 code (ISO 639-2:1998) 1998 (1998); International Organization for Standardization: Codes for the representation of names of languages. Part 3: Alpha-3 code for comprehensive coverage of languages (ISO 639-3:2007) 2007 (2007); International Organization for Standardization: Codes for the representation of names of languages. Part 4: Implementation guidelines and general principles for language coding (ISO 639-4:2010) 2010 (2010); International Organization for Standardization: Codes for the representation of names of languages. Part 5: Alpha-3 code for language families and groups (ISO 639-5:2008) 2008 (2008); International Organization for Standardization: Codes for the representation of names of languages. Part 6: Alpha-4 representation for comprehensive coverage of language variation (ISO 639-6:2009) 2009 (2009).

5 Abstimmungsprozesse gestalten

Im vorangegangenen Abschnitt bin ich auf verschiedene Anforderungen eingegangen, die sich aus der aktuellen Sicht des Betriebs und der Entwicklung von Discovery-Systemen an die Inhaltserschließung ergeben. Dabei handelt es sich aber lediglich um eine Momentaufnahme – sowohl die Weiterentwicklung der Suchmaschinenteknologie als auch ein grundsätzlicher Wechsel der Technologie hin zu einem neuen Paradigma kann viele dieser Anforderungen ändern, ablösen oder neue schaffen. Da die Komplexität sowohl auf technologischer Seite als auch im Bereich der Wissensorganisation mittelfristig nicht abnehmen wird, besteht auch weiterhin die Notwendigkeit, die Bereiche Erschließung und Entwicklung von bibliothekarischen Recherchesystemen inhaltlich und in Bezug auf Arbeitsprozesse aufeinander abzustimmen.

Als Grundlage dieses Austauschs möchte ich eine *zielorientierte* Kommunikation vorschlagen. Für moderne Bibliotheken muss gute Nutzbarkeit eine Selbstverständlichkeit sein, die sich bei webbasierten Systemen in guter Usability niederschlägt. Eine Verständigung auf gemeinsame, nutzungsorientierte Ziele ist daher eine sinnvolle Grundlage des Abstimmungsprozesses.

Welche Probleme für die Gestaltung eines solchen Prozesses gibt es und auf welcher Basis kann dieser dennoch umgesetzt werden?

5.1 Bewusstsein schaffen

Um einen gemeinsamen Abstimmungsprozess zu gestalten, ist es zunächst wichtig, die Notwendigkeit desselben zu kommunizieren. Es ist einfach, Inhaltserschließung ausschließlich anhand von theoretisch-wissensorganisatorischen Prinzipien zu betreiben oder IT-Entwicklung anhand der vorliegenden bibliografischen Daten aufzubauen, ohne die Möglichkeiten derselben auszuloten oder sogar gestaltend auf die Daten einzuwirken. Beides ist aber auf Dauer nicht zielführend, da Gestaltungspotential verschenkt wird. Als erster Schritt empfiehlt es sich daher, sich die Möglichkeiten und die Notwendigkeit von kooperativen Prozessen vor Augen zu führen, auch wenn diese auf den ersten Blick einen erhöhten Zeitaufwand benötigen.

5.2 Ressourcen

Eines der wichtigsten Alltagsprobleme an Bibliotheken ist zweifellos die Knappheit finanzieller oder personeller Ressourcen. Es ist daher notwendig, bei der Einführung neuer Services klar zu definieren, welche Leistung erwartet wird und welche Ressourcen dafür initial oder dauerhaft benötigt werden. Aufgrund der in diesem Beitrag dargestellten Verschränkung von Discovery-Systemen und Inhaltserschließung ist es nötig, Discovery-Projekte nicht als reine IT-Projekte zu verstehen, sondern die Bedeutung der Inhaltserschließung und die Aufwände für die notwendige Kommunikation zwischen den Expert:innen beider Bereiche deutlich zu machen.¹⁵

Interessanterweise wurde in beiden Bereichen in den vergangenen Jahren versucht, Synergieeffekte zur Reduzierung der Aufwände nutzbar zu machen. So wird schon seit Jahren innerhalb von Verbänden und auch über Verbundgrenzen hinaus kooperativ inhaltlich erschlossen. Die Entwicklung von Discovery-Systemen wird zentral von privatwirtschaftlichen Anbietern durchgeführt oder in Entwicklungsverbänden auf Basis von Open Source-Systemen wie z. B. VuFind, Lukida oder beluga core betrieben. Durch kooperative Strukturen oder Projekte wie finc oder das beluga-core-Konsortium lassen sich ebenfalls Entwicklungs- und Einrichtungsaufwände minimieren.

5.3 Gestaltung der Kommunikation

Selbst wenn ausreichend Ressourcen für den Abstimmungsprozess vorhanden sind und das Bewusstsein der Notwendigkeit desselben existiert, kann dieser dennoch fehlschlagen. Ein Grund hierfür kann sein, dass kein ausreichender expliziter oder zumindest impliziter inhaltlicher Konsens zwischen Erschließung und Produktentwicklung der Suchsysteme herrscht. Diese Probleme sind oft das Ergebnis separater professioneller Diskurse in der Inhaltserschließung und in der Entwicklung von Discovery-Systemen, oder werden durch diese zumindest verstärkt. Discovery-Systeme werden zumeist in IT-Abteilungen von Bibliotheken bzw. Verbänden oder von privatwirtschaftlichen Dienstleistern entwickelt. Die Inhaltserschließung findet dagegen in vielen Einrichtungen in anderen Abteilungen und Organisationseinheiten – zum Beispiel der Erwer-

¹⁵ Natürlich sind Discovery-Systeme auch mit zahlreichen anderen Arbeitsbereichen verknüpft, für die ebenfalls Ressourcen bereitgestellt werden müssen. In diesem Beitrag liegt der Fokus aber auf den Bereichen Inhaltserschließung und IT-Entwicklung von Discovery-Systemen.

bungsabteilung oder im Fachreferat – statt. Aus diesem Grund ist es nötig, zur Entwicklung von Discovery-Systemen in einen anhaltenden gemeinsamen Kommunikationsprozess zu treten. Dabei ist es zum einen wichtig, dass bei der Entwicklung von Discovery-Systemen die Besonderheiten von Inhaltserschließungssystemen berücksichtigt werden, wie weiter oben z.B. im Abschnitt *Darstellung und Vernetzung* skizziert wurde. Zum anderen müssen aber auch bei der Gestaltung von Inhaltserschließungssystemen die Möglichkeiten von Discovery-Systemen oder nachfolgenden Technologien berücksichtigt werden. Dies ist keine neue Idee – so war z. B. die Permutation von Schlagwortketten ein Mechanismus, der speziell für die Technologie der Zettelkataloge eingeführt wurde¹⁶ und der mittlerweile weitestgehend seinen Zweck verloren hat.

Die Etablierung eines professionellen Diskurses zwischen Inhaltserschließung und IT-Entwicklung bei Discovery-Projekten ist leider nur eine notwendige, aber keine hinreichende Bedingung für den Erfolg. Es ist erforderlich, diesen aufgabenorientiert zu gestalten, so dass die Lösung konkreter Probleme angegangen und von den fachgebietsspezifischen Diskursen abstrahiert werden kann. Kurz gesagt: Es muss Konsens geschaffen werden über die zu erreichenden operativen Ziele und Anforderungen. Um dies zu erreichen möchte ich die Einführung eines gemeinsamen Anforderungsmanagements vorschlagen. Mangelndes Anforderungsmanagement ist die größte Gefahr für IT-Projekte im Allgemeinen,¹⁷ daher ist auch für Discovery-Projekte ein solches sehr empfehlenswert. Die Ausweitung des Anforderungsmanagements auch auf den Diskurs Erschließung/IT-Entwicklung ist naheliegend.

6 Gemeinsames Anforderungsmanagement als Erfolgsfaktor

Wie kann ein Anforderungsmanagement, das Inhaltserschließung und die Entwicklung von Discovery-Systemen umfasst, gestaltet werden? Zunächst ist es sinnvoll, sich die klassischen Aufgabenbereiche des Anforderungsmanagements vor Augen zu führen und zu überprüfen, ob und wie diese berücksichtigt werden können. Diese sind¹⁸:

16 Gödert, Winfried: Verbale Sacherschließung und Probleme ihrer Koordination. In: Bibliothek. Forschung und Praxis 12 (1988). S. 325–336, S. 325.

17 The Standish Group: CHAOS Report 2015.

18 Grande, Marcus: 100 Minuten für Anforderungsmanagement. kompaktes Wissen nicht nur für Projektleiter und Entwickler. 2. Aufl. Wiesbaden: Springer Vieweg 2014, S. 7.

1. Ermitteln und Finden von Anforderungen
2. Dokumentieren von Anforderungen
3. Prüfen und Abstimmen von Anforderungen
4. Validieren von Anforderungen
5. Pflegen und Verwalten von Anforderungen
6. Pflegen der Anforderungsdokumente/Anforderungsdatenbasis
7. Beurteilen von Risiken
8. Arbeiten mit Werkzeugen für das Anforderungsmanagement

Die weiter oben geschilderte Notwendigkeit, eine gemeinsame kommunikative Basis zu gestalten, findet sich in dieser Auflistung höchstens mittelbar wieder. Dies mag darin begründet sein, dass klassisches Anforderungsmanagement auf personell und zeitlich klar abgegrenzte Projekte bezogen ist und den Fokus nicht auf teilweise über Jahre hinweg laufende Abstimmungen voneinander getrennter Arbeitsbereiche setzt. Es ist daher sinnvoll und nötig, diese klassischen Aufgaben zu erfüllen, aber darüber hinaus wie folgt zu erweitern:

1. Es müssen übergeordnete (strategische) Ziele definiert werden. Anhand dieser kann der Nutzen einzelner Anforderungen begründet und diese damit legitimiert werden.
2. Es müssen Werkzeuge und Techniken zum Einsatz kommen, die eine gemeinsame kommunikative Basis schaffen, anhand derer die einzelnen Anforderungen geprüft und diskutiert werden können.

Ein übergeordnetes Ziel muss die Nutzungsfreundlichkeit sein – diese ist ein erklärtes Interesse von Discovery-Systemen und sollte auch ein generelles Anliegen bei der Entwicklung bibliothekarischer Dienste für die Öffentlichkeit sein, unabhängig davon, ob diese auf der Basis von webbasierten Informationssystemen erfolgen oder nicht. Anforderungen, die kontrovers zwischen Erschließung und Entwicklung diskutiert werden, könnten auf Basis der Frage *Schafft die Umsetzung dieser Anforderung einen Mehrwert für die Usability?* bewertet werden. Natürlich sind auch andere übergeordnete Ziele denkbar und müssten ausgearbeitet und beschlossen werden.

Die Wahl der Werkzeuge ist nicht ganz einfach, allerdings wurden in den vergangenen Jahren zahlreiche vielversprechende Ansätze entwickelt. Ich möchte an dieser Stelle beispielhaft zwei gut miteinander kombinierbare Techniken vorstellen, die in den letzten Jahren in verschiedenen bibliothekarischen IT-Projekten erfolgreich eingesetzt wurden:¹⁹ *Personas* und *Scenario Based*

¹⁹ Vgl. z. B. Wendt, Kerstin u. Matthias Finck: *Scenario-based Design als Vorgehensmodell für Softwareentwicklung in Bibliotheken*. Berlin 2018 (108. Deutscher Bibliothekartag); Siegfried,

Design. Beide Techniken erlauben es, Interessen von Nutzer:innen festzustellen und die Erfüllung derselben sowohl abstrakt als auch in Bezug auf reale Informationssysteme zu diskutieren.

6.1 Personas

Der Einsatz der Persona-Technik speziell in Bibliotheken wird z. B. von Viveca Nyström detailliert beschrieben.²⁰ Diese Technik kann beim Treffen strategischer Entscheidungen, für das Marketing aber auch bei der Entwicklung webbasierter Systeme von großem Nutzen sein.

Personas sind Charakterisierungen fiktiver Personen, die Nutzer:innen des jeweiligen Service oder Serviceangebotes sind. Personas können sehr detailliert ausgearbeitet oder nur grob als sogenannte Proto-Personas skizziert werden. Wichtige, zu definierende Eigenschaften der Proto-Personas sind beispielsweise der Name, das Aussehen (in Form eines Profilbildes), ein oder mehrere typische Aussagen oder Zitate, die grundsätzliche Motivation der Persona und spezifische Aspekte der Persönlichkeit.

Personas repräsentieren Nutzungsgruppen und sollten daher anhand von empirischen Daten erstellt werden, zum Beispiel auf Basis von Umfragen oder Nutzungsgruppenanalysen. Kontextbezogene Interviews von Nutzer:innen zu konkreten Tätigkeiten sind sehr hilfreich, um die Motivation zu verstehen.

In der Arbeit mit Personas kann man Entscheidungen treffen, die die Personas als konkrete Beispiele heranziehen. So kann z. B. die Frage, ob und wie die RVK im Discovery-System dargestellt werden soll, daran geprüft werden, ob die Persona *Jörg Müller*, 42, Industriedesigner, aus der vorgeschlagenen Darstellung einen Nutzen zieht oder sich sogar daran stört.

Personas helfen, die Anliegen verschiedener Nutzungsgruppen im Gedächtnis zu behalten. Die ausschließliche Optimierung von Services für kleine oder ggf. nicht bedeutende Nutzungsgruppen wird so verhindert. Sie bieten eine sehr gute Diskussionsgrundlage sowohl bei strategischen Entscheidungen als auch für konkrete Fragestellungen. Weiterhin sind sie hervorragend mit Szenarien und dem Verfahren des Scenario Based Design kombinierbar.

Doreen: Mit Personas Empathie für Bibliotheksnutzer/innen schaffen. Frankfurt a. M. 2017 (106. Deutscher Bibliothekartag); Siegfried, Doreen: Personas in der ZBW. Frankfurt a. M. 2017 (106. Deutscher Bibliothekartag).

²⁰ Nyström, Viveca: *An evaluation of the benefits and value of libraries*. Oxford, UK: Chandos Pub 2012 (Chandos information professional series). S. 99 ff.

6.2 Scenario Based Design

Scenario Based Design (SBD) wurde initial von Rosson und Carroll beschrieben.²¹ Es ist ein Prozessablauf zur Entwicklung von interaktiven Produkten. Solche Produkte können webbasierte Informationssysteme sein, aber auch z. B. Informationsservices einer Bibliothek. Das Ziel der Methode ist es, die Produkte über das tiefgehende Verständnis von Aktivitäten der Nutzenden schrittweise und gezielt zu verbessern (Transformation).

In der Anwendung von SBD notiert man im Gegensatz zu vergleichbaren Methoden keine abstrakten Beschreibungen von Anforderungen, sondern erstellt sogenannte Szenarien. Szenarien beschreiben erzählerisch die konkrete Nutzung eines Produktes aus der Perspektive eines Nutzenden.

Zunächst werden anhand eines Grundkonzeptes und empirischer Daten Problemszenarien entwickelt, die den Status quo beschreiben. Im Rahmen einer Claims Analysis werden die wesentlichen Aspekte erfasst, die positive oder negative Auswirkungen auf die User Experience des Produktes haben.

Im nächsten Schritt wird iterativ anhand verschiedener Szenarientypen die Vision entwickelt – also eine genaue Vorstellung davon, wie das zukünftige, optimierte System aussehen soll. Folgende Szenarien werden dazu modelliert:

1. Das *Activity Design*: In diesem Schritt werden die gewünschten Funktionen des Systems entwickelt. Dabei wird aufgrund der Komplexität von User Interfaces zunächst von Aspekten der Bedienoberfläche des Systems abstrahiert – es geht an dieser Stelle darum zu erfassen, welche Aktivitätsszenarien für die Nutzer:innen befriedigend sind und welche Funktionen das System dafür bereitstellen muss.
2. Das *Information Design*: Hierfür werden Szenarien erstellt, die die für die Interaktion bereitgestellten Informationen darstellen. Dabei kann es sich unter anderem um konkrete Boxen und Objekte der Bedienoberfläche handeln, um Texte, Bedienelemente und die mit ihnen verbundenen Aufgaben.
3. Das *Interaction Design*: Hier werden Interaktionsszenarien beschrieben, um sicherzustellen, dass Interaktionsverläufe flüssig und erfolgreich ablaufen können.

Alle drei Schritte werden durch eine *Claims Analysis* ergänzt und erweitert. Jeder der Schritte erfolgt iterativ, auch die jeweilige *Claims Analysis* kann zu einer Anpassung führen.

²¹ Rosson, Mary Beth u. John M. Carroll: Usability engineering. Scenario-based development of human-computer interaction. 1. Aufl. San Francisco: Academic Press 2010 (Morgan Kaufmann series in interactive technologies).

Abschließend können auf der erarbeiteten Grundlage Prototypen entworfen werden, die durch Usability-Tests überprüft werden. Je nach Projektsituation können die Prototypen schließlich in den Produktivbetrieb übernommen oder erneut geprüft werden.

6.3 Anwendbarkeit der Techniken

Personas lassen sich aufgrund ihrer hohen Anschaulichkeit relativ einfach in einen Kommunikations- und Abstimmungsprozess aufnehmen. Somit erfüllen sie die oben definierte Anforderung, ein Werkzeug zu sein, auf dessen Grundlage man anschaulich das Erfüllen von Anforderungen diskutieren kann.

Aufgrund seiner Komplexität und des strukturierten Ablaufs ist Scenario Based Design etwas schwieriger in Abstimmungsprozesse aufzunehmen. Speziell bei der Entwicklung von Activity Designs und Information Designs können kooperativ Szenarien entwickelt werden, die das Fachwissen von Erschließung und IT-Entwicklung einbinden und in anschaulichen Beschreibungen (Szenarien) erfassen. Wie Abstimmungsprozesse im Detail zu gestalten sind, muss anhand der vorliegenden Situation und Organisationsform erfasst werden. Die beiden Techniken liefern aber eine spannende Grundlage für ein gemeinsames Anforderungsmanagement. Ihre leichte Erlernbarkeit und Anwendbarkeit im Kontext von Projekten an und mit Bibliotheken haben sie bereits unter Beweis gestellt, so dass sie einen guten Ausgangspunkt darstellen.

7 Fazit und Ausblick

Ich habe in diesem Beitrag argumentiert, dass Inhaltserschließung für die Entwicklung von Discovery-Systemen von großem Nutzen ist. Es besteht die Gefahr, dass beide Prozesse ohne die Verständigung auf gemeinsame Ziele nicht produktiv miteinander interagieren. Aus diesem Grund ist es sinnvoll, zunächst ein gemeinsames Bewusstsein dafür zu schaffen, dass eine Abstimmung Synergieeffekte freisetzt und eine notwendige Voraussetzung für das Verfügbarmachen von nutzungszentrierten Services darstellt.

Als nächstes müssen übergeordnete Ziele definiert werden, wobei das Ziel, Produkte und Services mit hoher Nutzungsfreundlichkeit anzubieten, eine zentrale Rolle spielen sollte. Um im Rahmen eines gemeinsamen Anforderungsmanagements Ziele aus Sicht der Nutzenden zu definieren und diese als Leitlinien

für die Arbeit beider Geschäftsprozesse zu verwenden, können Techniken wie Personas oder Scenario Based Design dabei von großem Nutzen sein.

Die Qualität der Inhaltserschließung – gemessen an Homogenität, semantischer Vernetzung, Abdeckung, Qualität des Datenaustauschs etc. – wirkt sich dabei direkt auf die Nutzbarkeit derselben in Discovery-Systemen aus. Es ist trivial anzunehmen, dass dies auch für die Nutzung der Inhaltserschließung durch zukünftige Technologien gilt. Dies ist wichtig, da sich abzeichnet, dass Discovery-Systeme in der heutigen Form als vorherrschende Technologie für bibliothekarische Recherchesysteme abgelöst oder zumindest stark weiterentwickelt werden. Auch die Verwendung von alten Datenformaten wie MARC 21 oder MARC-XML sollte in absehbarer Zeit eingestellt werden – der Satz *MARC must die* von Roy Tennant ist trotz des Alters von bald 20 Jahren so wichtig wie nie zuvor. Mögliche Nachfolger sind Formate und Systeme auf der Basis von Semantic-Web-Datenmodellen wie BIBFRAME oder dem Open Research Knowledge Graph. Durch die Verwendung neuer Technologien und Datenmodelle entstehen neue Berührungspunkte von Inhaltserschließung und Informationssystemen, die die eingangs genannten Punkte ersetzen oder ergänzen. Eine Ausrichtung von Inhaltserschließung und der Entwicklung von Informationssystemen in Hinblick auf gemeinsame Ziele und die Optimierung der Kommunikation in beiden Prozessen wird daher auf unbestimmte Zeit eine wichtige Aufgabe sein.

8 Literaturverzeichnis

- Apache Software Foundation: Apache Solr Reference Guide. MoreLikeThis. Version 8.4. https://lucene.apache.org/solr/guide/8_4/morelikethis.html (29.5.2020).
- Blenke, Martin: Inhaltserschließung als Navigationspunkte für den Suchprozess. eine Bibliothek gestaltet ihr Discovery System selbst... Bremen 2013 (Fachreferententagung Wirtschaftswissenschaft). https://m.suub.uni-bremen.de/uploads/cms/files/Martin_Blenke_Inhaltserschliessung_als_Navigationspunkte_fuer_den_Suchprozess_Sep_2012.pdf (20.4.2020).
- Gödert, Winfried: Verbale Sacherschließung und Probleme ihrer Koordination. In: Bibliothek. Forschung und Praxis (1988) Bd. 12 H. 3. S. 325–336. <https://doi.org/10.1515/bfup.1988.12.3.325>.
- Grande, Marcus: 100 Minuten für Anforderungsmanagement. kompaktes Wissen nicht nur für Projektleiter und Entwickler. 2. Aufl. Wiesbaden: Springer Vieweg 2014.
- Hacker, Rupert: Bibliothekarisches Grundwissen. 7. Aufl. München: Saur 2000.
- International Organization for Standardization: Codes for the representation of names of languages. Part 2: Alpha-3 code (ISO 639-2:1998) 1998 (1998).
- International Organization for Standardization: Codes for the representation of names of languages. Part 1: Alpha-2 code (ISO 639-1:2002) 2002 (2002).

- International Organization for Standardization: Codes for the representation of names of languages. Part 3: Alpha-3 code for comprehensive coverage of languages (ISO 639-3:2007) 2007 (2007).
- International Organization for Standardization: Codes for the representation of names of languages. Part 5: Alpha-3 code for language families and groups (ISO 639-5:2008) 2008 (2008).
- International Organization for Standardization: Codes for the representation of names of languages. Part 6: Alpha-4 representation for comprehensive coverage of language variation (ISO 639-6:2009) 2009 (2009).
- International Organization for Standardization: Codes for the representation of names of languages. Part 4: Implementation guidelines and general principles for language coding (ISO 639-4:2010) 2010 (2010).
- Koster, Lukas: Discovery tools: a rearguard action? Palma de Malloca 2012 (ELAG2012). <https://www.slideshare.net/lukask/discovery-tools-a-rearguard-action> (4.9.2020).
- Lewandowski, Dirk: Web Information Retrieval. Technologien zur Informationssuche im Internet. Frankfurt a. M.: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e. V. 2005 (=DGI-Schrift (Informationswissenschaft) 7).
- Nyström, Viveca: An evaluation of the benefits and value of libraries. Oxford, UK: Chandos Pub 2012.
- Rosson, Mary Beth und John M. Carroll: Usability engineering. Scenario-based development of human-computer interaction. 1. Aufl. San Francisco: Academic Press 2010.
- Siegfried, Doreen: Mit Personas Empathie für Bibliotheksnutzer/innen schaffen. Frankfurt a. M. 2017 (106. Deutscher Bibliothekartag). Abstract [Folien unveröffentlicht]. <https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/year/2017/docId/2855> (6.1.2021).
- Siegfried, Doreen: Personas in der ZBW. Frankfurt a. M. 2017 (106. Deutscher Bibliothekartag). <https://www.zbw.eu/fileadmin/pdf/veranstaltungen/2017-bibtag-siegfried-personas.pdf> (6.1.2021).
- Steilen, Gerald: Discovery-Systeme – die OPACs der Zukunft? Hamburg 2012 (101. Deutscher Bibliothekartag). <http://www.opus-bayern.de/bib-info/volltexte/2012/1188> (20.4.2020).
- Stumpf, Gerhard: „Kerngeschäft“ Sacherschließung in neuer Sicht. Was gezielte intellektuelle Arbeit und maschinelle Verfahren gemeinsam bewirken können. Augsburg: Universitätsbibliothek 2015. <https://nbn-resolving.org/urn:nbn:de:bvb:384-opus4-30027> (31.12.2020).
- The Standish Group: CHAOS Report 2015. https://www.standishgroup.com/sample_research_files/CHAOSReport2015-Final.pdf (31.12.2020).
- Wendt, Kerstin u. Matthias Finck: Scenario-based Design als Vorgehensmodell für Softwareentwicklung in Bibliotheken. Berlin 2018 (108. Deutscher Bibliothekartag). <https://nbn-resolving.org/urn:nbn:de:0290-opus4-35267> (29.5.2020).
- Wiesenmüller, Heidrun: Resource Discovery Systeme – Chance oder Verhängnis für die bibliothekarische Erschließung? Hildesheim 2012 (Jahrestagung der Deutschen Gesellschaft für Klassifikation). <https://doi.org/10.5445/IR/1000029081>.