

Gerhard Stumpf

„Kerngeschäft“ Sacherschließung in neuer Sicht

Was gezielte intellektuelle Arbeit und maschinelle Verfahren gemeinsam bewirken können

Einleitung

Dieser Aufsatz¹ ist der Sacherschließung im Allgemeinen gewidmet und soll ein Beitrag zu den laufenden Überlegungen zu ihrer Neuausrichtung im deutschsprachigen Raum sein. Ein wichtiger Aspekt ist dabei die Abstimmung von intellektueller und automatischer Erschließung. Hier soll deshalb versucht werden, die Perspektiven der intellektuellen und der maschinellen Verfahren im Allgemeinen zu skizzieren und in einen adäquaten Zusammenhang zu stellen.

Die Sach- oder Inhalterschließung gilt traditionell als „Kerngeschäft“ insbesondere der wissenschaftlichen Bibliotheken. Ein Kerngeschäft ist ein Arbeitsbereich, der konstitutiv für das Selbstverständnis eines Betriebs / einer Branche / eines Berufes ist, woraus diese einen Großteil ihrer Existenzberechtigung ableiten. In diesem Sinn, so scheint es, hat die Sacherschließung im Umfeld all dessen, womit Bibliotheken heute nützlich sein wollen, von ihrem früher beachtlichen Stellenwert einiges eingebüßt. Man braucht dazu nur die diversen neueren Beiträge zur Berufsbild-Debatte und zum Fachreferentensystem zu lesen.² Aber verbirgt sich dahinter vielleicht gar kein dauerhafter Relevanzverlust, sondern nur die Notwendigkeit einer Neujustierung der Ziele, Verfahren und Akteure, die der Sacherschließung langfristig neue Wertschätzung bringen kann?

Wenn wir uns im deutschsprachigen Raum in nächster Zeit stärker internationale, d. h. im Wesentlichen von der amerikanischen Tradition geprägte Prinzipien zu eigen machen, müssen wir uns bewusst sein, dass die Sacherschließung dort nicht eine jederzeit hinterfragbare Zusatzleistung, sondern integraler Teil des bibliothekarischen Kerngeschäfts der Erschließung ist. Mindestens eine verbale oder klassifikatorische Inhalterschließung wird auch nach RDA an jeder Titelaufnahme erwartet. Das entspricht auch der aktuellen Praxis der Library of Congress und anderer Nationalbibliotheken: Inhalterschließung breit und umfassend angelegt, aber in einem abgestuften System, wobei nur ein Kernbereich sowohl klassifikatorisch als auch verbal erschlossen wird – und anderswo natürlich noch stärker auf Kompetenz- und Datenlieferzentren konzentriert als es in der deutschen Bibliothekslandschaft zur Zeit der Fall ist.

Bezieht man die *user tasks* der FRBR (find – identify – select)³ auf die inhaltliche Suche, so umschreiben sie ein durchaus anspruchsvolles Aufgabenspektrum. Dass Sacherschließung notwendig ist, wird auch hierzulande kaum bestritten. In den letzten Jahrzehnten haben sich aber Entwicklungen, Fortschritte und veränderte Rahmenbedingungen ergeben, die zusammen dafür verantwortlich sind, dass ihre Form und Organisation insbesondere in den Bibliotheken im deutschsprachigen Raum nicht erst heute in Frage steht und ja bereits vielfache Anpassungen erfahren hat.

¹ Es handelt sich um den leicht überarbeiteten Text eines Vortrags bei der VDB-Fortbildungsveranstaltung "Wandel als Konstante: neue Aufgaben und Herausforderungen für sozialwissenschaftliche Bibliotheken" am 22./23. Januar 2015 in Berlin.

² Vgl. z. B. Bonte, Achim: Der wissenschaftliche Dienst in der Digitalen Bibliothek. Was kommt nach dem Fachreferentensystem? (Vortrag, 2014). http://www.slideshare.net/Achim_Bonte/der-wissenschaftliche-dienst-in-der-digitalen-bibliothek-was-kommt-nach-dem-fachreferentensystem-39650127

³ Tillett, Barbara: What is FRBR? A Conceptual Model for the Bibliographic Universe. 2004, p. 5. www.loc.gov/cds/downloads/FRBR.pdf

- Bibliotheken haben immer mehr Aufgaben, aber weniger Personalressourcen. Wirtschaftliche Zwänge lassen vor allem zu personalintensiven Arbeiten Alternativen suchen.
- Es gibt neue Retrievalsysteme, vor allem die Suchmaschinentechologie, mit Zusatzfeatures, die die Suchprozesse nach verschiedenen Kriterien unterstützen. Die bibliothekarische Sacherschließung verliert in diesem Kontext anscheinend an Bedeutung, zumindest an Exklusivität. Auch die Benutzererwartungen verändern sich; statt gründlicher Literaturrecherche gibt man sich gern mit dem leicht Erreichbaren zufrieden; die Bereitschaft, sich in spezifische Erschließungssysteme hineinzudenken, schwindet.
- Maschinelle Verfahren, die Erschließungsdaten generieren bzw. aufbereiten, treten konkurrierend neben die intellektuell basierte Erschließung bzw. werden als Alternativen in Betracht gezogen, weil sie qualitativ aufgeholt haben. Dazu gehört auch das Catalog Enrichment. Allerdings nährt dies auch die überzogene Erwartung, alles ließe sich automatisieren.
- Mehr Fremddaten aus heterogenen Quellen werden rascher verfügbar und die Fremddatenversorgung unübersichtlicher. Die Eigenleistung verliert weiter an Bedeutung; so wächst das Gefühl, die Sacherschließung werde ohnehin anderswo gemacht und „ready for use“ frei Haus geliefert.
- Linked open data machen die im Netz vorhandene Information sukzessive besser verfügbar; auch Bibliothekskataloge öffnen sich ins Semantic Web, Damit sind inhaltliche Informationen über Dokumente oft mit Internetsuchmaschinen bequemer zu finden. So werden für manche Fragestellungen Schlagwörter und Klassifikationen entbehrlich. Verfügbarkeitsinformationen sind oft unabhängig von einer gründlichen, proprietären Katalogsuche und Relevanzbeurteilung erhältlich.

Dagegen steht unverkennbar eine positive Entwicklung: Die Reichweite und letztlich die Effizienz jeder Art von Sacherschließung ist viel größer geworden. Metadaten in Web-Dimension erhalten weltweite Relevanz.

Machen wir uns noch einmal bewusst, was die Inhaltserschließung durch Menschen bedeutet: Die intellektuelle Inhaltsanalyse versucht die wirkliche Thematik eines Werkes verstehend zu erfassen, unabhängig von den Formulierungen des Titels o.ä. Sie umfasst eine Relevanzbeurteilung auf Grund fachlichen Vorwissens, und sie umfasst in den meisten Fällen eine Autopsie, also das kursorische Lesen im Text einer Publikation. Außerdem ist die intellektuelle Erschließung fähig, ihre Ergebnisse zielgerichtet auf Menschenlesbarkeit und –verstehbarkeit hin zu optimieren.

Menschenlesbarkeit und Verständlichkeit werden oft ausgeblendet, sind aber nicht wegzudiskutierende Anforderungen an jede Erschließung, die sich an Menschen richtet. Andererseits kann ein umfassendes Erschließungssystem, das Netzressourcen dezidiert einbezieht, nicht mehr ohne automatische Module und Techniken auskommen. Ein Erschließungssystem jedoch, das Netzressourcen außen vor lässt, ist heute verloren und nutzlos.

Deshalb meine Überzeugung: **Auch wenn die intellektuelle Sacherschließung als bibliothekarisches Kerngeschäft in Frage steht, bleibt sie für den Kern der Metadatenwelt unverzichtbar.** Obwohl Sacherschließung überall eher zurückgefahren wird, steigen die gewonnenen Mehrwerte. Das hat auch mit den Fortschritten bei der Datenaufbereitung und bei maschinellen und automatischen Verfahren zu tun. Maschinelle Verfahren sind also kein Ersatz für Sacherschließung oder ihr Tod, sondern tragen zu ihrer Wirkung bei, weil sie die quantitativen Defizite ausgleichen, die bei der intellektuellen Erschließung bleiben und ihr schaden.

Funktionen der Sacherschließung⁴

Sacherschließung dient im Allgemeinen (funktional betrachtet, also nach den Möglichkeiten in bibliothekarischen Ordnungssystemen und Katalogen)

1. zum primären Auffinden bei der Suche nach inhaltlichen Kriterien. Die üblichen Formen sind die sequenzielle Aufstellung (im Regal), eine Auflistung (Register) oder ein maschinell erstellter Index für Retrievalsysteme.
2. als deskriptive Aussage über eine Thematik, zur Einschätzung der inhaltlichen Relevanz einer Ressource an Hand einer angezeigten Dokumentbeschreibung,
3. zum sekundären Auffinden im Suchprozess („weitere Literatur zu diesem Thema“). Im Online-Katalog wird ein angezeigtes Datenelement zum neuen Suchterm; dieser Suchstrategie entspricht im Web ein semantisch definierter Hyperlink.
4. zur Selektion der inhaltlich für die Fragestellung relevantesten Ressourcen aus einer größeren Treffermenge (Auswahl eines inhaltlichen Aspekts, Facetten).

The screenshot shows a library search interface with the following elements:

- Search Query:** Ihre Suchanfrage Thema/Schlagwort = logische partikel
- Ausgewählte Suchfilter:** Themen/Schlagwörter (grob) with **Deutsch** selected.
- Treffer UB Augsburg (12):** 1. **Konditionalität und Kau:** Volodina, Anna, 2011.
- Treffer eingrenzen:**
 - Fach: Germanistik / N... (9), Sprach-/Literat... (6), Slavistik (1)
 - Themen/Schlagwörter (grob): Logische Partikel (12), Aufsatzsammlung (3), Kausalsatz (2), Syntax (2), **Textkohärenz (-)**
 - Themen/Schlagwörter (präzise): Adversativität ... (1), Adversativsatz ... (1), Deutsch / Adver... (1), Deutsch / Appos... (1), Deutsch / Aussa... (1)
 - Erscheinungsjahre: >= 2006 (3), 2002 - 2005 (3), 1994 - 2001 (3)
- UB Augsburg (9/12):**
 - Substitutionsprozesse in Junktionen mit "dass" oder "ob"**
 - Antworten auf Fragen zum Platzhalterproblem
 - VERFASSER: Bischoff, Anneliese
 - ORT, VERLAG, JAHR: Bielefeld, Bischoff, 1996
 - UMFANG: IV, 309 S. : graph. Darst.
 - SIGNATUR: 64/GC 7246 B621
 - SCHLAGWORT: Deutsch, Fachsprache, Komplexer Satz, Logische Partikel, dass, ob <Wort>
 - SCHLAGWORT 2: Deutsch, Komplexer Satz, Konjunktion, Korrelat
 - ISBN / ISSN: 3-925167-03-X
 - [in die Merklste](#)
- Exemplare:**

| zum Dokument | mehr zum Titel |
|-----------------------------|------------------------------------|
| Signatur Fußnoten | Standort Zentralmagazin |
| Zweigstelle UB | Status <u>bestellbar</u> |
| 01/GC 7246 B621 | |
- 6. Handbuch der deutsche** 2003, 64/GC 7117 P279

⁴ Ich beziehe mich vorwiegend auf konventionelle Bibliotheksmaterialien, also Publikationen in gedruckter und digitaler Form; die Überlegungen zur bibliothekarischen Sacherschließung sollten aber auch für Webressourcen aller Art, also z. B. Webseiten, zutreffen.

Nach einem anderen, perspektivisch vielleicht angemesseneren Schema lassen sich drei Grundleistungen unterscheiden.

| Suchbarkeit von Inhalten | Beschreibung von Inhalten | Verlinkung zwischen Themen, Entitäten und Ressourcen |
|---|--|---|
| über hinterlegte Indizes, maschinenlesbar | menschenlesbar und -verständlich | im Web, über Identifikatoren, Datenmodelle wie RDF |
| Metadaten: <ul style="list-style-type: none"> • Notationen und Benennungen (Klass.) • Bevorzugte und abweichende Namen • Codes | Metadaten: <ul style="list-style-type: none"> • Benennungen (Klass.) • Bevorzugte Namen • Schlagwortfolgen • Abstracts | Normdaten GND-interne Beziehungen GND – Wikipedia GND – VIAF usw. |
| Teile von Paratext und Volltext | Paratext Catalogue Enrichment | |

Bezogen auf die Entitäten der Gruppe 1 der FRBR könnte man Sacherschließungsdaten annähernd wie folgt strukturieren:

| Werk | Expression | Manifestation („Dokument“) | Exemplar |
|---|--|---|---|
| Inhaltsanalyse (erfolgt an Hand von Manifestationen) | | Paratext (von Autor / Hrsg. / Verlag) <ul style="list-style-type: none"> • Inhaltsverzeichnis • Klappentext • Register • Abstract | |
| <ul style="list-style-type: none"> • Zuweisung von Schlagwörtern für den Inhalt (Themen) • Klassifizieren des Inhalts | Formschlagwörter für die Art des Inhalts Klassifizieren mit formal definierten Klassen (z. B. RVK-Notationen LP-LY) | Formschlagwörter für den Datenträger oder Materialtyp | <i>Systematische Aufstellung gemäß Klassifikation</i> |
| | | Klassische Kataloganreicherung | |
| | | Volltext-Extraktion | <i>Provenienzerschließung (auch mit Normdaten)</i> |

Potenziale und Grenzen maschineller Erschließungsverfahren

Intellektuelle Sacherschließung gilt prinzipiell dem **Werk**. Sie kann allerdings, so lange vorliegende Manifestationen katalogisiert werden, nur stellvertretend an diesen verankert werden. Die für ein Werk einmal geleistete Inhalterschließung kann bei nahezu allen seinen Manifestationen in identischer Form maschinell angereichert, d. h. auf diese kopiert werden. Damit lässt sich für die Ausprägungen eines Werks und dessen Inhalt eine konsistente Erschließung sicherstellen.

Die automatische verbale Erschließung und ebenso die klassische Kataloganreicherung basieren in der Regel auf der **Manifestation**, besonders auf deren Paratext⁵ bzw. deren inhaltsauszeichnender Struktur und werden für unterschiedliche Manifestationen eines Werkes ggf. wiederholt.

Nun können textliche Unterschiede verschiedener Manifestationen, wenn daraus maschinell Metadaten extrahiert werden, leichter zu abweichender Inhaltsbeschreibung trotz gleichem Inhalt führen. Beispielsweise können für die Neuausgabe eines Werkes das Vorwort und der Klappentext ganz neu verfasst sein und anderes Vokabular enthalten. Ein Haupthindernis für werkbezogen konsistente Inhalterschließung ist jedoch vor allem die wechselnde Sprache der Manifestationen.

Wenn aber für jede Manifestation das zugehörige Werk ermittelt werden kann, dann könnten auch die originär maschinell erzeugten Sacherschließungsdaten auf andere Manifestationen dieses Werkes übertragen werden, wenn diese nicht bereits intellektuell hinreichend erschlossen sind.

Verschiedene Arten der Sacherschließung sind in unterschiedlicher Form automatisierbar. Die verbalen **automatischen Extraktionsverfahren** suchen und erzeugen Wörter, die auf Grund ihrer Häufigkeit und Position im Text und ihrer Übereinstimmung mit Wörterbucheinträgen und Normdaten als bevorzugte Benennungen indexiert und angezeigt werden sollen. Den Schritt von Wörtern, also Zeichenketten, zu Konzepten und Themen, den das menschliche Gehirn vollzieht, beherrscht die Maschine nicht bzw. sie ersetzt sie durch das Verfolgen von Wahrscheinlichkeiten. Ebenso fehlt den automatischen Prozessen in der Regel der Blick aufs Ganze als Korrektiv für Aktionen im Einzelnen. Die Lernfähigkeit eines automatischen Systems ist nicht vergleichbar mit dem Wissen, das ein menschlicher Erschließer durch seine Arbeit erwirbt. Im Idealfall erzielt die automatische verbale Erschließung anscheinend gleiche Ergebnisse wie eine intellektuelle Schlagwortvergabe, wenn sich beide am gleichen Vokabular, z. B. der GND, ausrichten. Von den Wörtern auf Themen zu schließen, wird aber – ohne Einschaltung des erschließenden Verstehens – ganz dem Metadaten-Nutzer überlassen.

Die **intellektuelle Inhaltsanalyse** geht ebenso von verbalen Elementen des Dokuments aus, kann aber in anderer Weise Konzepte und Themen erkennen, das Gemeinte interpretieren und die vorkommenden Einzelwörter auf Grund kontextueller Informationen gewichten. Nur die intellektuelle Erschließung kann nutzer- oder anwendungsbezogene Modifikationen vornehmen, z. B. auch eine der menschlichen Aufnahmefähigkeit gemäße Beschränkung und Redundanzvermeidung. Die intellektuelle Textanalyse kann auch diskursive und syntaktische Markierungen deuten. Wenn ein Autor seine Thematik beschreibt, indem er sie von anderen Themen abgrenzt und diese dabei benennt, dürfen in der Inhalterschließung des vorliegenden Dokuments die negativ markierten

⁵ Zum Begriff Paratext vgl. Genette, Gérard: Paratexte. Das Buch vom Beiwerk des Buches. Frankfurt am Main, Suhrkamp, 2001; Kreimeier, Klaus / Stanizek, Georg (Hrsg.): Paratexte in Literatur, Film und Fernsehen. Berlin, Akademie-Verlag, 2004.

Begriffe gerade nicht auftauchen. Eine maschinelle Extraktion wird solche Unterscheidungen nicht treffen können und ggf. irreführende Metadaten liefern.⁶

Die intellektuelle Inhaltsanalyse eines Dokuments als Ganzes wird in der Regel rasch zu einer groben **klassifikatorischen Erschließung** kommen können, so bald sie die Thematik als Ganzes erfasst und ins klassifikatorische Raster eingeordnet hat, weil dieses Raster im Vorwissen des Erschließers gespeichert sein kann. Systematische Zuordnungen wird ein automatisches Verfahren nicht direkt aus dem Wortmaterial erzeugen können, es sei denn, es finden sich im hervorgehobenen Vokabular eindeutige Übereinstimmungen mit den Benennungen von Klassen in einem Vergleichswortschatz. Die Chance des automatischen Klassifizierens liegt eher darin, dass innerhalb eines Ressourcenkorpus an eine Systemstelle oft schon viele Dokumente geknüpft sind; dann sind mit probabilistischen Verfahren unter Einbeziehung der verbalen Erschließung aus größeren Vergleichsmengen annähernd plausible Zuweisungen möglich. Das funktioniert vor allem mit groben Klassifikationen wie bei der automatischen Sachgruppenvergabe für Netzpublikationen im DNB-Projekt PETRUS.⁷ Es ist nur fraglich, welchen Wert eine solche Groberschließung für anspruchsvolle wissenschaftliche Inhaltsrecherchen wirklich hat.

Anreicherung von Titeldaten mit vorhandenen Sacherschließungsdaten

Die einfachste Form der Anreicherung ist das erweiterte, d. h. auch retrospektive Zuführen von Fremddaten⁸ aus der konventionellen bibliothekarischen Erschließung, wo vielfach noch großer Nachholbedarf besteht.

Seit einigen Jahren gibt es in deutschen Verbundkatalogen Anreicherungen nach dem **Werkcluster-Verfahren**,⁹ wie es von Magnus Pfeffer erstmals angewandt und in der Culturegraph-Umgebung weiter entwickelt wurde. Es setzt voraus, dass für mindestens eine Manifestation des Werkes im Erschließungsraum Sacherschließungsdaten vorliegen.

Neue Metadaten lassen sich auch aus anderen Metadaten z. B. über **Crosskonkordanzen** gewinnen. Je gröber die angestrebte Systematisierung sein darf, umso mehr Möglichkeiten ihrer Herleitung gibt es.

Bei all diesen Verfahren ist die Verbreitung klassifikatorischer Daten besonders erfolgreich, weil diese im internationalen Rahmen leichter zugänglich sind als die sprachabhängigen Schlagwörter.

Hier und da findet man Ansätze zur Nutzung fremdsprachiger Schlagwörter durch **maschinelle Übersetzung**. Bloße Übersetzungen sind zwar wegen der vielen ambivalenten Bedeutungen sehr problematisch, aber es lässt sich argumentieren, dass in einer Datenschicht, die ansonsten von der im Suchraum dominanten Sprachgemeinschaft nicht verstanden wird, Übersetzungsergebnisse mit einer Nützlichkeit von ca. 80 % einen nicht zu ignorierenden Mehrwert darstellen. Sie sollten als

⁶ Ein Beispiel: Aus dem Einleitungssatz „Neben Rilke ist Gottfried Benn der bedeutendste deutsche Lyriker des 20. Jahrhunderts. Hier soll aber nicht seine Lyrik behandelt werden, sondern seine Essays.“ würden maschinell extrahiert: Rilke + Benn + Lyrik + Essay. Irreführend wären Rilke + Lyrik. Intellektuell würden als Schlagwörter richtig „Benn, Gottfried“ und „Essay“ vergeben.

⁷ <http://www.dnb.de/DE/Wir/Projekte/Abgeschlossen/petrus.html>

⁸ Fremddaten sind zu verstehen als adäquate, zur Nutzung übernommene Metadaten aus anderen Erschließungsräumen, z. B. anderen Verbunddatenbanken.

⁹ Vgl. zuletzt Pfeffer, Magnus: Die Anreicherung der B3Kat-Echtdatenbank mit RVK-Notationen nach dem "Pfeffer-Verfahren". Vortrag beim RVK-Anwendertreffen 2014. http://rvk.uni-regensburg.de/images/stories/Conf2014/rvk_br3kat.pdf

zusätzliche Sucheinstiege indexiert, aber besser nicht als vermeintlich korrekte Inhaltsbeschreibung präsentiert werden.

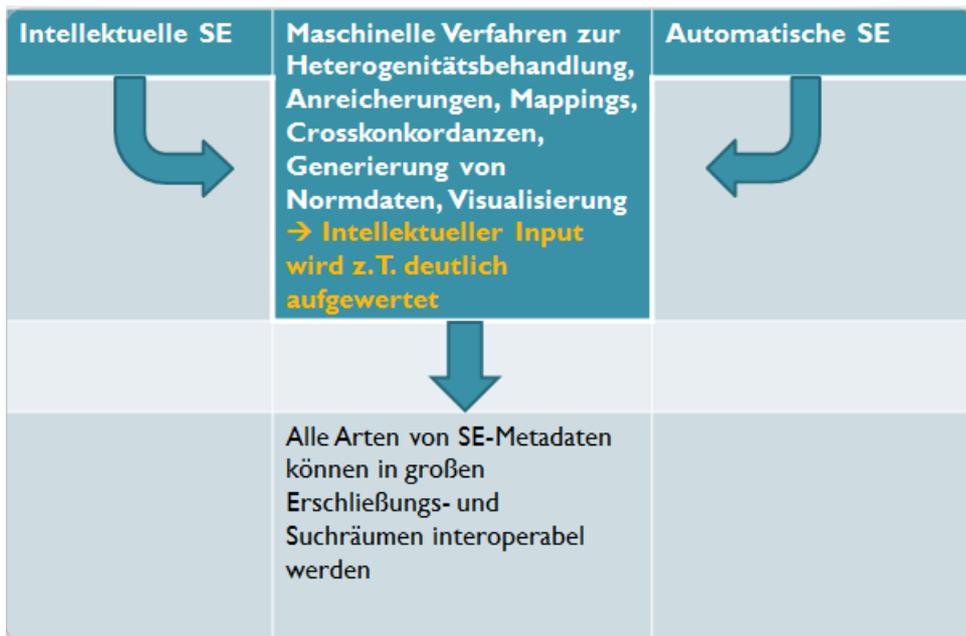
Alle Anreicherungen haben einen wichtigen quantitativen Aspekt: Bei der großen Zahl von Ressourcen im Suchraum, der mit Suchmaschinenkatalogen durchsucht wird, wird es immer wichtiger, dass in den Sacherschließungs-Suchfacetten die inhaltsrelevanten Begriffe hochfrequent bzw. umgekehrt die zuoberst angezeigten häufigsten Begriffe wirklich inhalts- und unterscheidungsrelevant sind.

Die heutige bibliothekarische Sacherschließung steht unter einem hohen Druck, quantitativ relevant zu werden, wenn sie in den heutigen Suchumgebungen nicht untergehen soll. Es lohnt sich, dieses Ziel zu verfolgen, auch wenn dabei mehr qualitative Inkonsistenz in Kauf zu nehmen ist. Die Konsequenz ist ein Szenario, in dem sowohl intellektuell als auch automatisch generierte Erschließung ihren Platz hat und beides für maschinelle Anreicherungen verwendbar ist.

Die Dichotomie von intellektuell vs. automatisch löst sich langsam auf. Dabei gelten die unterscheidenden Feststellungen durchaus weiter, müssen aber nicht zwingend zu einer schroffen qualitativen Abgrenzung beider Wege führen.

| Intellektuelle SE | Automatische SE |
|--|---|
| „Verstehende“ Inhaltsanalyse ist in der Lage, nicht adäquaten Paratext bzw. Metadaten zu erkennen | Linguistische Datenverarbeitung, z. B. Lemmatisierung, wörterbuchbasiert Grundlage: Volltext, Paratext, vorhandene Metadaten |
| Erzeugt Indexate in Anzeigeform (z. B. Schlagwortfolgen) mit dem Kriterium Verständlichkeit für Menschen | Erzeugt vor allem zusätzliche Sucheinstiege, nicht primär lesbare Inhaltsbeschreibung |
| Unabhängig von der Dokumentsprache, insofern konsistent | Immer abhängig von der Dokumentsprache |
| Kann tendenziell Unklarheit erkennen und Eindeutigkeit herstellen | Kann Verstehbarkeit im lesbaren Kontext nicht erkennen |
| Kann implizite Aspekte (z. B. geographische und sprachliche Bezüge) benennen | Wichtige, aber nicht explizit genannte Aspekte werden unterschlagen |
| Qualitativ überlegen | Qualitativ unterlegen |
| Quantitativ begrenzt | Großes quantitatives Potenzial |

Das heutige Szenario könnte man so beschreiben:



Exkurs: Sacherschließung ohne Metadaten?

Wegen der absehbaren Entwicklungsrichtung maschineller Erschließungsverfahren erübrigt sich beinahe die grundsätzliche Frage, ob angesichts der Volltextsuche Metadaten überhaupt noch notwendig sind. Denn die einzige Volltextauswertung, die nicht in irgendeiner Form zur Erzeugung von Metadaten führt, ist ein Display als Einblick in Teile des Originaldokuments zur Relevanzbeurteilung. Ein Beispiel hierfür sind die Google-Snippets, ein Highlighting hinterlegter Textausschnitte im Umfeld des Suchbegriffs, das bisher nur für Online-Ressourcen möglich ist.

Web
Bilder
Videos
Shopping
Bücher
Mehr ▾
Suchoptionen

Ungefähr 182 Ergebnisse (0,59 Sekunden)

Tankerunfälle und der Ersatz ökologischer Schäden: am ...

<https://books.google.de/books?isbn=3825895564>
Liliane C. Kappet - 2006 - Vorschau

62 Von einem „Umweltschaden“ spricht Art. 2 Ziff. I Richtlinie 2004/35/EG zur Umwelthaftung: „Umweltschaden“ ist: (a) Schädigung geschützter Arten ... Die Waffenstillstandsresolution 687 vom 3.4.1991 nach dem Golfkrieg 1990/91 (UN Doc.

Snippets basieren i.d.R. nicht auf vorhandenen Metadaten, sondern auf Volltext, und stellen auch keine Metadaten dar.

Der Golfkrieg (1990/91): eine Auswahlbibliographie - Seite 10

<https://books.google.de/books?isbn=3922852440>
Ingeborg Otto, Marianne Schmidt-Dumont - 1992 - Snippet-Ansicht - Mehr Ausgaben

Englisch Persischer Golf + Irak-Kuwait-Krise / -Krieg (2. August 1990 - 11. April 1991) + Verlauf internationaler Konflikte + Umweltzerstörung + Umweltkatastrophe + Folgen von Umweltschäden + Umweltverschmutzung + Kriegsfolgen ...

Der Deutsche Wirtschaftsanwalt 2008/2009: Handbuch Fur ...

<https://books.google.de/books?isbn=3939804029>
Ein Handbuch für Unternehmen, Lexxion - 2008 - Vollansicht

... bei der Sanierung von Umweltschäden, die durch den ersten Golfkrieg

Auch für Snippets¹⁰ sind allein strukturierte Auszeichnungsdaten geeignet, wirklich aussagekräftige Inhaltsinformationen über das Gesamtdokument zu liefern. Die Rezeption beliebiger kurzer Textschnipsel, die die gesuchten Wörter enthalten, überfordert in der Regel die kognitiven und ergonomischen Möglichkeiten des Suchenden. Im bibliographischen Kontext, also für bibliothekarische Zwecke, sind Metadaten für solche fortgeschrittene Anzeigetechniken grundsätzlich notwendig, und auch sonst dominieren Metadaten die Erschließungswelt im Web bei weitem. Da das Semantic Web aus Metadaten geflochten wird, dürfte die Frage, ob zur Inhaltserschließung auch digitaler Ressourcen langfristig Metadaten benötigt werden, im Grunde positiv beantwortet sein.

Neue Anforderungen an die Inhaltserschließung: „Kontrollierte Heterogenität“

Neben die unverändert gültigen Anforderungen (Finden – identifizieren – auswählen) treten neue. Der Metadaten-Input

- darf nicht (mehr) nur auf einen fest definierten Suchraum abzielen,
- darf nicht nur mit bestimmten Suchtechniken nutzbar sein,
- soll aber etablierten Standards folgen oder wenigstens mit solchen interoperabel sein,
- soll direkt oder nach Weiterverarbeitung zur Nutzung als Linked data im Web geeignet sein,
- soll neben Bibliotheken auch für andere Kultureinrichtungen attraktiv sein.

Nutzungsbezogene Konzepte der Sacherschließung können zwar innerhalb eines Erschließungs- und Suchraums oder für eine bestimmte Fachcommunity bestimmte Leitsysteme definieren, müssen aber prinzipiell alle Möglichkeiten der Inhaltserschließung einbeziehen, weil sonst regelmäßig große Potenziale ungenutzt blieben.

Auch Fachthesauri und –klassifikationen müssen ihr wertvolles Potenzial für die allgemeine Erschließungswelt entdecken und sich auch in diese Richtung öffnen, um nicht noch tiefer in einer Nische zu verschwinden. Umgekehrt müssen die Universalsysteme sich auf die Ergänzung durch fachspezifische Erschließung einlassen. Crosskonkordanzen sind hierzu ein guter Weg.

Den funktionalen Kontext dieser „kontrolliert heterogenen Welt“ beschreibt das **Schalenmodell**.¹¹

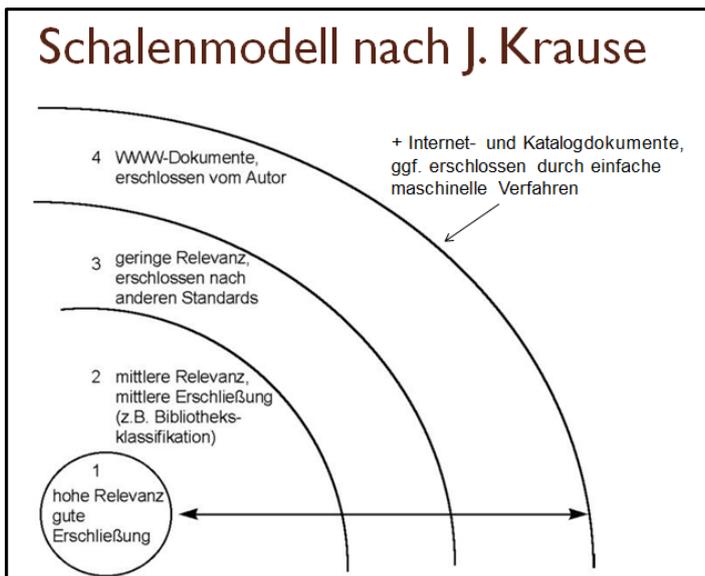
Die überwiegend intellektuelle Erschließung des Kerns, der auf einer bereits vorhandenen Menge vieler Millionen von Dokumenten bibliothekarischer Herkunft aufbaut, bildet eine qualitative Grundlage und kritische Masse für zahlreiche maschinelle Verfahren.

Dabei sind nicht in jeder Datenschicht feine Unterschiede in der Inhaltsbeschreibung darstellbar.

Den Kern bilden für die Universal-bibliotheken und Katalogverbände im deutschsprachigen Raum Titel mit Schlagwort-Indexaten aus kontrolliertem Vokabular sowie Daten aus Klassifikationen, die gepflegt werden.

¹⁰ Jedenfalls auf Metadaten basieren die sogenannten *rich snippets*, die zusätzliche Relevanzbewertungen auf der Basis der HTML-Markups von Webseiten ermöglichen.

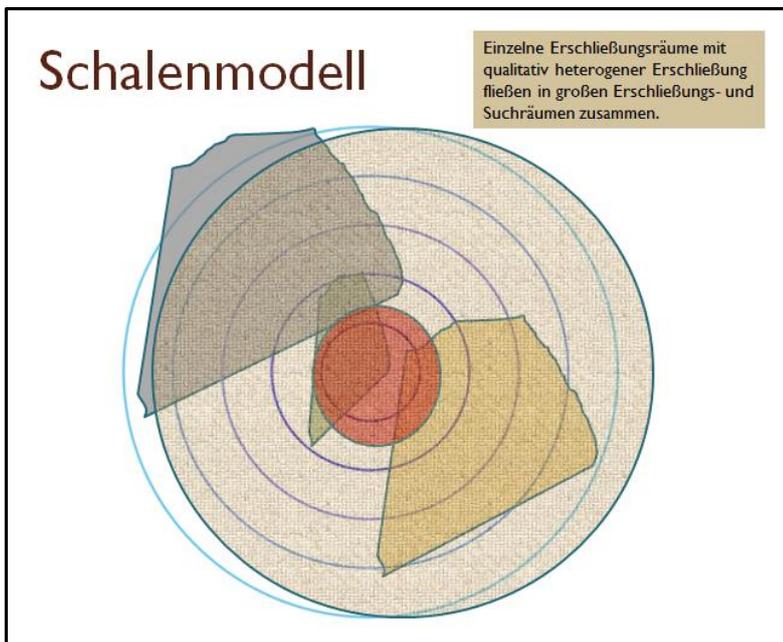
¹¹ Vgl. Krause, Jürgen: Kapitel D 13.2: Standardisierung und Heterogenität. In: Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation: ein Handbuch zur Einführung in die Informationswissenschaft und -praxis, 5. völlig neu gefasste Ausgabe 2003. München: Saur. (DGD-Schriftenreihe). Vgl. auch: Krause, J.: Sacherschließung in Deutschland. Lösungsansätze für eine Welt der polyzentrischen Informationsversorgung. Leitreferat zur 29. Jahrestagung der Gesellschaft für Klassifikation, Magdeburg 2005. http://archiv.tu-chemnitz.de/pub/2005/0061/data/juergen_krause.ppt



Neben gepflegten Klassifikationen und Normdaten müssen auch ungepflegte und inkonsistente Metadaten (z. B. freie Deskriptoren, fremdsprachige Termini, Extrakte, Abstracts) genutzt werden, und zwar als zusätzliche Sucheinstiege in einem Basic Index. So weit wie möglich sollten sie auf gepflegte Daten umgesetzt / gemappt werden, um die Erschließung qualitativ aufzubessern und die Anbindung möglichst vieler Ressourcen ans Semantic Web zu erleichtern. Es gibt bereits Mappings zu diesem Zweck, z. B. das multilinguale MACS = Multilingual Access to Subjects.¹²

Metadatentypen und -schichten sollten so weit wie möglich trennbar bleiben, damit sich je nach Suchraum und Suchinteresse wahlweise eine höhere virtuelle Konsistenz oder ein gewünschtes Verhältnis von Recall und Precision herstellen lässt.

Für neuere gedruckte Ressourcen sind weithin die Metadaten und urheberrechtsfreie Surrogate oder Ausschnitte aus dem Paratext die einzigen in Suchsystemen anzeigbaren Daten. Die Auszeichnung inhaltsrelevanter Daten durch Autoren, Verlage und andere Bearbeiter im



Publikationsprozess tritt zwar neben die bibliothekarische Erschließung, ist aber zunächst nur scheinbar qualitativ gleichwertig. Die bibliothekarische Erschließung hat den Vorteil der Standardisierung, insbesondere der Anbindung an Normdaten. Die z. B. im Buchhandel üblichen Schemata¹³ bleiben dahinter deutlich zurück. Wenn es gelingt, solche und besonders automatisch extrahierte Daten mittels Mappingtechniken zu Normdaten aufzurüsten, können die automatischen Verfahren eine neue Qualitätsstufe erreichen.

Einen Daten- und Suchraum mit inhaltsbeschreibenden Metadaten

aus gepflegten Systemen möglichst dicht zu durchsetzen, muss das vorrangige Ziel der bibliothekarischen Sacherschließung sein. Nach dem Schalenmodell ist die Aufwertung der Erschließung der äußeren Schalen vom Kern her wichtig. Trotz aller sonstigen, verbleibenden Heterogenität kommt es dabei auf eine weitgehende Interoperabilität mit den Normdaten an.

¹² Projekthomepage: http://www.nb.admin.ch/nb_professionnel/projektarbeit/00729/00733/index.html

¹³ Z. B. BIC und Thema. Vgl. <http://www.editeur.org/151/thema/en/welcome-to-thema/>

Die Sacherschließung wird also nach dem Schalenmodell in einem anderen Sinn zu einem „Kerngeschäft“. Wie auch immer man den Kern definiert und wie auch immer intellektuelle, semi-intellektuelle und automatische Verfahren ineinander greifen: sein Kennzeichen wird stets eine relativ hohe Metadatenqualität sein.

Exkurs: Sacherschließung und Discovery-Systeme

Wie stark muss die Discovery-Suche, die ganz auf große Suchmaschinenindizes setzt, die Entwicklung der Sacherschließung bestimmen?

Weil hier eine wirklich saubere inhaltliche Primärsuche nicht möglich ist, ist der Nutzer auf funktionierende Facetten zur Sucheinschränkung angewiesen. Hierfür braucht es hinreichende kritische Häufigkeit einzelner Deskriptoren bzw. Systematikstellen. Schlagwortvokabulare und Feinklassifikationen lassen sich dabei schwer zur Geltung bringen.

Nötig ist aber auch eine hinreichende Granularität, um der zunehmenden Differenzierung der wissenschaftlichen und der Sachthemen gerecht zu werden.

Zum Einstieg in thematische Zusammenhänge müsste in jedem Suchraum ein Leitsystem zur Verfügung stehen, das dem Nutzer je nach gewähltem Suchraumzuschnitt auch angeboten wird. Discovery-Systeme verhalten sich hier indifferent oder bieten gar keine Ansätze zum Navigieren mit Normdaten.

Ohne dass die Masse der Nutzer diesen Mangel wirklich bemerkt, zeigt er sich aus fachlicher Sicht deutlich, unabhängig vom Systemanbieter. Der eigentliche Paradigmenwandel in der heutigen Erschließung besteht ja gerade darin, dass sie, indem sie sich aus den Katalogen heraus ins Web erweitert, eine entitätsbezogene bzw. objektorientierte Richtung nimmt. Von den heutigen Discovery-Systemen könnte man kritisch sagen: es sind doch wieder nur gigantische Kataloge. Ihr Gott ist der Index; dieser ist gewaltig, aber die Flachheit der Ergebnisse wird dem Potenzial der Metadaten, das zum Nutzen des Informationssuchenden den bibliographischen Raum überschreiten könnte, meist nicht gerecht.

Das Entdecken, das sich diese Tools auf die Fahnen geschrieben haben, ist ein wichtiges Mittel der Wissensgewinnung, es kann aber nicht an die Stelle des Findens treten. Die angebliche Benutzerfreundlichkeit der Discovery sollte nicht zur Geringschätzung des Suchenden führen. Wer präzise Suchintentionen und ein Kerninteresse hat, darf nicht mit Nebensächlichkeiten abgespeist werden. Ein Kerninteresse muss mit den erwarteten hoch relevanten Kerninformationen bedient werden.

Mehrwert durch intellektuelle Erschließung

Den Output der Sacherschließung müssen sowohl Maschinen als auch Menschen verarbeiten können. Exemplarisch möchte ich die **Kontextualisierung** herausgreifen, die für die Verständlichkeit einer Inhaltsbeschreibung für Menschen eine große Rolle spielt. Das heißt, die einzelnen Metadaten-Elemente müssen in einem Kontext gezeigt werden, der es erlaubt, möglichst leicht Rückschlüsse auf die oft komplexen Inhalte zu gewinnen. Dazu gehört, dass die Kernthemen von den Nebenaspekten getrennt bzw. vor diesen hervorgehoben sind. Es wäre deshalb fragwürdig, die Elemente der Sacherschließung nur noch nach dem Kriterium einer mehr oder weniger konzidierten Nützlichkeit gleichordnend zu erfassen und zu präsentieren. Wo es möglich ist, sollen

Strukturen und Patterns geschaffen werden, die die visuell-kognitive Aufnahme und Selektion der Inhaltsinformationen erleichtern. In diesem Zusammenhang haben z. B. die Verbalisierung von Systematikenotationen, die Anzeige von Schlagwortfolgen, visualisierte Raum- und Zeitbezüge und wiederkehrende Muster bei Reihenfolgen und Anordnung in der Präsentation ihren Wert.

| | |
|---|---|
| 1 | Golfkrieg 1990-1991 Türkei Kurden Flucht Geschichte 1991 Jugendsachbuch |
| 2 | Golfkrieg 1990-1991 Umweltbelastung |
| 1 | Golfkrieg 1990-1991 Umwelthaftung Schadensersatzanspruch Auswirkung Internationales Umweltrecht Völkerrecht |
| 2 | Golfkrieg 1990-1991 Umweltkatastrophe |
| 7 | Golfkrieg 1990-1991 Umweltschaden |
| 1 | Golfkrieg 1990-1991 Umweltschaden Aufsatzsammlung |
| 1 | Golfkrieg 1990-1991 Umweltschaden Korallenriff Persischer Golf |
| 1 | Golfkrieg 1990-1991 Umweltschaden Ölunfall Meeresverschmutzung Aufsatzsammlung |
| 1 | Golfkrieg 1990-1991 Umweltverschmutzung |
| 4 | Golfkrieg 1990-1991 United Nations |

<< >> | SW-Ketten= "Golfkrieg 1990-1991 Umweltbelastung" OR SW-Ketten= "Golfkrieg 1990-1991 Umwelthaftung Schadensersatzanspruch Auswirkung Internationales Umweltrecht Völk...

[Sortiert nach Titel (Aufsteigend) dann Autor (Aufsteigend).] <<<<Titeldaten>>>>

| #/Treffer | Autor | Titel |
|-----------|-------------------------------|--|
| 1/14 | Hawley, T. M. | Against the fires of hell |
| 2/14 | | The environmental aftermath of the Gulf War |
| 3/14 | Golob, Richard S. | Environmental impact of war in the Arabian Gulf |
| 4/14 | Farīd, Ḥātim N. | Ġarīma deidda al-ḥayāt |
| 5/14 | Şādiq, Muḥammad | The Gulf War aftermath |
| 6/14 | Bāz, Fārūq al- | The Gulf War and the environment |
| 7/14 | Payne, Cymie R. | Gulf War reparations and the UN Compensation Commission |
| 8/14 | Bloom, Saul | Hidden casualties |
| 9/14 | Vogt, Helge Peter | Korallenriffe im Persischen Golf |
| 10/14 | Linden, Olof | Kriegets miljøeffekter |
| 11/14 | Kühnel, Matthias | Ökologische Auswirkungen des Golfkrieges |
| 12/14 | Sartorius, Rolf [Bearb.] | Ökologische Auswirkungen des Golfkrieges |
| 13/14 | Mackinnon, Michael | Quellen der Finsternis |
| 14/14 | *Abū-Zināda, *Abd-al-*Aziz H. | The status of coastal and marine habitats two years after the Gulf War oil spill |

Zur Kontextualisierung als Visualisierung von Zusammenhängen gibt es aber auch neue Möglichkeiten durch die Linked-data-Strukturen.

| Syntagmatisch-sequenzieller Kontext | Kontext im semantischen Netz |
|---|--|
| Fügt Einzelelemente (nicht zwingend Normdaten) zu einer Themenbeschreibung zusammen (präkombiniert oder präkoordiniert) | Führt von Einzelelementen der Inhaltsbeschreibung (Normdaten) in ein semantisches Umfeld, erschließt weitere Themen und Ressourcen |
| Sprachlich-begrifflich orientiert; Grundmuster: menschenlesbarer Text | an Entitäten orientiert; Grundmuster: maschinell auswertbares Geflecht von Aussagen über Ressourcen |

„Gezielte intellektuelle Arbeit“ einzusetzen, heißt zum einen, dass die intellektuell zu erschließenden Werke und Dokumente danach ausgewählt werden,

- wie groß die mit einem Erschließungsvorgang zu erzielende Wirkung (vor allem die Nachnutzbarkeit) voraussichtlich insgesamt ist,
- wie relevant der erzeugte Mehrwert für die Bedürfnisse der jeweils eigenen Community und Haupt-Nutzerschaft ist,
- aber auch: wie sinn- und wertvoll eine intellektuelle Erschließung im Kontext anderer Erschließungsmöglichkeiten erscheint.

Es heißt aber auch, intellektuelle Arbeit dort zu investieren, wo sie das Instrumentarium zur maschinellen Erschließung verbessert: zur Erstellung von Crosskonkordanzen, zum Ausbau des semantischen Netzes sowie zu Test und Evaluation automatischer Erschließungsverfahren. Intellektuelle Arbeit ist natürlich besonders wichtig für die Normdaten- und Klassifikationspflege. So ist es schwer vorstellbar, dass die Notwendigkeit neuer Termini oder das Obsoletwerden von Termini einer Dokumentationssprache im Licht der Entwicklung einer Fachterminologie automatisch erkannt werden könnte.

Automatische und maschinelle Verfahren müssen in Zukunft ihren festen Platz im Gesamtsystem der Erschließung haben. Ihr Output ist aber an den Standards der intellektuell gesteuerten Erschließung zu messen, nicht umgekehrt. Sonst kann die Vision eines semantischen Netzes, zu dem Bibliotheken einen wesentlichen qualitativen Beitrag leisten wollen, nicht Wirklichkeit werden.

Sowohl die beachtlichen Erschließungsleistungen der letzten Jahrzehnte als auch der internationale Kontext unserer Arbeit fordern uns auf, die bibliothekarische Sacherschließung zu einer effizienten, über das Web vernetzten und in vieler Hinsicht offen zugänglichen Mehrwert-Dienstleistung auszubauen. Intellektuelle und automatische Ansätze müssen nicht nur irgendwie interoperabel werden, sondern es gilt auch daran zu arbeiten, beide in ein neues System von Standards zu integrieren. Das bedeutet, auch automatische Erschließung nach Qualitätsstandards zu betreiben und, anstatt sie nur als Einsparmöglichkeit zu sehen, ihre Nutzungsperspektiven in der beschriebenen größeren Dimension im Blick zu haben.