Was ist Inhaltliche Erschließung?

Der Inhalt von Medien erschließt sich aus unterschiedlichen Quellen: aus dem Titel, aus dem Inhaltsverzeichnis, aus einem zusammenfassenden Text auf dem Cover eines Buches, aus Abstracts, aus einzelnen Texten eines Mediums oder sogar aus dem Volltext… -

Allen gemeinsam ist, dass daraus Begriffe (**Stichworte**) wahrgenommen werden, die den Inhalt mehr oder weniger repräsentieren, nach denen auch gesucht würde…

Oft sind diese Begriffe nicht ausreichend. Inhaltliche Erschließung ist daher der Versuch, ein vorliegendes Medium inhaltlich zu analysieren und in geeigneter Form strukturiert inhaltlich zu beschreiben.

Das geschieht primär durch Sprache, natürliche Sprache – allerdings mit **normiertem Vokabular**, normiertenBegriffen, die mit anderen Begriffen, die dasselbe meinen aber anders lauten (Synonyme) in einem Normdatensatz verknüpft sind und die so bei einer Suche (Retrieval) automatisch mitgesucht werden. In einem Normdatensatz sind auch verwandte Begriffe dokumentiert, Ober- und Unterbegriffe – ob assoziativ, parallel oder hierarchisch – unter Umständen ein ganzes thematisches Feld. Die **Gemeinsame Normdatei** (GND) wird im deutschsprachigen wissenschaftlichen Bibliothekswesen verwendet und umfasst sämtliche für die bibliografische Beschreibung nötigen normierten Begriffe (Normdatensätze), auch jene für die verbale inhaltliche Erschließung.

Diese Begriffe, die den Inhalt eines Mediums als Ganzes oder wesentliche thematische Aspekte beschreiben, nennt man **Schlagworte**. Sie werden in den bibliografischen Datensatz des Mediums eingegeben. Dies ist mindestens ein Schlagwort, es können aber auch mehrere Schlagworte sein, die miteinander eine oder mehrere **Schlagwortfolge(n)** bilden. Es gibt unterschiedliche Arten von Schlagworten: Schlagworte zu Personen (Personenschlagworte – Personennamen, Familien, Werke von Personen/Titeldatensätze,…), geografische Schlagworte, Sachschlagworte, Zeitschlagworte und Formangaben zum Genre bzw. zur Art des Mediums.

Die Vergabe der Schlagwörter erfolgt nach einfachen Regeln, den sog. **RSWK** (Regeln für die Schlagwortkatalogisierung) und folgt einer bestimmten **Reihenfolge** (Syntax): **P**ersonenschlagwort – **g**eografisches Schlagwort – **S**achschlagwort – **Z**eitschlagwort. Die **F**ormangabe und die Zielgruppe werden in eigenen Kategorien eingetragen und sind nicht (mehr) Teil der Schlagwortfolge. Als Faustregel hilft vielleicht die Vorstellung aus dem Fremdsprachenunterricht vom sinnerfassenden Lesen: bilde 2 bis 3 Sätze zum Inhalt des Mediums, extrahiere die Substantive und bilde eine Schlagwortfolge…

Die Tradition, Schlagwortfolgen zu bilden, ist jedoch nicht in allen deutschsprachigen Verbünden normative Praxis – es werden auch Einzelschlagwörter aus der GND vergeben und verknüpft. Sie bilden zumindest für die unterschiedlichen Discovery Systeme einen Anker für die thematische Suche – Schlagwortfolgen sind hingegen auch für die LeserInnen auf einen Blick leichter zu erfassen.

Die Kreation neuer Schlagwörter (Ansetzung/Erfassung) in der GND (s. o.) folgt im Wesentlichen dem Regelwerk **RDA** (Resource Description and Access) und wird gegenwärtig überarbeitet (Stand September 2018).

Zum Nachweis, dass ein Begriff in der Fachwelt wissenschaftlich akzeptiert ist und dem neuesten Wissensstand entspricht, gibt die sog. **Liste der Nachschlagewerke** (NSW) eine Reihenfolge sämtlicher für die einzelnen Fachgebiete zu konsultierenden Nachschlagewerke vor, die für die Definition bzw. den wissenschaftlichen Nachweis eines neuen Begriffes relevant sind. Selbstverständlich sind auch Internetquellen zulässig.

Als Schlagwort wird zunächst das **engste Schlagwort**, also das am präzisesten zutreffende vergeben; es können mehrere Schlagwortfolgen vergeben werden, wenn eine Folge nicht ausreicht, um die unterschiedlichen Aspekte eines Werkes abzubilden. Je nach vergebenen Schlagworten ergibt sich auch die sog. Indexierungstiefe, also die Granularität der inhaltlichen verbalen Erschließung – mit welchen Begriffen wird der Sucheinstieg gelingen…?! Hilfreich ist dabei immer die Überlegung, sich in die Position der LeserIn zu versetzen – wie würde ich suchen… - als ExpertIn, als Laie…? Die Genauigkeit bzw. das Ausmaß an inhaltlicher Erschließung ist natürlich immer eine Zeitfrage – dennoch hängt davon die Qualität des Suchergebnisses ab. Es mögen viele Treffer kommen (der Recall ist also gut) – stimmt das Ergebnis aber auch mit dem tatsächlich Gesuchten und Erwarteten überein (das ist die Precision)? Zwischen diesen Parametern bewegt sich also die adäquate Schlagwortvergabe…

Zusätzlich zur **verbalen inhaltlichen Erschließung** durch Schlagworte besteht die Möglichkeit, einzelne thematische Bereiche mehr oder weniger umfassend mit Hilfe von Klassifikationssystematiken – also durch Einteilung in Klassen – in Form von **Notationen** inhaltlich zu kennzeichnen – die **klassifikatorische inhaltliche Erschließung**. Notationen sind Codes, die aus Zahlen und/oder Buchstaben (numerisch, alphanumerisch,…) bestehen können und die auch eine verbale Benennung haben.

Es gibt unterschiedliche Systeme – im Österreichischen Bibliothekenverbund werden die **BK (Basisklassifikation)**, die **RVK (Regensburger Verbundklassifikation)**, die **DDC (Dewey Dezimalklassifikation)** und die **MSC (Mathematics Subject Classification)** verwendet (Näheres dazu s. unter Verbundklassifikation). Je nach Verwendungszweck werden gröbere oder feiner strukturierte Klassifikationssystematiken verwendet. Wenn z. B. die Notation Teil einer zu vergebenden Signatur sein soll und die Medien als Freihandbestand systematisch aufgestellt werden, empfiehlt sich eine feine Systematik, wie z. B. die RVK, die im deutschsprachigen Raum seit 50 Jahren die am weitesten verbreitete Klassifikation in wissenschaftlichen Bibliotheken darstellt. Im Angloamerikanischen Raum kommt der DDC diese Bedeutung zu – sie ist auch die älteste Klassifikation und findet mittlerweile ebenfalls im deutschsprachigen Raum partielle Anwendung.

Zur großflächigeren virtuellen Diversifizierung von Fachgebieten eignen sich hingegen gröbere Systematiken, wie z. B. die BK, die thematische Teilbestände gut isoliert repräsentieren und durchsuchbar machen kann.

Zur geschichtlichen Entwicklung der unterschiedlichen inhaltlichen Erschließungsmethoden und Regelwerke -> s. Link zu ULG Skript E 3.1 Erschließungsmethoden und Metadatenmanagement.

Automatische inklusive intellektueller inhaltlicher Erschließung

Gegenwärtig befindet sich die Inhaltliche Erschließung in einer Umbruchsphase. Die konventionellen intellektuell-manuellen Verfahren finden Grenzen in ihrer Anwendbarkeit auf umfangreiche Medienpakete elektronischer Medien (E-books,

E-journals,…). Die maschinellen Verfahren automatischer Inhaltserschließung (Vergabe von Notationen, Schlagworten) sind allerdings qualitativ noch nicht so weit, dass jegliche intellektuelle Arbeit inhaltlicher Erschließung obsolet wäre. Eine Kombination von beiden (Methodenmix) ist das Mittel der Wahl – die richtige Gewichtung von redaktionellem Input und maschineller Verarbeitung großer Datenmengen gilt es noch zu finden (halbautomatische Verfahren, Kontrollroutinen durch Clustering, d. h. vergleichende Verfahren, Ähnlichkeitssuchen und fallbasiertes Schließen/Abgleichen

z. B. mit Hilfe von Konkordanzen klassifikatorischer Inhaltserschließung, aber auch Abgleiche von ähnlichen Titeln und Beschlagwortungen – co-occurence u. a. m.).

Auch auf diesem Hintergrund ist die intellektuell-manuelle inhaltliche Erschließung – sowohl die verbale als auch die klassifikatorische – als Parametrisierung für maschinelles Lernen (Definition von Standards) weiterhin wichtig, wiewohl sich die Erschließungsarbeit in Zukunft vielleicht mehr auf die Metaebene (redaktionelle Betreuung maschineller Verfahren, Relationierung von Normdaten, Erstellung von Thesauri u. a. m.) verlagern oder sich weiter fachspezifisch als Nische etablieren wird (FID – Fachinformationsdienste, inhaltliche Erschließung und Dokumentation von Spezial- und Sondersammlungen,…).

Stand: 6.12.2018