



Die Sacherschließung in der Datenbank „Zeitschrifteninhaltsdienst Theologie“ Hilger Weisweiler

Über die Datenbank *Zeitschrifteninhaltsdienst Theologie (ZID)* der UB Tübingen wurde in dieser Zeitschrift schon des öfteren berichtet. Diesmal soll von einer Spezialfrage die Rede sein, nämlich von der Konzeption der Sacherschließung.

Jedem Bibliothekspraktiker ist bekannt, daß die perfekte und stringente Sacherschließung einer fest umrissenen Datenmenge zwar das Ideal, daß sie aber, wie dies mit Idealen der Fall zu sein pflegt, kaum zu erreichen ist. Zu dieser Diskrepanz von Ideal und Wirklichkeit führen meist historische Gründe. So auch beim ZID. Die folgenden Zeilen wollen die historische Entwicklung der Sacherschließung im ZID bis hin zu ihrer gegenwärtigen Form darstellen, auf deren Stärken hinweisen und ihre Schwächen nicht verschweigen.

Die gedruckten Hefte des ZID alten Typs, eines klassischen Current-Contents-Dienstes, verfügten - neben einem Autorenregister - über Indizes der in den angezeigten Dokumenten behandelten Personen und Bibelstellen: der pragmatische Ansatz einer Sacherschließung, wie sie mit dünner Personaldecke zu leisten war. Bei der Umstellung vom Cut-and-paste-Verfahren auf Datenbankfassung (1995) war zunächst nur an die Fortführung dieser Minimalsacherschließung gedacht; denn die Personalsituation war unverändert (gemessen an dem Anwachsen der zu bewältigenden Arbeit hatte sie sich sogar erheblich verschärft). Allerdings wurde die Indizierung von Bibeltexten ausgeweitet auf die Versebene (zuvor hatte sie sich nur auf die Kapitelebene erstreckt), vor allem jedoch auf Texte aus dem Umfeld des Alten und Neuen Testaments (Apokryphen, Pseudepigraphen, Apostolische Väter, Qumran-Texte, Nag-Hammadi-Schriften), die ähnliche Literaturgattungen wie die biblischen Schriften repräsentieren, und auf innerbiblische Traditionsschichten (z. B. Spruchquelle). Darüber hinaus bestand nun mit der Möglichkeit der Stichwortrecherche in der Datenbank eine wenn auch unzulängliche Quasi-Sacherschließung, die ein reines Druckprodukt natürlich niemals bieten kann.

Daß eine weitere Ausweitung der ansatzweisen verbalen Sacherschließung zunächst unrealistisch erschien, führte dazu, daß für Personenschlagwörter und Textschlagwörter, die als Einzeldesktoren in die Titeldatensätze geschrieben wurden, jeweils getrennte Kategorienbereiche vorgesehen wurden: 100 bis 105 für Personen- und 150 bis 155 für Textschlagwörter (biblische und parabiblische Texte). Für die beiden Kategorienbereiche wurden in der allegro-Datenbank getrennte Indizes vorgesehen.

Vordringlich und mit dem geringen personellen Potential leichter realisierbar erschien die Einführung einer klassifikatorischen Erschließung. Nur die Existenz dieser Komponente gewährleistet ja eine besonders wichtige Funktion jeder praktikablen Datenbank: die Möglichkeit der Recherche nach Publikationen zu einem Fach oder einer Disziplin. Um dieser Möglichkeit willen wurde die Klassifikation bewußt grob angelegt: Sie besteht aus zwölf teils zwei-, teils dreistufig untergliederten Hauptgruppen mit insgesamt 132 Klassen. Ein Dokument darf nach der internen Konvention bis zu vier Klassen zugeordnet werden. Theoretisch wäre natürlich auch die Vergabe von mehr als vier Notationen möglich. Für die Notationen wurde die Kategorie 191 vorgesehen, die gegebenenfalls mehrfach belegt wird. In Stammsätzen wurden neben den Notationen selbst bis zu fünf Klassenbeschreibungen erfaßt, mit deren Hilfe auch eine verbale Suche, sozusagen mittels weiten Schlagworts, möglich wurde. Für Klassenbeschreibungen und Notationen wurde zunächst ein gemeinsamer Index generiert; von den verbalen Klassenbeschreibungen führte der allegro-typische automatische Indexsprung zu den Notationen und damit

zu der unter diesen nachgewiesenen Literatur. Die Klassifizierung rezenter Dokumente wurde im Mai 1995 aufgenommen und wird bis heute, im wesentlichen unverändert, fortgeführt.

Obwohl sich an der Personalsituation zunächst nichts änderte (und bis Oktober 1998 auch nichts ändern wird), führte die von uns als unbefriedigend angesehene Sacherschließungssituation zu weiteren Aktivitäten. Ausgleich für das Fehlen von Personal schaffte erhöhter persönlicher Einsatz, auf die Dauer natürlich kein haltbarer Zustand. Im März 1996 wurde mit der vollen verbalen Sacherschließung der rezenter Dokumente nach der SWD begonnen. Seitdem werden pro Monat durchschnittlich 1.500 aktuelle Zeit- und Festschriftenaufsätze intensiv verschlagwortet. Literatur, die von der Verschlagwortung - etwa wegen sogenannter minderer Wichtigkeit oder zu geringer Seitenzahl - ausgenommen wäre, gibt es nicht. Für Schlagwörter, die sich nicht auf Personen oder (para)biblische Texte beziehen, wurden die Kategorien 156 bis 159 vorgesehen, in denen jeweils an den RSWK-Regeln orientierte Schlagwortketten erfaßt werden können, in der Regel jedoch nur eine Kette pro Dokument. Indiziert wurden die Schlagwörter zunächst als Einzeldescriptoren in ein eigenes Register. Die postkoordinierende Sachrecherche war nun - auch indexübergreifend - möglich.

Der nächste Schritt war eine tiefgreifende Erweiterung des Indexprogramms, die die Darstellung und automatische innere Permutation der Schlagwortketten im Sachschlagwortregister ermöglichte, wohl der entscheidende Schritt auf dem Weg zu einer dem Ideal der RSWK nahekommenden Sacherschließung. Die erste CD-ROM mit automatischer Sachschlagwortketten-Permutation wurde im Juli 1997 ausgeliefert. Bis zu acht Glieder einer Sachschlagwortkette permutieren automatisch. Auch Aspekt- und Forms Schlagwörter permutieren mit, was sich durchaus als sinnvoll erwiesen hat. Daß mit dieser Lösung über die Anforderungen der RSWK hinausgegangen wurde, war beabsichtigt. Außen vor blieben zunächst die Personen- und Textschlagwörter, die sich nicht in die automatische volle Permutation einbinden ließen, es sei denn um den Preis endloser Bedingungsprüfungen, die das Indexprogramm allzu stark belastet hätten. Das Fehlen von Personen- und Textschlagwörtern führte natürlich in nicht seltenen Einzelfällen zu einer inneren Unplausibilität der Sachschlagwortketten, was von Rezensenten mit Recht kritisiert worden ist, in Tübingen jedoch längst vor dem Verlauten dieser Kritik bekannt war.

Für das dargestellte Problem wurde folgende Lösung gefunden:

Die Trennung der verschiedenen Schlagwortindizes wurde aufgegeben. Die erste CD-ROM mit einem einheitlichen Schlagwort(ketten)index wurde im Juli 1998 ausgeliefert. Übrigens enthält dieser Index mittlerweile auch die verbalen Klassenbeschreibungen zu den klassifikatorischen Notationen als Verwei-

sungsbegriffe; von ihnen aus gelangt der Nutzer mittels automatischen Indexsprungs in den Notationenindex.

Existieren in einer Dokumentbeschreibung ein bzw. mehrere Personenschlagwörter oder ein bzw. mehrere Textschlagwörter und zusätzlich eine oder mehrere Sachschlagwortketten, so stellt das Indexprogramm hinter jedes existierende Personenschlagwort bzw. Textschlagwort die Sachschlagwortkette, die außerdem auch ohne voranstehendes Personen- oder Textschlagwort im Index abgelegt wird und permutiert.

Existieren mehrere Personenschlagwörter, so permutieren diese zusätzlich miteinander; ebenso tun dies mehrere Textschlagwörter und schließlich auch Personenschlagwörter mit Textschlagwörtern. Es entstehen also, je nach Schlagwortvergabe, zahlreiche Indexketten, weit über die Regeln der RSWK hinaus und im Sinne einer wirklich multidimensionalen präkoordinierenden verbalen Sacherschließung. Die anfängliche Entscheidung, Personen- und Textschlagwörter in getrennte Kategorienbereiche zu schreiben statt in die bereits in der Dokumentbeschreibung koordinierten Schlagwortketten, erwies sich im nachhinein sogar als segensreich; denn nur dadurch wurde die Kettengenerierung in ihrer ganzen Vielfalt möglich.

Ein sehr einfaches Beispiel: Aus dem Personenschlagwort „Diana <Wales, Prinzessin>“ und der Sachschlagwortkette „Idol ; Religionsersatz“ wird im Index eine nichtpermutierende Kette „Diana <Wales, Prinzessin> ; Idol ; Religionsersatz“. Die daneben im Index abgelegte Sachschlagwortkette „Idol ; Religionsersatz“ permutiert darüber hinaus zu einer Kette „Religionsersatz ; Idol“. Beide reinen Sachschlagwortketten sind für die Sachrecherche ebenso nützlich und plausibel wie die Kette mit dem voranstehenden Personenschlagwort, je nach dem die Recherche leitenden Interesse des Nutzers.

Übrigens: Das Semikolon wurde als Trennzeichen zwischen den Kettengliedern gewählt, um die allegro-typische sogenannte variable Trunkierung zu ermöglichen: in diesem Fall die Reduzierung der Indexketten (jeder Kombinations- oder Permutationsstufe) auf einen Einzeldescriptor.

Ein Sonderfall: Existieren ein oder mehrere Personen- oder Textschlagwörter und zusätzlich ein einzelnes Sachschlagwort in einer der Kategorien 156 bis 158, so wird unterstellt, daß dieses Sachschlagwort nur ein Aspektschlagwort zu dem Personen- bzw. Textschlagwort ist. In diesem Fall wird es nicht einzeln im Index abgelegt, sondern permutiert mit den Personen- bzw. Textschlagwörtern. Beispiel: Aus den in getrennten Kategorien erfaßten Schlagwörtern „Augustinus, Aurelius“ und „Literaturbericht“ werden für den Index die Ketten „Augustinus, Aurelius ; Literaturbericht“ und „Literaturbericht ; Augustinus, Aurelius“ generiert.

Schließlich besteht - allerdings erst seit neuestem - die Möglichkeit, Sachschlagwörter, deren Kombination oder Permutation mit Personen- bzw. Textschlagwörtern nicht gewünscht wird, in eine entsprechend parametrisierte Kategorie (159) zu schreiben. Liegt in 159 eine Schlagwortkette, so permutiert diese jedoch in sich selbst.

Ich hatte eingangs versprochen, nicht nur die Stärken der Sacherschließungskonzeption des ZID zu preisen, sondern auch die Schwächen nicht zu verschweigen. Sie seien im folgenden aufgezählt:

Mißlich ist vor allem der unterschiedliche Erschließungsgrad von Dokumentnachweisen in ein und derselben Datenbank. Dokumentnachweise, die aufgrund der laufend durchgeführten Retrokonversion des gedruckten ZID alten Typs in die Datenbank gelangen, sind - von wenigen Ausnahmefällen abgesehen - verbal nur durch Personen- und Textschlagwörter und überhaupt nicht klassifikatorisch erschlossen. Eine vollgültige sachliche Erschließung wäre hier ohne Autopsie nicht in seriöser Weise zu leisten. Es verbietet sich für uns jedoch auch nur der Gedanke an eine autoptische Bearbeitung dieser Dokumente. Rein quantitativ betrachtet, ist der Sacherschließungsgrad innerhalb der Datenbank nicht schlecht: Nur etwa 25 % der nachgewiesenen Dokumente sind überhaupt nicht sachlich erschlossen; bei diesen handelt es sich durchweg um ältere, vor Frühjahr 1995 erschienene Literatur.

Die automatische Generierung von Ketten aus Personen- bzw. Textschlagwörtern und Sachschlagwörtern kann in (seltenen) Einzelfällen zu unplausiblen Indexketten führen. Dieser Mißstand muß jedoch um der zahlreichen Vorteile des Verfahrens willen in Kauf genommen werden.

Wie eingangs angedeutet: Ideale wollen angestrebt sein; erreicht werden sie selten. Oft ist es unumgänglich, Entscheidungen zu treffen, um deren Nichtidealität man weiß, wenn dadurch mehr erreicht wird als durch die Fixierung auf das Ideal.

