

# Automatisierte und intellektuelle Verschlagwortung am Beispiel psychologischer Fachliteratur im Vergleich

Michael Gerards und Günter Krampen

## 1 Einleitung

› Nicht nur wegen der stark zugenommenen und weiter zunehmenden Publikationszahlen in den Wissenschaften, sondern auch und vor allem wegen dem explosionsartigen Anstieg digitaler Medien sind Katalogisierungen, Klassifizierungen und Verschlagwortungen von Neuerscheinungen nicht mehr alleine intellektuell und händisch, also via „Kopf und Hand“ zu bewältigen, da dafür die Ressourcen fehlen. Notwendig wurde und wird immer stärker ein mehr oder weniger automatisiertes Vorgehen, bei dem sprachanalytische Software unterstützend oder womöglich auch alleine für die Vergabe von Klassifikationen und Deskriptoren für die Ein- und Zuordnung wissenschaftlicher Literatur die Aufgaben von Bibliothekaren und Fachwissenschaftlern ganz oder zumindest teilweise übernimmt. Die Qualität der automatisierten, teilautomatisierten bzw. intellektuellen Verschlagwortung von Neuerscheinungen ist für deren Genauigkeit (Differenziertheit und Auflösung) und Validität (Korrektheit) ausschlaggebend und dafür (mit-)entscheidend, ob die Quellen in der heute automatisiert via Internet oder Intranet durchgeführten Literatur- und Informationsrecherche aufgefunden werden. Qualitativ hochwertige, gültige Verschlagwortungen tragen entscheidend dazu bei, dass dem Nutzer weder „endlose“ sogenannte „Trefferlisten“ (die auf ungenauen Recherchestrategien, aber auch auf zu groben, wenig validen Verschlagwortungen mit zu geringer Auflösung beruhen) noch ihr Gegenteil, nämlich keine oder nur sehr wenige „Treffer“ (was ebenso auf ungeeignete Recherchestrategien, aber auch auf zu spezifische Verschlagwortungen mit zu hoher Auflösung zurückzuführen ist) geboten werden. Beide Rechercheergebnisse behindern und verzögern die Bibliotheksarbeit und Informationsrecherche von Studierenden und Wissenschaftlern, da sie auf jeden Fall die wertvolle Ressource „Zeit“ strapazieren, darüber hinaus aber auch zu Frustration und De-Motivierungen oder – noch schlimmer – zu Fehleinschätzungen des

Über die Evaluation des automatischen Indexierungssystems AUTINDEX, das seit 2006 im Rahmen der Dokumentation psychologischer Literatur und Medien in der Datenbank PSYINDEX eingesetzt wird, wird berichtet. AUTINDEX generiert aus den vorgegebenen Titeln, Zusammenfassungen und Autorenschlagworten Deskriptorvorschläge, die vom fachwissenschaftlichen Dokumentar in sein intellektuell erstelltes Indexat übernommen werden können. Eine Analyse von 11.887 bzw. 13.941 Dokumenten der Publikationsjahrgänge 2002 bis 2010, die ohne bzw. mit Unterstützung durch AUTINDEX verschlagwortet wurden, macht deutlich, dass (1) es mit der Einführung von AUTINDEX zu einem leichten Anstieg der Anzahl vergebener Deskriptoren gekommen ist, (2) ein durchaus substantieller Anteil an Deskriptoren im automatisch bzw. intellektuell erstellten Indexat übereinstimmen und (3) trotzdem eine Überarbeitung des automatisch erstellten Indexats durch den Dokumentar notwendig ist, um die Qualitätsansprüche einer Informationsdatenbank zu erfüllen.

This is a report on the evaluation of the automated indexing system AUTINDEX, which has been used since 2006 in the context of the documentation on psychology literature and media in the PSYINDEX data base. From predefined titles, summaries and authors' indexes, AUTINDEX generates proposed descriptors, which the specialized documentalists can integrate into his or her intellectually compiled index. An analysis of 11,887 and 13,941 documents indexed, respectively, without or with support from AUTINDEX showed that (1) the introduction of AUTINDEX resulted in a slight increase of the number of assigned descriptors; (2) a substantial part of descriptors in the intellectually or automatically compiled index were identical; and, (3) in any case, revision of the automatically index by a documentalists is indispensable to fulfill the quality expectations of a data base.

aktuellen Wissensstandes zu einem Thema bzw. bei einer Fragestellung führen können.

Im Folgenden werden die empirischen Befunde einer exemplarischen systematischen Studie zum Vergleich der intellektuellen und der (teil-)automatisierten Verschlagwortung vorgestellt. Das Beispiel bezieht sich auf die Dokumentation (inklusive Verschlagwortung) von Neuerscheinungen aus der Psychologie im deutschsprachigen Bereich, die zu den zentralen Aufgabenbereichen des *Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID)* gehören. Kontinuierlich und zeitnah werden im ZPID alle

für die Psychologie relevanten Neuerscheinungen der Fachliteraturgattungen Monografien, Dissertationen, Sammelwerke, Beiträge in Fachzeitschriften und Editionen, Institutsberichte („graue Literatur“) sowie audiovisuelle Medien erfasst, die im deutschsprachigen Bereich in deutscher oder englischer Sprache konventionell und/oder digital publiziert werden. Die daraus resultierende Fachliteraturdatenbank PSYNDEX, die über alle großen Hochschulbibliotheken und auch für Individualnutzer (über [www.zpid.de](http://www.zpid.de)) online direkt verfügbar ist, umfasst neben den vollständigen bibliographischen Angaben Metadaten, die sich u.a. auf eine Zusammenfassung der Publikation, Deskriptoren, Klassifikationscodes (für psychologische Subdisziplinen) und mehr oder weniger standardisierte Phrasen mit Zusatzinformationen (vor allem bei empirischen Studien) beziehen.

Für die Deskriptorvergabe, die intellektuell nach dem Autopsie-Prinzip durch Fachwissenschaftler vorgenommen wird, wird seit 2006 im ZPID das vom Institut der Gesellschaft zur Förderung der Angewandten Informationsforschung e.V. (IAI) entwickelte Softwarepaket AUTINDEX (= Automatic Indexing) zur automatischen Extraktion von Schlagworten aus deutschen und englischen Texten zur Unterstützung des Indexierungsprozesses für die Datenbank PSYNDEX ergänzend eingesetzt. Auf der Grundlage eines mittels einer etwa 23.500 Begriffe umfassenden Indikatorenliste erweiterten Thesaurus generiert AUTINDEX anhand intelligenter linguistischer Verfahren in Kombination mit statistischen Funktionen aus den Titeln, Zusammenfassungen und den ggf. von den Autoren aufgeführten Schlagworten einer Neuerscheinung Deskriptorvorschläge, die dem fachwissenschaftlichen Dokumentar zur Auswahl angezeigt und von ihm – sofern er sie als zum Dokumentinhalt passend bewertet – übernommen werden. Als Ergänzung und Erweiterung von bereits 2004 und 2006 erfolgten Evaluationen des Einsatzes von AUTINDEX im ZPID<sup>1</sup> wurde eine weitere Evaluationsstudie durchgeführt, die die Leistungen von AUTINDEX anhand einer großen Dokumentmenge ermittelt und über deren Befunde hier berichtet wird. Dabei stehen zwei Fragestellungen im Vordergrund: (1) Hat die Einführung der automatischen Indexierung das Indexierungsverhalten der fachwissenschaftlich qualifizierten Dokumentare verändert? (2) In welchem Ausmaß schlägt AUTINDEX Deskriptoren vor, die vom fachwissenschaftlichen Dokumentar als zur Neuerscheinung passend bewertet und von ihm übernommen werden?

<sup>1</sup> <http://www.zpid.de/download/PSYNDEXmaterial/autindex.pdf> [9. Juli 2012]

## 2 Datenbasis der Studie

In die Auswertungen gingen insgesamt 13.941 Dokumente aus PSYNDEX ein, die die folgenden Kriterien erfüllen:

- Die Publikation wurde zwischen 2006 und Juni 2010 veröffentlicht, also zu einem Zeitpunkt, an dem AUTINDEX bereits im ZPID eingesetzt wurde.
- Alle Dokumente sind deutschsprachig, da AUTINDEX bislang nur für diesen Teilbereich der Dokumentation an die Bedürfnisse des Faches Psychologie angepasst wurde.
- Im Feld ASH1 (Abstract-Herkunft 1) ist vermerkt, dass bereits vor der inhaltlichen Erschließung eine Zusammenfassung der Publikation vorlag, die von AUTINDEX zur Generierung von Deskriptorvorschlägen benutzt wurde. Typische Einträge sind etwa „Zeitschrift“, „Buch“, „Verlag“ oder „Autor“.
- Im Feld ASH2 (Abstract-Herkunft 2) ist das Kürzel eines ZPID-internen Dokumentars eingetragen.
- Für die Neuerscheinung sind sowohl automatisch generierte Deskriptorvorschläge als auch vom ZPID-Dokumentar vergebene Deskriptoren vorhanden.

In der Gesamtstichprobe wurden alle für PSYNDEX ausgewerteten Dokumentarten (Zeitschriftenbeiträge, Bücher und Sammelwerke, Sammelwerksbeiträge, Dissertationen, Institutsberichte, Audiovisuelle Medien) berücksichtigt, die von neun verschiedenen Dokumentaren mit Studienabschluss im Hauptfach Psychologie inhaltlich erschlossen wurden. Die Dokumentationserfahrung variierte zwischen 2 und 32 Jahren und betrug im Durchschnitt 17 Jahre und 7 Monate. Ein Dokumentar war in der Regel für die inhaltliche Erschließung von ein bis zwei Literaturgattungen zuständig.

Bevor die Ergebnisse zur Indexierungskonsistenz dargestellt werden, sollen Befunde berichtet werden, die zeigen, ob und ggf. wie sich durch die Einführung der automatischen Indexierung das Indexierungsverhalten der fachwissenschaftlichen Dokumentare verändert hat.

## 3 Unterschiede zwischen der Indexierung ohne und mit AUTINDEX-Unterstützung

### 3.1 Methoden

Um zu überprüfen, ob die Implementierung der Dokumentations-Unterstützung durch AUTINDEX Auswirkungen auf die Zahl und die Verteilung der vergebenen Deskriptoren hatte, wurden für die Publikationsjahre 2002 bis 2005 insgesamt 11.887 Dokumente untersucht, die mit den oben genannten Selektionskriterien für die Dokumente 2006 bis 2010 übereinstimmten. Allerdings stand für die inhaltliche

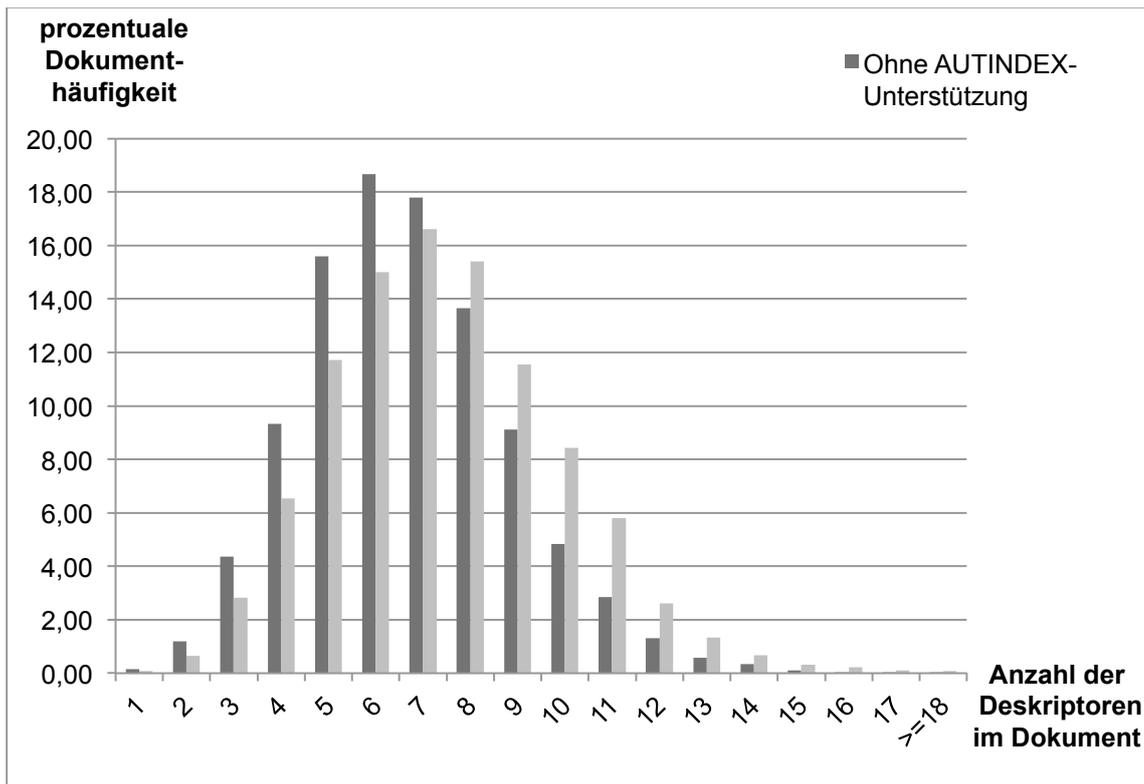


Abbildung 1: Verteilung der prozentualen Häufigkeit von Dokumenten, die ohne bzw. mit AUTINDEX-Unterstützung bearbeitet wurden, für unterschiedliche Anzahlen von Deskriptoren im Dokument (Ausgangsbasis: 11.887 Dokumente der Publikationsjahrgänge 2002–2005 (ohne AUTINDEX verschlagwortet); 13.941 Dokumente der Publikationsjahrgänge 2006–2010 (mit Unterstützung durch AUTINDEX verschlagwortet))

Erschließung dieser Dokumente AUTINDEX noch nicht zur Verfügung. Die Dokumente wurden mit einer Ausnahme von den gleichen Fachdokumentaren verschlagwortet. Ebenso stimmte die Zuordnung zwischen Dokumentar und der von ihm bearbeiteten Literaturgattung weitgehend überein.

### 3.2 Ergebnisse

Bei den 11.877 Dokumenten, die ohne AUTINDEX-Unterstützung bearbeitet wurden, wurden insgesamt 79.653 Deskriptoren vergeben. Davon waren 34.549 gewichtete Haupt-Deskriptoren, die besonders zentrale Inhalte der Publikation abbilden, und 45.104 ungewichtete Neben-Deskriptoren. Für die Gesamtzahl der vergebenen Deskriptoren ergibt sich ein Durchschnitt von  $M = 6.7$  ( $SD = 2.22$ ) Deskriptoren pro Neuerscheinung, mit  $M = 2.9$  ( $SD = 1.18$ ) für die Haupt- und  $M = 3.8$  ( $SD = 2.00$ ) für die Neben-Deskriptoren pro Dokument.

Bei den 13.941 Dokumenten der Publikationsjahrgänge 2006 bis 2010, die mit Unterstützung von AUTINDEX verschlagwortet wurden, wurden 104.146 Deskriptoren vergeben. Davon waren 45.387 Haupt- und 58.759 Neben-Deskriptoren. Der Durchschnitt von  $M = 7.5$  ( $SD = 2.49$ ) Deskriptoren pro Dokument liegt somit leicht über dem Durchschnitt der ohne AUTIN-

DEX-Unterstützung bearbeiteten Dokumente. Dieser hoch signifikante Unterschied spricht dafür, dass die Einführung der automatischen Indexierung dazu geführt hat, etwas mehr Deskriptoren pro Neuerscheinung (knapp ein Schlagwort mehr im Durchschnitt) zu vergeben. Dieser Befund gilt für beide Arten von Deskriptoren: Der Durchschnitt von  $M = 3.3$  ( $SD = 1.27$ ) Haupt- und  $M = 4.2$  ( $SD = 2.26$ ) Neben-Deskriptoren pro Dokument liegt um jeweils 0.4 höher als bei Dokumenten, die ohne AUTINDEX-Unterstützung verschlagwortet wurden.

Dieser Ergebnistrend wird bestätigt, wenn die Verteilung der prozentualen Häufigkeit von Dokumenten für unterschiedliche Anzahlen von Deskriptoren, die in einem Dokument vergeben wurden, betrachtet wird: Die prozentuale Dokumenthäufigkeit liegt für Dokumente, bei denen ein bis sieben Deskriptoren vergeben wurden und die ohne AUTINDEX-Unterstützung bearbeitet wurden, jeweils höher als für die Dokumente, die mit AUTINDEX-Unterstützung bearbeitet wurden. Werden im Dokument acht oder mehr Deskriptoren vergeben, kehrt sich das Bild um: In diesen Kategorien kommen (prozentual) häufiger Dokumente vor, die mit AUTINDEX-Unterstützung bearbeitet wurden, als Dokumente, bei denen AUTINDEX noch nicht zur Verfügung stand (siehe Abbildung 1).

Die Auswirkung der AUTINDEX-Unterstützung auf die durchschnittliche Anzahl vergebener Deskriptoren zeigt sich auch, wenn die einzelnen Literaturgattungen betrachtet werden (siehe Tabelle 1): Bei den Dissertationen, bei Zeitschriftenbeiträgen und bei Monografien kommt es zu einem hoch signifikanten Anstieg der Deskriptoren nach Einführung der teilautomatischen Indexierung. Bei Sammelwerken und Sammelwerksbeiträgen fällt dieser Anstieg geringer, aber immer noch hoch signifikant aus. Lediglich bei den Institutsberichten und bei den audiovisuellen Medien, für die allerdings nur relativ wenige Dokumente verglichen werden können, werden nach der Einführung von AUTINDEX etwas weniger Deskriptoren vergeben.

**Tabelle 1:**  
Durchschnittliche Anzahl vergebener Deskriptoren bei Dokumenten, die ohne bzw. mit AUTINDEX-Unterstützung verschlagwortet wurden, für einzelne Dokumentarten

	Durchschnittliche Anzahl vergebener Deskriptoren bei Dokumenten, die ohne AUTINDEX-Unterstützung verschlagwortet wurden (PY 2002-2005)	Durchschnittliche Anzahl vergebener Deskriptoren bei Dokumenten, die mit AUTINDEX-Unterstützung verschlagwortet wurden (PY 2006-2010)	Ergebnisse t-Test
Alle Dokumentarten	6.7 (SD = 2.22) (11.887 Dokumente)	7.5 (SD = 2.49) (13.941 Dokumente)	t (25764.258547) = 26.295***
Zeitschriftenbeiträge	6.8 (SD = 2.12) (7.850 Dokumente)	7.5 (SD = 2.38) (10.034 Dokumente)	t (17569.751516) = 21.112***
Monografien	6.8 (SD = 2.59) (807 Dokumente)	7.6 (SD = 2.79) (1.170 Dokumente)	t (1811.581426) = 6.470***
Sammelwerke	5.3 (SD = 1.99) (956 Dokumente)	5.8 (SD = 2.21) (796 Dokumente)	t (1618.681350) = 4.556***
Sammelwerksbeiträge	5.9 (SD = 1.88) (1.222 Dokumente)	6.3 (SD = 1.96) (840 Dokumente)	t (2060) = 4.865***
Dissertationen	8 (SD = 2.16) (805 Dokumente)	9.6 (SD = 2.23) (831 Dokumente)	t (1634) = 14.338***
Institutsberichte	7.8 (SD = 2.69) (178 Dokumente)	7.2 (SD = 2.38) (141 Dokumente)	t (317) = 2.281
Audiovisuelle Medien	5.8 (SD = 1.68) (69 Dokumente)	5.5 (SD = 1.83) (129 Dokumente)	t (196) = 1.038

\*\*\*  $p < .001$

### 3.3 Zwischenfazit

Insgesamt deutet der Vergleich der Dokumentationen darauf, dass die Einführung der teilautomatischen Indexierung seit 2006 mit einem leichten Anstieg der pro Dokument vergebenen Deskriptoren einhergeht. Es sei aber darauf verwiesen, dass dieser Befund auch dadurch zustande gekommen oder mitbedingt sein kann, dass den Dokumentaren nach dem *Thesaurus of Psychological Index Terms*<sup>2</sup> bis Ende 2005 insgesamt 5.416 standardisierte psychologische

Fachbegriffe für die Verschlagwortung zur Verfügung standen und zwischen 2006 und 2010 aufgrund der Veränderungen in den Wissenschaften 557 neue Fachbegriffe hinzukamen, für die Dokumentation somit 10.3% mehr standardisierte psychologische Indexterme zur Verfügung standen.

## 4 Übereinstimmung zwischen von AUTINDEX vorgeschlagenen und vom Auswerter vergebenen Deskriptoren (Indexierungskonsistenz)

Ergänzend zur Gesamtzahl der von AUTINDEX vorgeschlagenen bzw. vom Dokumentar vergebenen Deskriptoren wurde zur Untersuchung der Konsistenz zwischen intellektueller und automatischer Indexie-

rung in einer ersten Auswertung das „Übernahmeverhalten“ der Auswerter bezüglich der automatisch vorgeschlagenen Deskriptoren analysiert. Dazu wurde ermittelt, wie häufig von den Dokumentaren während der inhaltlichen Erschließung in der Arbeitsoberfläche STAR-Client das Unterfeld „Übernehmen“ (des Deskriptorvorschlags) ausgefüllt wurde<sup>3</sup>. Durch das Ausfüllen dieses Felds wird im STAR-Client der von AUTINDEX vorgeschlagene Deskriptor in die Liste der vom Dokumentar vergebenen Deskriptoren eingetragen.

2 GALLAGHER TULEVA, Lisa: *Thesaurus of Psychological Index Terms* (11<sup>th</sup> ed.), Washington, DC 2007.

3 <http://www.dipf.de/de/pdf-dokumente/bildungsinformation/fisbildung/tagung2010.pdf/> [9. Juli 2012].

#### 4.1 Methoden

Vorab sei angemerkt, dass diese einleitend beschriebene Auswertungsstrategie der Ergänzung bedarf, da sie aus verschiedenen Gründen fehlerbehaftet ist: Zum einen schreiben die Mitarbeiter häufiger die Deskriptoren selbst in die Erfassungsmaske, füllen das Feld „Übernehmen“ also nicht aus. Gründe dafür können sein, dass es bei vielen AUTINDEX-Vorschlägen zu aufwendig ist, die komplette Liste durchzugehen, dass „naheliegende“ bekannte Deskriptoren manuell eingetragen werden, bevor die AUTINDEX-Liste durchgeschaut wird und dass AUTINDEX-Vorschläge übersehen werden. Zum anderen verwenden einige Mitarbeiter nicht die STAR-Client-Erfassung, sondern arbeiten mit einer WEB-Erfassungsmaske, in der AUTINDEX-Vorschläge mittels „Drag & Drop“ in die Liste der vom Auswerter vergebenen Deskriptoren „gezogen“ werden. Da eine Häufigkeitsauszählung der übernommenen Deskriptoren in der WEB-Erfassung nicht möglich ist, konnten die Daten dieser Mitarbeiter nicht berücksichtigt werden.

Aus diesen Gründen wird hier zusätzlich eine andere Form der Häufigkeitsauszählung verwendet: Durch eine Datenbanksuche wurde mittels einer „Look-Up“-Prozedur festgestellt, ob ein in dem Feld CTAI

(AUTINDEX-CT-Vorschläge) eingetragener Deskriptor auch als identischer Textstring in dem Feld CT (Controlled Term aus der Liste der vom Auswerter vergebenen Deskriptoren) gefunden wird. Somit wird für jedes Dokument gezählt, wie häufig in den beiden unterschiedlichen Feldern CTAI und CT die gleichen Deskriptoren eingetragen sind. Dies ermöglicht eine vom Übernahmeverhalten des Dokumentars unabhängige Bewertung der Leistung von AUTINDEX: Unter Berücksichtigung der Zahl der vorgeschlagenen und der vergebenen Deskriptoren zeigen hohe Übereinstimmungszahlen der Einträge in CTAI und CT an, dass AUTINDEX häufig Vorschläge generiert, die sich auch in der Liste der vergebenen Deskriptoren wiederfinden. Diese Zählung kann zwar nicht bestimmen, ob die Dokumentare einen Deskriptor wirklich aus der Vorschlagsliste übernommen haben (was in der Regel der Fall sein dürfte), aber sie belegt anhand der Übereinstimmung, inwieweit in der AUTINDEX-Liste für das Dokument passende Deskriptoren vorhanden waren.

Die „Unterstützungsleistung“ der automatischen Indexierung lässt sich quantitativ dabei anhand verschiedener Indikatoren bestimmen: (1) Wie viele der vom fachwissenschaftlichen Dokumentar vergebene-



## ENDLICH EIN KASSENAUTOMAT MIT „EINE SORGE WENIGER“-TECHNOLOGIE.

Immer eine Idee voraus hat Crown einen Kassenautomaten entwickelt, der für Sie arbeitet: den Crown BGT. Perfekt geeignet für öffentliche Einrichtungen und Bibliotheken ermöglicht er Ihren Kunden problemlos das selbstständige Erledigen der Zahlungsvorgänge. Mit individueller Softwareschnittstelle lässt sich der Automat in bestehende IT-Systeme einbinden und übernimmt die Verbuchung. Mit dem Crown BGT brauchen Sie sich um den Zahlungsverkehr keine Sorgen zu machen und haben mehr Zeit für die wichtige Kundenberatung. Mehr unter [www.crown-systems.de](http://www.crown-systems.de)



**crown**  
**systems**

smart ideas that work for you

nen Deskriptoren wurden auch von AUTINDEX vorgeschlagen? (2) Wie viele Deskriptoren, die nicht von AUTINDEX vorgeschlagen wurden, hat der Dokumentar zusätzlich vergeben? (3) Wie viele Deskriptoren hat AUTINDEX vorgeschlagen, die nicht vom Dokumentar übernommen wurden? Die Leistung von AUTINDEX wird somit umso positiver bewertet, je mehr der vergebenen Deskriptoren auch von AUTINDEX vorgeschlagen wurden, je weniger der Dokumentar von AUTINDEX nicht vorgeschlagene Deskriptoren zusätzlich vergeben hat und je weniger AUTINDEX Vorschläge generiert hat, die sich aus Sicht des Dokumentars als unpassend für das Dokument erweisen. Als ein Maß, das diese verschiedenen Indikatoren integriert, wird der *Rolling-Index* (RI) herangezogen, der in der Literatur zur Bestimmung der Indexierungskonsistenz berechnet wird.

Nach Rolling<sup>4</sup> hängt die Indexierungskonsistenz nicht nur von der Anzahl der Übereinstimmungen der Deskriptoren in zwei Indexaten ab (s.o.), sondern auch von der Gesamtzahl der in den zwei Indexaten vorkommenden Deskriptoren. Der Rolling-Index berechnet sich nach der Formel  $RI = 2c/(a+b)$ , wobei c die Anzahl der in beiden Indexaten übereinstimmenden Deskriptoren und a bzw. b die Gesamtzahl der in jedem einzelnen Indexat vorkommenden Deskriptoren ist. Ein optimaler Wert von  $RI = 1$  (bzw. 100 %) wird erreicht, wenn alle Deskriptoren in beiden Indexaten übereinstimmen. Dieser Wert wird umso stärker negativ beeinflusst, je mehr in einem oder in beiden Indexaten nicht übereinstimmende Deskrip-

einander, da bei der inhaltlichen Erschließung dem fachwissenschaftlichen Dokumentar die Ergebnisse der automatischen Indexierung vorlagen und ihn bei seiner Verschlagwortung beeinflussten. Trotzdem soll der Rolling-Index verwendet werden, wobei sich hier hohe Konsistenz zwischen den zwei Indexaten daraus ergibt, dass vom Dokumentar einerseits relativ viele AUTINDEX-Vorschläge übernommen wurden und andererseits er relativ wenige eigene Vorschläge hinzugefügt bzw. AUTINDEX relativ wenige unpassende Vorschläge generiert hat.

#### 4.2 Ergebnisse

Die Auswertung der 13.941 Dokumente der Publikationsjahrgänge 2006 bis 2010, die mit AUTINDEX-Unterstützung verschlagwortet wurden, ergab, dass bei diesen Dokumenten insgesamt 104.146 Deskriptoren vergeben und von AUTINDEX 124.420 Deskriptoren vorgeschlagen wurden. Obwohl – wie der Vergleich der Text-Strings in den Feldern CTAI und CT zeigte – 45.829 vorgeschlagene und vergebene Deskriptoren übereinstimmten, wurde nur bei 15.567 vorgeschlagenen Deskriptoren das Subfield „Übernehmen“ von den Dokumentaren angekreuzt. Die Gründe für dieses „Übernahmeverhalten“ wurden oben erläutert. Auf Grund der großen Diskrepanz werden im Weiteren nur die Daten ausgewertet, die auf dem Vergleich der Text-Strings der vorgeschlagenen und vergebenen Deskriptoren basieren.

Tabelle 2 zeigt die Ergebnisse, die sich im Durchschnitt für ein Dokument ergeben. Ergänzend zu den

**Tabelle 2:**  
Anzahl vergebener und vorgeschlagener Deskriptoren für ein „durchschnittliches“ PSYINDEX-Dokument  
(Ausgangsbasis: 13.941 Dokumente der Publikationsjahrgänge 2006 bis 2010)

Anzahl der vom Dokumentar vergebenen Deskriptoren	7.5 (SD = 2.49)
Anzahl der von AUTINDEX vorgeschlagenen Deskriptoren	8.9 (SD = 6.22)
Anzahl der Deskriptoren, die sowohl vorgeschlagen als auch vergeben wurden	3.3 (SD = 2.03)
Prozentanteil der vergebenen Deskriptoren, die auch vorgeschlagen wurden	44 %
Prozentanteil der vorgeschlagenen Deskriptoren, die <b>nicht</b> vergeben wurden	62.9 %
Rolling Index (RI)	40.1 %

toren vorhanden sind. Zu beachten ist, dass bei der in der Literatur berechneten Indexierungskonsistenz die Übereinstimmungen zwischen zwei unabhängig voneinander tätigen Indexierern ermittelt werden. Dagegen sind in der vorliegenden Studie die beiden zu vergleichenden Indexate nicht unabhängig von-

Durchschnittswerten wird zur Bestimmung der Indexierungskonsistenz der Rolling-Index (RI) berechnet, der die Abhängigkeit der Übereinstimmungszahl von der Gesamtanzahl der angeführten Deskriptoren berücksichtigt. Nach Rolling wird im Zähler die Zahl der Übereinstimmungen von vorgeschlagenen und vergebenen Deskriptoren verdoppelt, im Nenner wird die Summe der vorgeschlagenen und der vergebenen Deskriptoren angeführt und das Ergebnis der Division

<sup>4</sup> ROLLING, L.: Indexing consistency, quality and efficiency, in: Information Processing and Management 17 (1981) p. 69 – 76.

wird mit 100 multipliziert, um Prozentangaben zu ermöglichen.

Im Durchschnitt haben die Dokumentare  $M = 7.5$  Deskriptoren vergeben, von denen – bei einer Übereinstimmung zwischen den vergebenen und automatisch vorgeschlagenen Deskriptoren von  $M = 3.3$  Schlagworten 44 % auch von AUTINDEX vorgeschlagen wurden. AUTINDEX generierte zu den Dokumenten im Schnitt  $M = 8.9$  Deskriptoren, von denen sich allerdings fast 63 % bei den vergebenen Deskriptoren nicht wiederfinden. Der Rolling Index für die Indexierungskonsistenz beträgt 40.1 % und erreicht damit den nahezu gleichen Wert wie bei der AUTINDEX-Evaluation 2006, in der für 63 psychologische Zeitschriftenbeiträge eine Indexierungskonsistenz von  $R_I = 40.2$  % ermittelt wurde.

Eine Berechnung des Rolling-Index für die einzelnen Publikationsjahrgänge 2006 bis 2010 zeigt, dass der Index für 2006 mit 35.5 % etwas niedriger liegt als in den Folgejahrgängen (PY 2007: 41.7 %, PY 2008: 40.8 %, PY 2009: 41.7 %, PY 2010: 41.5 %), in denen sich nach der Phase der Einführung und Anpassung des Systems bzw. des Dokumentationsverhaltens die Werte auf ungefähr gleichem Niveau stabilisieren.

Tabelle 3 differenziert die Ergebnisse der 13.941

ausgewerteten Dokumente für die unterschiedlichen Literaturgattungen. Es zeigt sich, dass – abgesehen von den Institutsberichten und Audiovisuellen Medien, die in der gesamten Stichprobe einen relativ kleinen Anteil haben – über 40 Prozent der vom Dokumentar vergebenen Deskriptoren auch von AUTINDEX vorgeschlagen wurden: Für die Monografien liegt dieser Anteil bei 40.5 %, für die Zeitschriftenbeiträge bei 43.6 %, für die Sammelwerksbeiträge bei 45.6 %, für die Dissertationen bei 45.6 % und für die Sammelwerke bei 53.2 %. Dass der Rolling-Index gerade für Sammelwerke und – in etwas geringerem Maß auch für Monografien – mit 25.6 % bzw. 33.3 % niedriger ausfällt als bei Zeitschriften- (42.5 %) und Sammelwerksbeiträgen (43.3 %) sowie bei Dissertationen (42.6 %) liegt vor allem an der stark unterschiedlichen Indexierungsbreite zwischen intellektueller und automatischer Indexierung und dem hohen Anteil an Deskriptorvorschlägen, die nicht im intellektuellen Indexat berücksichtigt werden: Bei den Sammelwerken werden von AUTINDEX im Durchschnitt über 18 Vorschläge gemacht, von denen aber nur etwas mehr als 3 Vorschläge übernommen und um 3 Vorschläge ergänzt werden, bei den Monografien finden sich von knapp 11 Vorschlägen ebenso nur 3 im in-

# BEWEGTE BÜCHER



Gilgen Logistics entwickelt massgeschneiderte Komplettlösungen mit Eigenprodukten für die In-house Logistik. Kunden aus Dienstleistung, Handel und Industrie schätzen unsere Kompetenz - und dies seit mehr als 35 Jahren.

Was auch immer Sie bewegen wollen - wir steuern, kontrollieren und optimieren Materialflüsse in der gesamten Intralogistik.

### Logistik für Bibliotheken

- Automatisches Buchtransportsystem zwischen Archiven und Ausgabe-/Rückgabestellen mecom®
- Leistungsfähige Mediensortierung
- Buchsorter mit automatischer Rückführung in die Magazin- und Freihandbereiche
- 24-h-Medienausgabe und Rückgabeterminale
- Flexible Zwischenlagerung für Vorreservierungen
- Retrofit und Modernisierung




**www.gilgen.com / info@gilgen.com**  
**Gilgen Logistics AG**, Wangentalstrasse 252, CH-3173 Oberwangen  
 Tel. +41 31 985 35 35, Fax +41 31 985 35 36  
**Gilgen Logistics GmbH**, Haverf 20, D-44227 Dortmund  
 Tel. +49 231 9750 5010, Fax +49 231 9750 5040

	Zeitschriften- beiträge (10.034 Dok.)	Monografien (1.170 Dok.)	Sammelwerke (796 Dok.)	Sammelwerks- beiträge (851 Dok.)	Dissertationen (831 Dok.)	Instituts- berichte (141 Dok.)	Audiovisuelle Medien (129 Dok.)
Anzahl der vergebenen Deskriptoren	7.5 (SD = 2.38)	7.6 (SD = 2.79)	5.8 (SD = 2.21)	6.3 (SD = 1.96)	9.6 (SD = 2.23)	7.2 (SD = 2.38)	5.5 (SD = 1.83)
Anzahl der vor- geschlagenen Deskriptoren	7.9 (SD = 4.36)	10.9 (SD = 7.57)	18.3 (SD = 12.75)	7.0 (SD = 3.79)	12.1 (SD = 6.71)	7.0 (SD = 4.94)	5.6 (SD = 3.53)
Anzahl der über- einstimmenden Deskriptoren	3.3	3.1	3.1	2.9	4.6	2.0	1.8
Prozentanteil der vergebenen Deskriptoren, die auch vorgeschlagen wurden	43.6 %	40.5 %	53.2 %	45.6 %	47.7 %	28.7 %	32.8 %
Prozentanteil der vorgeschla- genen De- skriptoren, die <b>nicht</b> vergeben wurden	58.5 %	71.7 %	83.2 %	58.8 %	61.7 %	70.8 %	67.6 %
Rolling Index (RI)	42.5 %	33.3 %	25.6 %	43.3 %	42.6 %	28.8 %	32.6 %

*Tabelle 3: Anzahl vergebener und vorgeschlagener Deskriptoren für PSYINDEX-Dokumente verschiedener Dokumentararten (Ausgangsbasis: 13.941 Dokumente der Publikationsjahrgänge 2006 bis 2010)*

tellektuellen Indexat, das der Dokumentar um 4,5 eigene Deskriptoren anreicht. Prozentual liegt somit der Anteil von Deskriptorvorschlägen, die nicht vom Dokumentar berücksichtigt wurden, bei den Sammelwerken bei 83.2 % und bei den Monografien bei 71.7 %. Diese Werte liegen deutlich höher als bei den Zeitschriftenbeiträgen (58.5 %), den Sammelwerksbeiträgen (58.8 %) und den Dissertationen (61.7 %). Die Differenzen erklären sich insbesondere aus Unterschieden in den Zusammenfassungen der Neuerscheinungen, die von AUTINDEX für die Generierung der Deskriptorvorschläge herangezogen werden: Bei den Sammelwerken und Monografien werden die jeweiligen Inhaltsverzeichnisse in das Abstract-Feld eingegeben. Diese Inhaltsverzeichnisse weisen in der Regel thematisch eine breite Streuung auf. So werden etwa bei den Sammelwerken die Titel aller einzelnen Beiträge der Autoren, die ein Thema unter verschiedenen Perspektiven beleuchten können, aufgenommen. In diesen Titeln sind häufig zahlreiche Begriffe enthalten, die sich als Indikatoren in AUTINDEX finden und daher als Thesaurusvorschläge qualifizieren. Als Beispiel mag das 2010 von V. Arolt und A. Kersting herausgegebene Sammelwerk „Psychotherapie

in der Psychiatrie. Welche Störung behandelt man wie?“<sup>5</sup> dienen. Hier werden unter anderem im Inhaltsverzeichnis sechs psychotherapeutische Verfahren, zwölf Störungsbilder und acht Behandlungskontexte angeführt, die in den einzelnen Beiträgen vorgestellt und diskutiert werden. AUTINDEX generiert für die unterschiedlichen Verfahren, Störungen und Behandlungskontexte (sowie weitere Aspekte der Publikation) insgesamt 41 Deskriptorvorschläge, von denen der Dokumentar nur zwei Begriffe (Psychotherapie und Psychiatrie) übernimmt und sein Indexat um den Deskriptor „Mental Disorders“ ergänzt (AUTINDEX schlägt den „breiteren“ Begriff „Disorders“ vor). Bei diesem für die inhaltliche Erschließung von Sammelwerken in PSYINDEX üblichen Vorgehen wird in der Regel nur das zentrale Thema (hier: Psychotherapie psychischer Störungen in der Psychiatrie) beschrieben. Von einer Verschlagwortung der Inhalte der einzelnen Beiträge wird abgesehen, da die Sammelwerksbeiträge in PSYINDEX separat erfasst und dann dort adäquat verschlagwortet werden. Im Vergleich zu den

<sup>5</sup> AROLT, Volker / KERSTING, Anette (Hrsg.): Psychotherapie in der Psychiatrie. Welche Störung behandelt man wie? Berlin 2010.

Zusammenfassungen der Sammelwerke und Monografien weisen die Abstracts von Zeitschriften- und Sammelwerksbeiträgen sowie von Dissertationen einen thematisch engeren Fokus auf, was sich vor allem darin zeigt, dass von AUTINDEX weniger „irrelevante“ Deskriptoren vorgeschlagen werden. Wie das Beispiel der Sammelwerke zeigt, heißt „irrelevant“ aber nicht notwendigerweise „falsch“. In der Dokumentationspraxis ist es durchaus angemessen, zu sehr ins Detail gehende Vorschläge durch einen Oberbegriff zusammenzufassen oder – umgekehrt – einen Vorschlag weiter zu spezifizieren. Da in der vorliegenden Studie aber die wörtliche Übereinstimmung der Begriffe in den beiden Indexaten als wesentliches Gütekriterium verwendet wurde, lassen sich derartige Prozesse mit den vorliegenden Daten nicht abbilden. Die 141 ausgewerteten Institutsberichte erreichen mit 28.7 % einen relativ niedrigen Rolling-Index. Nur etwas mehr als 28 % der Deskriptoren finden sich in beiden Indexaten, ein Wert, der deutlich unter den Werten der anderen Literaturgattungen liegt. Dagegen werden über 70 % der Vorschläge von AUTINDEX vom Dokumentar nicht berücksichtigt. Während für dieses Ergebnis keine eindeutige Ursache auszumachen ist, lässt sich der ebenfalls niedrige Rolling-Index von 32.6 % bei den Audiovisuellen Medien vor allem auf Unzulänglichkeiten der AUTINDEX zur Verfügung stehenden Titel- und Abstract-Daten zurückführen. Die Titel der Medien sind häufig wenig aussagekräftig und als Zusammenfassungen werden die meist kurzen, vom Dokumentar im inhaltlichen Erschließungsprozess aufwändig zu überarbeitenden Medienbeschreibungen der Verlage AUTINDEX vorgegeben. Entsprechend generiert AUTINDEX bei den audiovisuellen Medien deutlich weniger Deskriptorvorschläge als bei den Literaturgattungen, von denen dann allerdings auch nur wenige Vorschläge übernommen werden.

## 5 Diskussion

Insgesamt deutet der Vergleich der intellektuellen mit der teilautomatisierten Verschlagwortung psychologischer Fachliteratur zunächst darauf, dass die Einführung der teilautomatischen Indexierung mit einem leichten Anstieg der pro Dokument vergebenen Deskriptoren (um im Durchschnitt zirka einen Deskriptor mehr) einhergegangen ist und damit potentiell zu einer differenzierteren Deskriptorvergabe mit einer etwas erhöhten Feinauflösung beigetragen haben kann. Wie oben ausgeführt, kann dies aber auch dadurch (mit-)bedingt sein, dass den Dokumentaren nach dem *Thesaurus of Psychological Index Terms* durch Fortschritte in der Psychologie und Differenzierungen

ihrer Fachsprache nach der Einführung der teilautomatisierten Indexierung 557 neue, zusätzliche standardisierte Indexterme (plus 10.3 %) zur Verfügung standen. Wird dies aus der (leichten) Zunahme in der Anzahl der Deskriptoren nach der Einführung der Unterstützung durch AUTINDEX herausgerechnet, so bleibt der durch sie gewonnene Zuwachs an Schlagworten marginal.

Gleichwohl zeigen die Befunde unserer empirischen Analysen der Übereinstimmung zwischen der teilautomatisierten und der intellektuellen Verschlagwortung für eine große Anzahl psychologischer Neuerscheinungen eine Übereinstimmung, die den Befunden anderer vergleichbarer Konsistenzstudien sehr gut entspricht<sup>6</sup>. Nach dem Rolling-Index liegt der Gesamtwert der Übereinstimmung für die analysierten über 13.000 psychologischen Neuerscheinungen bei über 40 %. Auffällig ist, dass der Rolling-Index bei Zeitschriften- und Sammelwerksbeiträgen mit 42.5 % bzw. 43.3 % noch höher liegt. Grund dafür könnte sein, dass vor allem psychologische Zeitschriftenbeiträge in aller Regel bereits über vom Autor vergebene Schlagworte verfügen, die von AUTINDEX direkt abgegriffen werden. Da dies zunehmend auch für Sammelwerksbeiträge gilt und diese zudem sehr häufig Überblicks- oder Lehrtext-Charakter mit einer Fokussierung auf ein mehr oder weniger enges Thema (mit relativ klaren Schlagworten) haben, könnte dies auch bei ihnen als Erklärung für den vergleichsweise ebenfalls hohen Übereinstimmungsindex gelten.

Bei den anderen Literaturgattungen liegen seltener von den Autoren selbst vergebene Schlagworte vor, so dass der automatischen Indexierung diese Informationen nicht zur Verfügung stehen, was zu den niedrigeren *RI*-Werten bei Monografien, Institutsberichten und AV-Medien führen kann (26 % <*RI*< 34 %; siehe Tabelle 3). Der niedrigste *RI*-Wert findet sich mit *RI* = 25.6 % für die 840 analysierten Sammelwerke. Da sich aber gerade bei dieser Literaturgattung die höchste prozentuale Übereinstimmung von Deskriptoren in beiden Indexaten findet, erklärt sich der niedrige Rolling-Index durch die sehr hohe Anzahl an Deskriptorvorschlägen, die bei der Verschlagwortung nicht berücksichtigt werden. Dass der Dokumentar den Spezifizierungen, die AUTINDEX bei den Sammelwerken aufgrund der vorliegenden Art der Zu-

<sup>6</sup> GROSS, Thomas / FADEN, Manfred: Automatische Indexierung elektronischer Dokumente an der Deutschen Zentralbibliothek für Wirtschaftswissenschaften, in : Bibliotheksdienst 44 (2010), S. 1120 - 1135.

KEMPF, Andreas Oskar: Automatische Indexierung in der sozialwissenschaftlichen Fachinformation. Eine Evaluationsstudie zur maschinellen Erschließung für die Datenbank SOLIS (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft), Berlin 2012.

sammenfassungen vorschlägt, nicht folgt, ist gut, da dadurch überdifferenzierte Verschlagwortungen (mit zu hoher, artifizierlicher Auflösung) die Nutzer bei der Literaturrecherche unnötig durch nicht oder nur marginal relevante „Treffer“ frustrieren oder zumindest in die Irre und auf Umwege führen, vermieden werden. Im Sinne einer Optimierung der automatischen Indexierung ist zu überlegen, ob bei Sammelwerken die Generierung von Deskriptorvorschlägen durch AUTINDEX ausschließlich auf der Basis der meist aussagekräftigen Titelinformationen und – leider nur sehr selten vorhandenen – Autorenschlagworten erfolgen soll, während das Abstract aus der Ermittlung der Vorschläge ausgeschlossen wird.

Zusammenfassend kann konstatiert werden, dass die von fachwissenschaftlich qualifizierten Dokumentaren vorgenommene intellektuelle Verschlagwortung psychologischer Neuerscheinungen durch eine automatisierte Indexierung sinnvoll unterstützt werden kann. Ergebnis ist ein teilautomatisierter Indexierungsprozess, der vor allem bei solchen Zeitschriften- und Sammelwerksbeiträgen, die selbst schon vor der Indexierung über intellektuell, nämlich vom Autor vergebene Schlagworte verfügt, zu guten Ergebnissen führt. Liegen dagegen keine (oder wenig geeignete) Schlagworte vor, so ist die Übereinstimmung zwischen rein automatisierter und intellektueller Deskriptorvergabe deutlich geringer und genügt kaum den Erfordernissen, die für ihre Verwendung in Literatur- und Informationsrecherchen von Nutzern gegeben sein müssen.

Da in unsere Übereinstimmungsanalysen nur solche Dokumente eingegangen sind, die von fachwissenschaftlich qualifizierten Dokumentaren mit langjähriger Berufserfahrung indexiert wurden, kann am vorliegenden Datensatz der Einfluss der Berufserfahrung und Kompetenz auf die Indexierung und die Übereinstimmung nicht untersucht werden. Zu vermuten ist allerdings, dass unerfahrene Dokumentare mehr AUTINDEX-Vorschläge als erfahrene übernehmen, dass von ihnen auch dann AUTINDEX-Vorschläge („im Zweifelsfall“) übernommen werden, wenn sie nur randständige Themenaspekte der Publikation behandeln und dass ggf. auf notwendige Differenzierung zu globaler AUTINDEX-Vorschläge verzichtet wird. Im schlimmsten Fall wird durch AUTINDEX-Vorschläge

verhindert, dass sich ein Dokumentar mit wenig Berufserfahrung inhaltlich differenziert und vertieft mit der Publikation sowie – allgemeiner – mit dem Thesaurus und dem jeweiligen Forschungsfeld beschäftigt, sich also zu sehr oder vollständig auf die automatische Verschlagwortung verlässt und dadurch selbst keine wirkliche Berufserfahrung mit fachlich fundierten Dokumentations-Kompetenzen erwirbt. Dies mag überdies durch die Bequemlichkeit der Übernahme automatisch angebotener Schlagwort-Listen verstärkt werden. Wenn dies passieren würde, könnten die in der Zukunft in weiter zunehmender Menge zur Verfügung stehenden wissenschaftlichen Fachinformationen in der Tat zu „Informations-Wüsten“ werden, in denen sich Nutzer nicht mehr zurecht finden, also die Orientierung verlieren und womöglich sogar auf die eine oder andere informationelle Fata Morgana reinfallen könnten. Dies kann ggf. aber auch durch aktuelle informationstechnologische und kognitionspsychologische Weiterentwicklungen der bislang vor allem an einzelnen Begriffen ausgerichteten automatisierten Verschlagwortung zu einer auf semantischen Technologien<sup>7</sup> basierenden Informations- und Literaturrecherche verhindert oder zumindest abgemildert werden. |



**Dipl.-Psych. Michael Gerards**

Funktion  
Stellvertretende Leitung  
Dokumentation  
gerards@zpid.de



**Prof. Dr. Günter Krampen**

Leibniz-Zentrum für Psychologische  
Information und Dokumentation  
(ZPID), Trier

<sup>7</sup> DOMS, Andreas/ Schroeder, Michael: Semantic Search with GoPubMed, in: François Bry/ Jan Maluszynski (Eds.): Semantic Techniques for the Web (Lecture Notes in Computer Science. Vol. 5500) Berlin/ Heidelberg 2009, pp. 309-342.  
Lu, Zhiyong: PubMed and beyond: a survey of web tools for searching biomedical literature, in: Database: the journal of biological databases and curation baq036 (2011) <http://www.ncbi.nlm.nih.gov/pmc/articles/MC3025693/?tool=pubmed> [9. Juli 2012].

**Besuchen Sie uns auf der Frankfurter Buchmesse Halle 4.2 L431**

**swets**