Erstellung einer DNB-Retrieval-Testkollektion

Bachelorarbeit

Bibliothekswissenschaften

Fakultät für Informations- und Kommunikationswissenschaften

Technische Hochschule Köln

vorgelegt von:

Johanna Munkelt

am 18.05.2018 bei

Prof. Dr. Philipp Schaer Prof.

Dr. Klaus Lepsky

Technology
Arts Sciences
TH Köln

Abstract

Seit Herbst 2017 findet in der Deutschen Nationalbibliothek die Inhaltserschließung bestimmter Medienwerke rein maschinell statt. Die Qualität dieses Verfahrens, das die Prozessorganisation von Bibliotheken maßgeblich prägen kann, wird unter Fachleuten kontrovers diskutiert. Ihre Standpunkte werden zunächst hinreichend erläutert, ehe die Notwendigkeit einer Qualitätsprüfung des Verfahrens und dessen Grundlagen dargelegt werden. Zentraler Bestandteil einer künftigen Prüfung ist eine Testkollektion. Ihre Erstellung und deren Dokumentation steht im Fokus dieser Arbeit. In diesem Zusammenhang werden auch die Entstehungsgeschichte und Anforderungen an gelungene Testkollektionen behandelt. Abschließend wird ein Retrievaltest durchgeführt, der die Einsatzfähigkeit der erarbeiteten Testkollektion belegt. Seine Ergebnisse dienen ausschließlich der Funktionsüberprüfung. Eine Qualitätsbeurteilung maschineller Inhaltserschließung im Speziellen sowie im Allgemeinen findet nicht statt und ist nicht Ziel der Ausarbeitung.

Since autumn 2017, content indexing for certain groups of media works has been carried out purely by machine in the German National Library. The quality of this technique, which can significantly affect the process organization of libraries, is a controversial subject among experts. Their views will first be adequately explained before the need for a quality review of the technique and its fundamentals is presented. A central component of a retrieval test is a test collection. Their creation and the documenatation is the focus of this work. In this context, the genesis and requirements of successful test collections are also treated. Finally, a retrieval test is realised, which proves the suitability of the developed test collection. Its results are for functional review only. A quality assessment of automatic content indexation in particular as well as in general does not take place and is not the aim of the preparation.

Inhaltsverzeichnis

1. Einleitung

Die Digitalisierung prägt die Arbeitswelt zusehends. Keine Branche, keine Institution, keine Nische bleibt von ihr unberührt. Moderne Technologien formen Arbeitsabläufe neu und verschieben die Grenzen des Möglichen – so auch in deutschen Bibliotheken. Die maschinelle Inhaltserschließung zählt zu den Innovationen, die den Arbeitsalltag dort mehr und mehr verändern. In der Theorie verspricht sie eine Katalogisierungsleistung, die mit Menschen kaum zu erbringen wäre. Das hat auch die Deutschen Nationalbibliothek (DNB) erkannt und setzt zunehmend auf die maschinelle Erschließung von Medienwerken. Das wirft die Frage auf: Kann Technik den Menschen bereits bei dieser Aufgabe ersetzen?

Die Deutsche Nationalbibliothek hat im Jahr 2017 einen großen Schritt in diese Richtung gemacht, der über zwei Papiere[1] angekündigt wurde. Die Ankündigung verlief eher verhalten, der Inhalt der Papiere gestaltet sich dafür umso brisanter: Die Deutsche Nationalbibliothek verzichtet seit dem 01. September 2017 auf eine intellektuelle Sacherschließung der Reihen B und H und stellt für diese Reihen auf ein maschinelles Verfahren um. Die Sacherschließung der Netzpublikationen (Reihe O) ist bereits seit 2010 rein maschinell, ab September 2017 werden auch Monografien und Periodika, die außerhalb des Verlagsbuchhandels erscheinen, und Hochschulschriften nicht mehr von Menschen inhaltlich erschlossen. Ein Zeitungsartikel von Dr. Klaus Ceynowa, dem Generaldirektor der Bayerischen Staatsbibliothek, der Ende Juli 2017 in der Frankfurter Allgemeinen Zeitung erschien[2], bot der Umstellung und den daraus resultierenden Folgen eine große Bühne. Der Artikel löste eine Diskussion über die Entscheidung der Deutschen Nationalbibliothek aus. Vor

---

[1] Deutsche Nationalbibliothek (Hrsg.): Strategische Prioritäten 2017-2020, 2017. http://www.d-nb.de/strategie (15. Mai 2018).;
Ulrike Junger und Ute Schwens: Die inhaltliche Erschließung des schriftlichen kulturellen Erbes auf dem Weg in die Zukunft : automatische Vergabe von Schlagwörtern in der Deutschen Nationalbibliothek, August 2017. http://www.dnb. de/SharedDocs/Downloads/DE/DNB/inhaltserschliessung/automatischeInhaltserschliessung.pdf?__ blob=publicationFile (15. Mai 2018).

[2] Klaus Ceynowa: Deutsche Nationalbibliothek: In Frankfurt lesen jetzt zuerst Maschinen. In: FAZ.NET, 2017. http://www.faz.net/1.5128954 (15. Mai 2015).

allem die Qualität der Inhaltserschließung mittels automatisierter Verfahren wird dabei hinterfragt. Ist die maschinelle Inhaltserschließung den bibliothekarischen Ansprüchen also bereits gewachsen? Zur Überprüfung dieser Fragestellung ist es notwendig, Tests durchzuführen, die die Qualität der inhaltserschließenden Merkmale evaluieren. Erst sie ermöglichen einen Qualitätsvergleich zwischen maschinell und intellektuell vergebenen Schlagworten.

Diese so genannten Retrievaltests sind ein gängiges Mittel, um die Qualität der Inhaltserschließung, Indexierungsstrategien verschiedener Systeme oder andere Fragestellungen im Information Retrieval einordnen zu können. Essenzieller Bestandteil eines Retrievaltests ist die Testkollektion. Eine auf die oben genannte Forschungsfrage abgestimmte und dementsprechend entworfene Testkollektion ist die wichtigste Grundlage für einen erfolgreichen Retrievaltest. Ist es möglich, eine Testkollektion zu konstruieren, die zur Beurteilung der oben gestellten Fragestellung beitragen kann? Welche Eigenschaften muss eine solche Testkollektion besitzen? Bisherige Retrievaltests und die dazugehörigen Testkollektionen dienten der Prüfung anderer Fragestellungen und sind daran angepasst erstellt worden. Es besteht die Notwendigkeit einer speziell auf diese Fragestellung abgestimmten Konstruktion einer neuen Testkollektion.

Diese Bachelorarbeit ist eine Dokumentation der Erstellung einer Testkollektion, die als Grundlage für einen Retrievaltest zur Qualitätsprüfung der inhaltserschließenden Merkmale in Titelaufnahmen aus dem Katalog der Deutschen Nationalbibliothek genutzt werden kann. Mit Beendigung dieser Arbeit soll eine Testkollektion mit allen relevanten Bestandteilen, die für einen erfolgreichen Retrievaltest benötigt werden, vorliegen.

Dazu werden zu Beginn der Arbeit die Hintergründe, die der Notwendigkeit eines Retrievaltests und der Erstellung einer geeigneten Testkollektion dafür zu Grunde liegen, dargelegt. Zusätzlich werden zwei Standpunkte, die sich kritisch und relativierend zu der Entscheidung der Deutschen Nationalbibliothek verhalten, erläutert.
Im Weiteren werden Anforderungen für eine gelungene und einsatzfähige Testkollektion aus der Fachliteratur herausgearbeitet. Diese Anforderungen

werden anschließend auf einen Datensatz aus Daten der DNB übertragen und in Handlungsstrategien übersetzt. Die Durchführung der Handlungsstrategien wird beschrieben, dies entspricht der Erstellung der Testkollektion. Alle Schritte der Konstruktion werden nachvollziehbar dokumentiert.
Zur Prüfung der erfolgreichen Erarbeitung einer einsatzbereiten Testkollektion wird zudem ein beispielhafter Retrievaltest durchgeführt. Anhand der Ergebnisse des Tests wird beurteilt, inwieweit die Testkollektion und ihre Konstruktion gelungen sind. Im Anschluss wird ein Fazit gezogen.

2. Hintergrund zur Entscheidung der DNB

Maßgeblich relevant für die Entstehung dieser Bachelorarbeit ist eine Entscheidung der Deutschen Nationalbibliothek, im Weiteren DNB, zur grundsätzlichen Verfahrensweise ihrer inhaltlichen Erschließung sowie diverse Standpunkte diesbezüglich, die in diesem Kapitel beleuchtet werden.

Im Jahr 2017 veröffentlichte die DNB zwei Papiere, die die Öffentlichkeit über folgende Entscheidung hinsichtlich der inhaltlichen Erschließung in Kenntnis setzten: Die deutsch- und englischsprachigen Medienwerke der Reihen B (Monografien und Periodika außerhalb des Verlagsbuchhandels) und H (Hochschulschriften) werden inhaltlich ab dem 01. September 2017 nicht mehr intellektuell erschlossen, dies solle fortan maschinell geschehen. Bereits seit 2010 erschließt die DNB Netzpublikationen mit Hilfe der Metadaten der Verlage und vergibt mittels maschineller Verfahren inhaltserschließende Merkmale. Wissenschaftliche digitale Publikationen werden auf diese Weise mit vollständigen DDC-Notationen und Schlagwörtern versehen.[3]

Auf einer Webseite der DNB ist nachzulesen, dass für die Reihen B und H vor Inkraftsetzung der neuen Erschließungspraxis „die Auswertung vorhandener digitaler Informationen […] erprobt“[4] wurde, weiter werden hierzu keine Informationen gegeben. Aus den Erprobungen folgt, signalisiert durch das Wort „daher“, die Änderung der Erschließungsstrategie. Ab dem 01. September 2017 wird die „inhaltliche Erschließung der in den Reihen B und H verzeichneten Medienwerke […] auf maschinelle Verfahren umgestellt“[5].
Schlagwörter aus der Gemeinsamen Normdatei (GND) und die Sachgruppen der Dewey-Dezimalklassifikation (DDC) werden weiterhin aufgenommen, die

---

[3] vgl. Junger und Schwens: Die inhaltliche Erschließung des schriftlichen kulturellen Erbes auf dem Weg in die Zukunft : automatische Vergabe von Schlagwörtern in der Deutschen Nationalbibliothek, 2017.

[4] „Deutsche Nationalbibliothek - Inhaltserschließung - Grundzüge und erste Schritte der künftigen inhaltlichen Erschliessung von Publikationen in der Deutschen Nationalbibliothek“, Homepage, Juli 2017. http://www.dnb.de/DE/Erwerbung/Inhaltserschliessung/grundzuegeInhaltserschliessungMai2017.html (15. Mai 2018).

[5] „Deutsche Nationalbibliothek - Inhaltserschließung - Grundzüge und erste Schritte der künftigen inhaltlichen Erschliessung von Publikationen in der Deutschen Nationalbibliothek“, 2017.

„klassifikatorische Tiefenerschließung mit der Dewey-Dezimalklassifikation“[6] jedoch wurde ebenfalls zum 01. September 2017 eingestellt. Die DNB erarbeitet neue DDC-Kurznotationen, die als Ersatz zu der bisher vergebenen vollständigen DDC-Notation eingesetzt werden sollen. Laut der DNB sollen die Daten mit einem „entsprechenden Herkunftskennzeichen“[7] versehen.

In der Frankfurter Allgemeinen Zeitung (FAZ) erschien am 31.07.2017 ein kritischer Artikel von Dr. Klaus Ceynowa, dem Generaldirektor der Bayerischen Staatsbibliothek, der sich mit der geplanten Änderung der inhaltlichen Erschließung der DNB auseinandersetzt. Ebendieser Artikel hat eine Diskussion zur Entscheidung der DNB und der daraus resultierenden Konsequenzen für die Bibliotheken Deutschlands ausgelöst. In den folgenden Unterkapiteln werden sowohl die Beweggründe der DNB dargelegt als auch die Kritik von Klaus Ceynowa erläutert. Heidrun Wiesenmüller, Professorin an der Hochschule der Medien, nimmt in ihrem Blog „Basiswissen RDA“, der sich mit der RDA und dem von ihr verfassten gleichnamigen Lehrbuch befasst, ebenfalls ausführlich Stellung dazu. Ihr Standpunkt zu dem Thema wird im Folgenden zusätzlich beleuchtet.

2.1. Beweggründe der DNB

Die DNB hat mit „Die inhaltliche Erschließung des schriftlichen kulturellen Erbes auf dem Weg in die Zukunft – Automatische Vergabe von Schlagwörtern in der Deutschen Nationalbibliothek“ von Ulrike Junger und Ute Schwens ein Papier veröffentlicht, das sowohl den Hintergrund als auch die geplanten Änderungen im Sacherschließungskonzept erklären soll. Die Autorinnen gehen darin zunächst auf die allgemeinen, mit dem Fortschritt der Technik einhergehenden Änderungen in unserem heutigen Alltag ein und kommen zu dem Schluss, dass in diesem Zusammenhang die „Grundlage vieler Prozesse […] lernende

[6] ebd.

[7] ebd.

Algorithmen [sind]“[8]. Für Bibliotheken seien vielfältige Möglichkeiten entstanden, Erschließungsverfahren zu erneuen. Die Autorinnen geben an, dass die DNB seit einigen Jahren die Perspektiven, die sich aus diesen neuen Möglichkeiten für eine verbesserte Erschließung neuer Medienwerke ergeben, diskutiere. Der Einsatz von automatisierten Verfahren spiele dabei eine Rolle.[9]

Die Inhaltserschließung und ihre Zweckmäßigkeit an sich werden von der DNB nicht in Frage gestellt, ihnen wird nach wie vor ein großer Nutzen zugeschrieben, auch wenn es Argumente für eine kritische Auseinandersetzung mit dem Status der Inhaltserschließung gibt (beispielsweise die Volltextsuche). Über eine Darstellung der bisherigen Praxis der inhaltlichen Erschließung, die vor allem darauf Bezug nimmt, dass die Reihe O (Online-Publikationen) der DNB seit 2010 nicht mehr intellektuell inhaltlich erschlossen wird, gelangen die Autorinnen zu einem der Hauptargumente für die angekündigte Änderung: Belegt mit Zahlen und einer Grafik (s. Abbildung 1) wird darauf aufmerksam

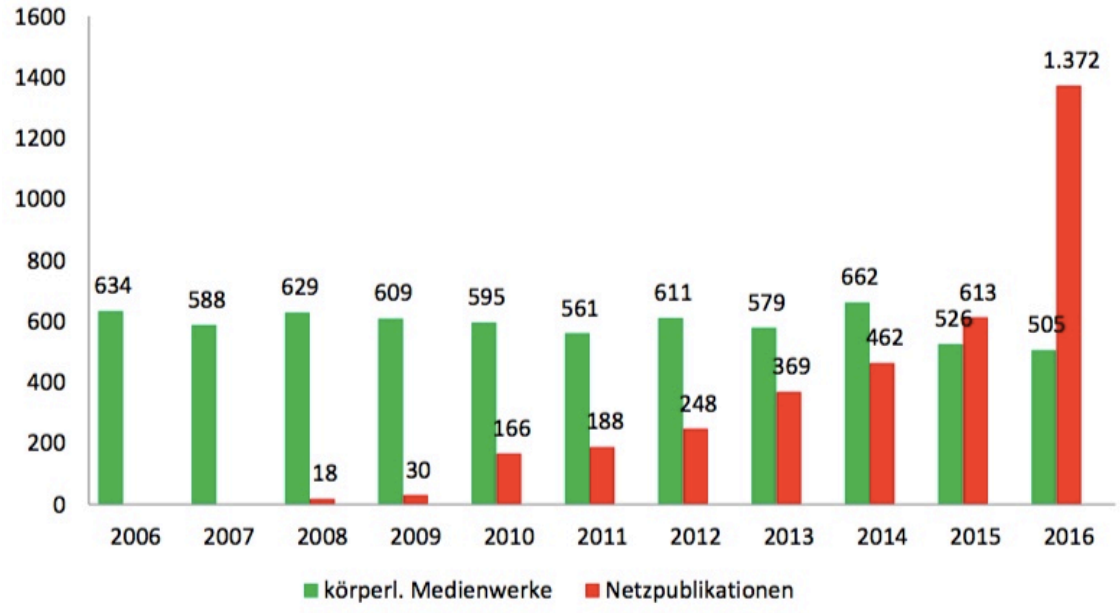

Abbildung 1: Zugang an körperlichen Medienwerken und Netzpublikationen in der DNB seit 2006.
(aus: Junger und Schwens: Die inhaltliche Erschließung des schriftlichen kulturellen Erbes auf dem Weg in die Zukunft : automatische Vergabe von Schlagwörtern in der Deutschen Nationalbibliothek, 2017.)

[8] Junger und Schwens: Die inhaltliche Erschließung des schriftlichen kulturellen Erbes auf dem Weg in die Zukunft : automatische Vergabe von Schlagwörtern in der Deutschen Nationalbibliothek, 2017, S.1

[9] vgl. ebd. S. 1.

gemacht, dass der Zuwachs an Netzpublikationen enorm[10] steigt, die Zugangszahlen zu den gedruckten Medienwerken hingegen gleich bleiben. Hier schwankte der Zugang in den Jahren 2006-2016 zwischen 500.000 und 660.000 Medienwerken. Bedingt durch den steigenden Zugang an Netzpublikationen ergibt sich ein zunehmender Anteil maschinell erschlossener Werk. Der Anteil von intellektuell erschlossenen Werken sinkt im Vergleich dazu. Die DNB spricht konkret von einem Anteil von aktuell 20 Prozent intellektuell erschlossener Medienwerke, wenn man die „Gesamtheit der maschinenlesbar vorliegenden Katalogdaten [betrachtet]“[11].

In der daraus resultierenden Heterogenität im Bereich der Inhaltserschließung sehen die DNB-Autorinnen Junger und Schwens eine Gefahr für den per Gesetz an die DNB ergangenen Auftrag der nationalbibliografischen Verzeichnung. Der starken Arbeitsbelastung durch den Anstieg an Netzpublikationen und dem Nutzerwunsch nach einer inhaltlichen Erschließung aller Medienwerke sei nur durch den Einsatz maschineller Verfahren zu begegnen.[12] Als Ziel legt das DNB-Papier fest, alle Publikationen einheitlich sprich: maschinell zu erschließen.

Bisher solle die maschinelle Erschließung allein in den Sachgruppen angewandt werden, in denen eine bestimmte Qualität der Erschließung gegeben und die Fehlerquote gering ist. Eine konkrete Quote als Richtwert wird nicht genannt. Das Gegenargument der höheren Qualität intellektueller Inhaltserschließung im Einzelfall relativieren die Autorinnen mit dem Hinweis auf den Nutzen einer gleichartigen Erschließung aller Werke. Dies sei für die Nutzerinnen und Nutzer bei der Recherche positiv. Auch führe der erhöhte Anteil erschlossener Medienwerke zu einer Aufwertung der thematischen Suche. Als Bespiel wird hierzu die Recherche nach Conference-Proceedings

---

[10] laut Papier der DNB: in 2016 gingen 1,3 Mio. Publikationen ein, doppelt so viele wie im Vorjahr

[11] Junger und Schwens: Die inhaltliche Erschließung des schriftlichen kulturellen Erbes auf dem Weg in die Zukunft : automatische Vergabe von Schlagwörtern in der Deutschen Nationalbibliothek, 2017, S. 2.

[12] vgl. ebd.

und Zeitschriftenartikeln angeführt, die mit den neuen Verfahren möglich gemacht werde.

Als weiteres Argument für die maschinelle Inhaltserschließung führt das DNB-Papier an, dass der Erschließungsvorgang nicht länger als ein abgeschlossener Prozess gelte, sondern nun die Chance habe, dynamisch zu werden. Sollten sich die maschinellen Erschließungsverfahren qualitativ steigern, könnten bereits erschlossene Medienwerke erneut erschlossen und somit laut DNB „besser zugänglich“[13] gemacht werden. Die auf diese Weise entstehenden Daten sollen „den anderen Bibliotheken und weiteren Daten-Nutzern frei und kostenlos zur Verfügung gestellt [werden]“[14].

Intellektuell erschlossen werden sollen darüber hinaus die Neuzugänge der Fachgebiete, in denen das maschinelle Erschließungsverfahren noch keine ausreichenden Ergebnisse erzielt. Von großer Bedeutung wird die GND bleiben, die für intellektuelle und maschinelle Inhaltserschließung eingesetzt wird. Ein neuer Aufgabenbereich innerhalb der DNB wird die Aktualisierung dieser Normdatei sein, denn nur ein aktuelles Vokabular kann eine gute Grundlage für die automatische Sacherschließung sein. Grundsätzlich werden sich die Arbeitsbereiche sich zum Qualitätsmanagement und zur Normdatenpflege orientieren, die Einzelfallbearbeitung im Bereich Inhaltserschließung wird in Zukunft abnehmen.[15] Ein Artikel der DNB-eigenen Zeitschrift „Dialog mit Bibliotheken“ aus dem Jahr 2010 enthält bereits eine ähnliche Aussage. Die Passage „Diesen gestiegenen Anforderungen stehen nicht im gleichen Maße gestiegene Personalressourcen gegenüber“[16] bezieht sich hier auf den gestiegenen Anteil der Netz-

---

[13] Junger und Schwens: Die inhaltliche Erschließung des schriftlichen kulturellen Erbes auf dem Weg in die Zukunft : automatische Vergabe von Schlagwörtern in der Deutschen Nationalbibliothek, 2017, S. 4.

[14] ebd.

[15] ebd.

[16] Renate Gömpel, Ulrike Junger und Elisabeth Niggemann: Veränderungen im Erschließungskonzept der Deutschen Nationalbibliothek. In: Dialog mit Bibliotheken, 2010, S. 20.

publikationen und den Anspruch, auch Zeitschriftenartikel oder einzelne Songs einer CD (auf „Einzeldokumentebene“[17]) zu erschließen.

Zusammengefasst lässt sich sagen, dass die DNB die Inhaltserschließung per maschinellem Verfahren auf alle Medienwerke ausweiten möchte oder dies teilweise seit Herbst 2017 auch schon tut. Die maschinelle Erschließung soll die intellektuelle Erschließung in allen Fachgebieten ablösen, sobald die Verfahren dazu geeignet sind. Die Umstellung trägt somit zu einer homogenen Erschließung und einer dadurch verbesserten thematischen Recherche bei. Durch den Wegfall eines Großteils der intellektuellen Erschließung sollen Arbeitsstellen entlastet und das Qualitätsmanagement der DNB im Bereich Inhaltserschließung und Normdatenpflege gestärkt werden. Auch die Tatsache, dass Katalogisate mit neuen Erschließungsdaten angereichert werden können, wenn sich durch den Einsatz besserer Techniken qualitative Unterschiede abzeichnen, bewertet die DNB positiv.

2.2. Standpunkt Dr. Klaus Ceynowa

Herr Ceynowa hat mit seinem Artikel, der am 31.07.2017 in der FAZ erschien, die Diskussion über bibliothekarische Erschließung in einer überregional erscheinenden Tageszeitung angeregt. „In Frankfurt lesen jetzt zuerst Maschinen“ lautet der Titel des Artikels, darauf folgt die kritische Sichtweise Herrn Ceynowas auf die geplanten Änderungen, die die DNB im vorherigen Teilkapitel erläuterten Papier veröffentlicht hat.

Die Titelunterschrift ist zugleich ein scharfer Angriff von Herrn Ceynowa auf die DNB: sie mache mit ihrer Interpretation der „Digitalisierung Wissen unzugänglich“[18]. Zu Beginn seines Artikels verweist der Autor auf das Gesetz von 2006, das den Sammlungsumfang der DNB festlegt. Daraus schließt Ceynowa, dass die DNB im Gegensatz zu anderen Bibliotheken mit

[17] Gömpel, Junger und Niggemann: Veränderungen im Erschließungskonzept der Deutschen Nationalbibliothek, 2010, S. 1.

[18] Ceynowa: Deutsche Nationalbibliothek: In Frankfurt lesen jetzt zuerst Maschinen, 2017.

Sammelauftrag – z.B. Forschungsbibliotheken – nicht darüber entscheiden kann, was sie sammelt: „Die DNB sammelt nicht, sie sammelt an.“[19] Doch die Sammlung allein sei nicht die Hauptaufgabe, vielmehr ginge es um eine sachdienliche formale und sachliche Erschließung aller Zugänge, die Wahrnehmung der Pflicht als Herausgeber der Nationalbibliografie. Anhand einer beispielhaft angeführten Verschlagwortung eines Titels wird der Nutzen einer solchen inhaltlichen Erschließung markiert, die in dem Beispiel intellektuell erfolgte. Laut Herrn Ceynowa sei vor allem durch die Digitalisierung die Bedeutung dieser inhaltlichen Merkmale gestiegen. Der Autor weist ferner darauf hin, dass die Bayerische Staatsbibliothek und andere große Bibliotheken internationale wissenschaftliche Fachliteratur erschließen, was nicht in den Aufgabenbereich der DNB falle.[20]

Herr Ceynowa erklärt kurz die geplanten Änderungen, die die DNB an ihrem Sacherschließungskonzept durchsetzen will. Eines seiner Hauptargumente gegen die geplante Umstellung ist die Qualität der maschinellen Inhaltserschließung. Diese sei ihm zu wenig konstant und nicht in allen Fachgebieten angemessen:

> Man muss sich klarmachen, was hier eigentlich geschieht: Es ist nicht irgendeine deutsche Bibliothek, die sich vom als lästig und ressourcenintensiv empfundenen Geschäft der Vergabe von Sachschlagwörtern entlasten will, sondern das nationalbibliographische Zentrum Deutschlands, das seiner Kernaufgabe einer hochqualitativen Inhaltserschließung offenbar überdrüssig geworden ist.[21]

Herr Ceynowa wirft der DNB vor, ihre festgelegte Hauptaufgabe nicht mehr zu erfüllen bzw. nicht weiter erfüllen zu wollen. Des Weiteren ist der Autor der Meinung, dass ein wachsender Anteil an Netzpublikationen kein Argument für die maschinelle Inhaltserschließung sein könne, da diese Publikationen bereits maschinell erschlossen würden. Er vertritt die Ansicht, dass ein konstant bleibender Anteil an gedruckten Medienwerken weiterhin von Menschen intellektuell erschlossen werden soll, da der Arbeitsaufwand nicht zunimmt. Die damit einhergehende Heterogenität in der Inhaltserschließung, die es nahezu in

---

[19] ebd.

[20] vgl. ebd.

[21] ebd.

jedem Bibliothekskatalog geben soll, sei für Bibliotheksbenutzer bisher kein Hindernis gewesen. Es wird herausgestellt, dass eine tiefe inhaltliche Erschließung nicht für jedes Medienwerk von großem Nutzen ist, beispielsweise bei Reiseführern.[22]

Es ergeht der Vorwurf an die DNB, in der Zukunft alle Medienwerke „gleichmäßig auf niedrigem Niveau“[23] zu erschließen und damit einhergehend einen Kompromiss in Kauf zu nehmen, der zwar die inhaltliche Erschließung aller Zugänge der DNB möglich mache, bei dem jedoch große qualitative Einbußen zu erwarten seien. Erneut macht Herr Ceynowa darauf aufmerksam, dass die DNB sich diese Strategie im Gegensatz zu jeder anderen deutschen Bibliothek aufgrund ihres Sammlungsauftrags nicht erlauben dürfe.

Vor allem die Darstellung der Inhaltserschließung als dynamischer Prozess wirft bei Herrn Ceynowa zusätzlich Fragen und Probleme auf: Zum einen sei die technische Realisierung aufwändig und komplex, zum anderen bräuchte er als Generaldirektor der Bayerischen Staatsbibliothek mehrere zusätzliche Arbeitskräfte, die eine Inhaltserschließung für die Neuzugänge mit dem gewohnten Standard gewährleisten könnten[24]. Es entsteht der Eindruck, Herr Ceynowa fühle sich und seine Bibliothek von der DNB und ihrer Rolle als zuverlässiger Lieferant für Katalogdaten von hoher Qualität im Stich gelassen. Er stellt die DNB in seinem Artikel als eine Organisation dar, die ihre Aufgaben auf andere Bibliotheken umverteilt, ohne dass es eine Diskussion mit allen Beteiligten gab, die von dieser Entscheidung betroffen sind.

Der Standpunkt von Herrn Ceynowa lässt sich als offene Kritik an der DNB zusammenfassen, die er öffentlichkeitswirksam platziert. Herr Ceynowa macht darauf aufmerksam, dass die Bibliothek in Deutschland, die seiner Meinung nach die meiste Verantwortung in Bezug auf eine lückenlose inhaltliche Erschließung trägt, sich von einer ihrer Hauptaufgaben bzw. dem

---

[22] vgl. ebd.

[23] ebd.

[24] vgl. ebd.

ursprünglichen Arbeitsumfang dieser Aufgabe distanzieren möchte und duldet, dass die Qualität des geplanten neues Erschließungsverfahren für viele Medienwerke abnimmt. Das DNB-Argument einer in Zukunft ausgedehnteren Medienerschließung kontert er mit der Aussage, dass dies nur möglich sei, indem die Erschließung auf einem allgemein niedrigeren Niveau vollzogen werde. Dies bedeute einen niedrigeren Sacherschließungsstandard als bisher, unter dem auch die von der DNB mit Daten belieferten Bibliotheken leiden würden.

2.3. Standpunkt Prof. Heidrun Wiesenmüller

Die Autorin des Lehrbuchs „Basiswissen RDA“ Heidrun Wiesenmüller, Professorin für Bibliotheks- und Informationsmanagement an der Stuttgarter Hochschule der Medien, betreibt einen Blog, der sich mit ihrem Lehrbuch und den aktuellen Entwicklungen zum internationalen Katalogisierungsstandard Resource Description and Access (RDA) befasst. Der Blog thematisiert auch andere aktuelle Entwicklungen in der Bibliothekswelt, mehrheitlich aus Deutschland. Zwei Tage nach der Veröffentlichung des Artikels von Herrn Ceynowa in der FAZ erschien eine Replik von Frau Wiesenmüller in ihrem Blog, die zu der Kritik von Herrn Ceynowa und den anstehenden Veränderungen in der DNB-Sacherschließungspraxis ausführlich Stellung nimmt.

Beispielsweise Herrn Ceynowas Argument der Qualitätseinbußen bei maschinell vergebenen Schlagworten kommt Frau Wiesenmüller mit ihrer Einschätzung der Situation entgegen: Sie führt dazu Ergebnisse einer Studie von Sandro Uhlmann an, für die 2013 zu mehreren Sachgruppen die maschinell vergebenen Schlagworte auf ihre Richtigkeit geprüft wurden[25]. In der Studie wurde herausgearbeitet, dass die vergebenen Schlagwörter nur teilweise nützlich seien und oftmals einige relevante Schlagwörter fehlen würden. Frau Wiesenmüller fordert einen gemeinsam gefundenen Konsens zu der

[25] vgl. Sandro Uhlmann: Automatische Beschlagwortung von deutschsprachigen Netzpublikationen mit dem Vokabular der Gemeinsamen Normdatei (GND). In: Dialog mit Bibliotheken, 2013, S. 26–36.

Fragestellung, wie fehlertolerant die maschinelle Inhaltserschließung betrachtet werden dürfe. Gäbe es einen bestimmten Grenzwert für Fehlerquoten, wäre es simpler, zu beurteilen, ob die maschinellen Schlagworte als akzeptable Ergebnisse gewertet werden können oder nicht.[26]

Herrn Ceynowas Aussagen zur angestrebten Homogenität in der inhaltlichen Erschließung sowohl gedruckter als auch online vorliegender Medienwerke begegnet Frau Wiesenmüller mit Verständnis für den Ansatz der DNB. Durch diese angestrebte Homogenität könne die qualitative Abgrenzung von gedruckten Publikationen zu digitalen Publikationen, die sich vor allem durch die Tiefe der Erschließung ausdrücke, abgeschafft werden. Wie Herr Caynowa empfindet auch Frau Wiesenmüller den Ansatz der DNB jedoch dahingehend als ungeeignet, als dass er als Konsequenz eine Erschließung aller Publikationen auf einem gleich schlechten Niveau beinhalte. Sie wünscht sich eine ausgewählte Menge gedruckter und digitaler Medienwerke, die für eine qualitative intellektuelle Inhaltserschließung angebracht ist. Dazu sie ein Kriterienkatalog nötig, der dann entwickelt werden müsse.[27]

Frau Wiesenmüller relativiert zwar mit einem Verweis auf die komplexe Personalsituation in der DNB, positioniert sich aber auch sogleich auf der Kritikerseite mit folgender Aussage über die maschinelle Inhaltserschließung:

> Dies ist fraglos der radikalste Weg: Er lässt die höchsten Personaleinsparungen erhoffen, birgt aber auch die größten Risiken für die Datenqualität.[28]

Die Autorin hält ein maschinell unterstütztes Verfahren in der Inhaltserschließung für einen guten Kompromiss zwischen Personalabbau und Erhaltung der Schlagwortqualität. Darüber hinaus fordert sie die verstärkte Nutzung von Fremddaten aus den Bibliotheksverbünden, die eine Arbeitsersparnis berge, in der aktuellen Praxis der DNB aber keine Beachtung finde.[29]

---

[26] vgl. Heidrun Wiesenmüller: Das neue Sacherschließungskonzept der DNB in der FAZ. In: Basiswissen RDA, 2017. http://www.basiswissen-rda.de/neues-sacherschliessungskonzept-faz/ (15. Mai 2015).

[27] vgl. Wiesenmüller: Das neue Sacherschließungskonzept der DNB in der FAZ, 2017.

[28] ebd.

[29] vgl. ebd.

Herrn Ceynowas Kritik der Arbeitsverlagerung in die Bibliotheken, die Daten der DNB beziehen, stimmt Frau Wiesenmüller zu. Sie skizziert für die Zukunft ein Szenario, in dem Bibliotheken die Sacherschließungsdaten der DNB nacharbeiten müssten, um eine gewohnt gute Inhaltserschließung für ihre Nutzer zu leisten, was wiederum zu mehr Personalbedarf führe.
In unterschiedlichen Bibliotheksverbünden sei möglicherweise zu erwarten, dass die gelieferten Daten auf unterschiedliche Weise überarbeitet würden, als Folge könne eine sinkende Datenqualität auftreten. Die geplante Gestaltung der Inhaltserschließung als zyklischen Prozess kann Frau Wiesenmüller nicht nachvollziehen, da ihrer Ansicht nach die technischen Voraussetzungen für die Verarbeitung der zu erwartenden umfangreichen Datenmengen nicht gegeben seien. Zu erwarten sei auch die Problemstellung, dass bereits von Bibliotheken überarbeitete DNB-Daten durch neue Daten der DNB ausgetauscht werden könnten. Hier fordert Frau Wiesenmüller den Kontakt zwischen den Bibliotheken und der DNB, alle von der geplanten Änderung betroffenen Institutionen müssten gemeinsam darüber diskutieren, Regeln konzipieren und Voraussetzungen zur Umsetzung schaffen.[30]

Der Blogeintrag von Frau Wiesenmüller entschärft die breite Kritik von Herrn Ceynowa in einigen Punkten und präsentiert Vorschläge zur Verbesserung der Situation. Die Autorin stellt sich mit dem Eintrag in Summe aber auf die Seite der Kritiker und drückt vor allem aus, dass die DNB alle Betroffenen über ihre Entscheidung informiere, es aber an einer der Möglichkeit mangele, die Sachlage mit der DNB zu diskutieren. Der Blogeintrag verweist auf mehrere Themenbereiche, die nach Frau Wiesenmüller einer breiteren Diskussion mit mehr Teilnehmern bedürfen, wie z.B. die Erschließung als zyklisches Verfahren, die Nutzung von Fremddaten oder Richtwerte für die Qualität von maschineller Sacherschließung. Darüber hinaus sei eine Diskussion über eine alternative Strategie in ihren Augen angebracht.

[30] vgl. ebd.

3. Testkollektionen in der Fachliteratur

In diesem Kapitel wird zunächst die Konzeption einer Testkollektion, wie sie beispielsweise in den Retrievaltests der Text Retrieval Conference (TREC) genutzt wird, vorgestellt und anhand von Fachliteratur mit Merkmalen ausgestattet. Jedes Teilkapitel befasst sich mit einem der drei Bausteine einer Testkollektion. Im letzten Teilkapitel werden Beispiele für erfolgreich aufgebaute Testkollektionen vorgestellt.

Die Konferenzreihe TREC wurde 1992 vom National Institute of Standards and Technology (NIST, Nationales Institut für Standards und Technologie) als Teil des TIPSTER Textprogramms der Defense Advanced Research Projects Agency (DARPA, Behörde für Forschungsprojekte der Verteidigung vom US-Verteidigungsministerium) gegründet. Das TIPSTER Textprogramm begann im Jahr 1991 und setzte sich mit der Verbesserung von Information Retrieval für die amerikanischen Behörden auseinander. Zu diesem Zweck arbeiteten Wissenschaftler und Entwickler aus Regierung, Industrie und der akademischen Welt[31] zusammen. Während das TIPSTER Textprogramm 1998 beendet wurde, existiert die TREC-Konferenzreihe bis heute und tagt einmal jährlich. Ursprünglich sollte TREC eine möglichst große Testkollektion mit über einer Millionen Volltext-Dokumenten[32] für Retrievaltests im Information Retrieval bereitstellen.
Heute sind die Aufgaben breiter gestreut: TREC fördert die Forschung im Bereich Retrievaltest/Testkollektion, ist ein offenes Austauschforum für Regierung, Industrie und Wissenschaft, verbessert Evaluationsmethoden für Industrie und Wissenschaft und beschleunigt die Umsetzung von theoretischen Ideen in die Praxis durch die Bereitstellung einer realistischen Testumgebung für Produkte[33].

---

[31] vgl. „NIST - TIPSTER Text Program - Overview“, Homepage. https://www-nlpir.nist.gov/related_projects/tipster/ (15. Mai 2018).

[32] vgl. Ellen M. Voorhees und Donna K. Harman: TREC Experiment and Evaluation in Information Retrieval, 2005, S. 5.

[33] vgl. „Text REtrieval Conference (TREC) Overview“, Homepage. https://trec.nist.gov/overview-.html (15. Mai 2018).

TREC setzt seit vielen Jahren Maßstäbe für die Durchführung von Retrievaltests. In einem Artikel von Ellen M. Vorhees, einer der Projektmanagerinnen der TREC-Konferenzreihe, aus dem Jahr 2007 spricht die Autorin von einer Verdopplung der Retrievaleffektivität seit dem Start von TREC[34].
Vor allem der sogenannte „ad hoc task“ repräsentiert einen Standard-Retrievaltest wie er beispielsweise in einer Bibliothek von Nutzen sein könnte. Die Ad-Hoc-Aufgabe untersucht das Retrieval eines Systems, über welches unterschiedliche Suchanfragen an eine unveränderliche Dokumentenkollektion formuliert werden.[35] Diese Aufgabe ist vergleichbar mit der Situation eines Nutzers, der ein Informationsbedürfnis befriedigen möchte und eine Suchanfrage an die Datenbank einer Bibliothek stellt. Einige der im Folgenden formulierten Richtlinien zur Erstellung von Testkollektionen beziehen sich daher auf die Gestaltung der Testkollektionen von TREC, insbesondere der ad hoc-Testkollektionen.

Für einen Standard-Retrievaltest und seine erfolgreiche Durchführung sind drei Grundbausteine von großer Bedeutung. Die benötigte Testkollektion besteht aus

- einer Dokumentenkollektion,
- reellen Informationsbedürfnissen, die mit Hilfe von sogenannten Topics[36] mit einer kurzen Beschreibung (Description) und einer genaueren Erläuterung (Narrative) imitiert werden,
- Relevanzurteilen, die für die Ergebnismengen aus der Kollektion zu den jeweiligen Topics gefällt werden.[37]

Wenn diese drei Bedingungen erfüllt sind, entsteht aus einer Dokumentenkollektion eine Testkollektion, mit der Retrievaltests durchgeführt werden können.

---

[34] vgl. Ellen M. Voorhees: TREC: Continuing Information Retrieval’s Tradition of Experimentation. In: Communications of the ACM, 2007, S. 51–54, hier: S. 2.

[35] vgl. Voorhees und Harman: TREC Experiment and Evaluation in Information Retrieval, 2005, S. 80.

[36] vgl. Voorhees und Harman: TREC Experiment and Evaluation in Information Retrieval, 2005, S. 23.

[37] vgl. Christopher D. Manning, Prabhakar Raghavan und Hinrich Schütze: Introduction to Information Retrieval, 2008, S.140.

Die folgenden drei Unterkapitel beschäftigen sich mit je einer der Bedingungen und den Richtlinien, die dazu in der Fachliteratur zu finden sind.

3.1. Richtlinien für die Dokumentenkollektion

Eine Testkollektion besteht nahezu immer aus einer Teilkollektion der Dokumentenkollektion, im Weiteren Ausgangskollektion, die es mit dem Retrievaltest auf zuvor aufgestellte Thesen zu prüfen gilt. Diese Teilkollektion sollte die Ausgangskollektion bestmöglich repräsentieren.[38]
Je nach Datenbank muss eine Dokumentenkollektion, auch Korpus genannt, also unter anderem eine bestimmte Größe haben, damit sie als valide Datenbasis gewertet werden kann. Auch die Tatsache, dass Ergebnismengen für unterschiedliche Informationsbedürfnisse (Topics) stark in ihrer Größe und Struktur variieren können[39], spricht für einen möglichst großen Korpus. Je größer der Korpus ist, desto größer ist die Wahrscheinlichkeit, dass für ein Topic Dokumente gefunden werden. Wenn nicht die komplette Ausgangskollektion genutzt werden kann, weil sie zu groß ist, muss ein ausreichend großer Teil der Dokumente ausgewählt werden.

In einem Artikel von Paul Clough und Mark Sanderson, der 2013 in der Fachzeitschrift Information Research erschien, wird darauf verwiesen, dass sich die inhaltliche Struktur des Korpus' an der Struktur der Ausgangskollektion orientieren sollte[40], dazu geeignet sind zufällige Zusammenstellungen, die sich über das gesamte Sachgebietsspektrum der Ausgangskollektion erstrecken. Ein thematisch abgestecktes Gebiet als Grundlage für die Testkollektion kann aber auch von Nutzen sein, wenn die Fragestellung des Retrievaltests dies hergibt.

---

[38] vgl. Paul Clough und Mark Sanderson: Evaluating the Performance of Information Retrieval Systems Using Test Collections. In: Information Research, 2013. http://www.informationr.net/ir/18-2/paper582.html#.U2unLPldXTp (15. Mai 2018).

[39] vgl. Manning, Raghavan und Schütze: Introduction to Information Retrieval, 2008, S.140.

[40] vgl. Clough und Sanderson: Evaluating the Performance of Information Retrieval Systems Using Test Collections, 2013.

Für die Zusammenstellung der Dokumente kann der Retrievaltest und die damit einhergehende Fragestellung im Hinblick auf einen weiteren Gesichtspunkt von Bedeutung sein: Soll die Testkollektion nach Durchführung des Tests weiterhin nutzbar sein, ist es zu empfehlen, die zugrunde liegenden Dokumente alle statisch zu hinterlegen[41]. Die Ergebnisse verschiedener Retrievaltests sind nur dann miteinander vergleichbar, wenn der Korpus stets unverändert bleibt. Ebenfalls von Relevanz ist die Urheberrechtsfrage im Fall einer in anderen Kontexten weiterverwendbaren Kollektion. Hier muss entschieden werden, ob Dokumente im Volltext vorliegen oder allein die entsprechenden Metadaten genutzt werden.

3.2. Richtlinien für die Topics

Topics für einen Retrievaltest sind die Verschriftlichung eines Informationsbedürfnisses wie es ein Nutzer in der Realität formulieren könnte. Diese imitierten Informationsbedürfnisse haben in der Regel vier Bestandteile: die Topic-Nummer, der Titel, eine Beschreibung und eine Schilderung dessen, was relevant ist, im Weiteren Description und Narrative. Die Topic-Nummer ordnet jedem Topic einen eindeutigen Identifikator zu, der vor allem für die technische Durchführung der Relevanzurteile von Bedeutung ist. Der Titel benennt das Topic in einem, zwei oder sehr wenigen Worten, z.B. „Zimmerpflanzen“. Die Description formuliert das Informationsbedürfnis in einem Satz, z.B. „Finde Dokumente, die sich mit Zimmerpflanzen befassen.“ Im Narrative wird festgelegt, welche Dokumente relevant, welche nur teilweise relevant und welche nicht relevant für das Topic sind, z.B. „Relevante Dokumente beschreiben die Artenvielfalt, Aufzucht und Pflege verschiedener Zimmerpflanzen und geben Anleitungen für ihr langes Bestehen. Teilweise relevante Dokumente handeln von einer einzelnen Zimmerpflanze (z.B. Orchidee). Nicht relevante Dokumente beschäftigen sich mit Gartengestaltung, Gartenpflanzen oder ähnlichem.“[42]

---

[41] vgl. ebd.

[42] vgl. ebd.

Topics sind folglich keine Suchanfragen. Sie beschreiben allein ein Informationsbedürfnis, das dann in eine passende Suchanfrage übersetzt werden muss.
Für die Ad-Hoc-Aufgabe bei TREC werden Topics und die Relevanzurteile zu dem jeweiligen Topic von derselben Person geschaffen. Aus einer großen Auswahl von eingereichten Topics werden von NIST 50 Topics ausgewählt, die in dem Test verwendet werden.[43]
In Christopher D. Mannings „Introduction to Information Retrieval“ werden ebenfalls 50 Topics als Minimum für einen validen Retrievaltest festgelegt.[44] Sanderson und Clough weisen in ihrem Artikel daraufhin, dass das Sachgebietsspektrum der Topics möglichst breit gefächert sein sollte. Ebenso empfehlen sie, bei der Formulierung der Suchanfragen zu variieren und selbige nicht allzu gleichförmig zu gestalten. Die Autoren des Artikels weisen darauf hin, dass die Topics so realistisch wie möglich sein sollen, wenn als Grundlage dafür beispielsweise Aufzeichnungen von Suchanfragen, die bereits an die Ausgangskollektion gestellt wurden, zu Rate gezogen werden.[45]

3.3. Richtlinien für die Relevanzurteile

Der dritte Bestandteil einer Testkollektion sind die Relevanzurteile, die für jedes Topic gefällt werden müssen. Hierbei wird festgelegt, welches Dokument relevant für das dargestellte Informationsbedürfnis ist. Relevanzurteile stellen den strittigsten und arbeitsintensivsten Part bei der Erstellung einer Testkollektion für einen Retrievaltest dar und es gibt verschiedene Ansätze, ein möglichst valides Ergebnis zu produzieren.
Die ideale Grundlange für einen Retrievaltest ist vergleichbar mit dem Aufbau einer der ersten Testkollektionen, der Cranfield-Kollektion. Diese Kollektion wurde von Cyril Cleverdon um 1960 erstellt und besteht aus 1398 Dokumenten (Abstracts von Zeitschriftenartikeln aus dem Fachgebiet Aeronautik) und 225

---

[43] vgl. Voorhees und Harman: TREC Experiment and Evaluation in Information Retrieval, 2005, S. 33.

[44] vgl. Manning, Raghavan und Schütze: Introduction to Information Retrieval, 2008, S.140.

[45] vgl. Clough und Sanderson: Evaluating the Performance of Information Retrieval Systems Using Test Collections, 2013.

Topics[46]. Es liegen Relevanzurteile für jedes Topic-Dokument-Paar vor, was bei dieser vergleichbar kleinen Kollektion schon zu einem Arbeitsaufwand von über 300.000 Relevanzurteilen[47] geführt hat. Für heutige Testkollektionen mit Dokumentenmengen jenseits der 100.000 Dokumente ist ein solches Szenario nicht denkbar, bei z.B. 50 Topics würden für 100.000 Dokumente schon 5 Millionen Relevanzurteile anfallen. Das Konzept der Cranfield-Kollektion (Dokumentenkorpus, Topics, dazu erstellte Relevanzurteile) ist seitdem zur gängigen Methode geworden, um Retrievaltests aufzubauen, und führte zur Namensgebung „Cranfield-Paradigma".

Die Konferenzreihe TREC ist mit ihren großen Testkollektionen gleichermaßen mit dieser Problemstellung konfrontiert. TREC greift auf das Pooling-Verfahren[48] zurück, um den Arbeitsaufwand für die Relevanzurteile möglichst gering zu halten, ohne große Einbußen in der Analyse der Retrievaltestergebnisse in Kauf nehmen zu müssen. Der TREC-Pool, wie er z.B. für die Ad-Hoc-Aufgabe gebildet wird, setzt sich wie folgt zusammen: Alle teilnehmenden Organisationen nutzen dieselbe Datenbasis und lassen zu einem Topic eine Suchanfrage durch ihr System laufen. Durch die Nutzung verschiedener Systeme und unterschiedlicher Rankingverfahren fällt jede Ergebnismenge ein wenig anders aus. Die ersten X gerankten Ergebnisse werden an TREC übermittelt und dort in einem großen Pool verarbeitet. Die Dubletten werden herausgefiltert und das Ranking wird entfernt. Der so entstandene Dokumentenpool wird dann von der Person, die das Topic eingereicht hat, bewertet.[49]

Pooling als Methode für eine übersichtlichere Anzahl an Relevanzbewertungen findet in der Praxis regelmäßigen Einsatz. Kritisiert wird an der Methode, dass alle Dokumente aus der Kollektion, die sich nicht im Pool finden lassen, bei der Evaluation des Retrievaltests zumeist automatisch als nicht relevant eingestuft werden. Pooling ist also ein Kompromiss, der Arbeit und Zeit bei der Erstellung

---

[46] vgl. Manning, Raghavan und Schütze: Introduction to Information Retrieval, 2008, S.141.

[47] 1398 Dokumente x 225 Topics = 314.550 Relevanzurteile

[48] Karen Spärck Jones und Cornelius J. van Rijsbergen: Report on the Need for and Provision of an 'ideal' Information Retrieval Test Collection, 1975.

[49] vgl. Voorhees und Harman: TREC Experiment and Evaluation in Information Retrieval, 2005, S. 33.

der Relevanzurteile spart, jedoch nie den Anspruch auf eine vollständige und damit ideale Bewertung der Dokumente erheben darf.

Christopher D. Manning formuliert, dass die Dokumenteninhalte sich im Inhalt der Topics wiederfinden sollen, nicht in der Formulierung der Suchanfrage[50]. Genauer formuliert: Das Auftreten aller in der Suchanfrage genutzten Suchbegriffe beispielsweise im Dokumententitel ist keine Garantie dafür, dass ein Dokument tatsächlich relevant für das jeweilige Thema ist.
Relevanz im Allgemeinen ist immer subjektiv, mehrere Relevanzurteile von unterschiedlichen Personen zu einem Topic können sehr unterschiedlich ausfallen. Selbst eine einzelne Person kann für dasselbe Dokument zu demselben Topic an verschiedenen Tagen verschiedene Urteile fällen. In einer Studie aus dem Jahr 1998 führte Ellen M. Vorhees Tests mit unterschiedlichen Relevanzbewertungen von mehreren Personen mit unterschiedlichem Vorwissen durch, das relative Leistungsverhalten der Ergebnisse blieb aber nahezu konstant.[51] Daraus kann geschlossen werden, dass eine einzelne Person ein Topic beurteilen sollte, die Beteiligungen mehrerer Person an einem einzelnen Topic hingegen nicht zielführend ist.
Hilfreich seien ausformulierte Beschreibungen dessen, was sich hinter einem Informationsbedürfnis verberge, zitieren Sanderson und Clough ein Papier von K.A.Kinney et al[52]. Je mehr Vorwissen die bewertende Person habe, desto valider seien die daraus resultierenden Relevanzurteile.[53]

Relevanz kann mittels zwei Verfahren festgehalten werden: Die binäre Relevanz beinhaltet nur relevant (1) oder nicht relevant (0) als mögliche Werte. Eine abgestufte Relevanz kennt mehr als einen Grad von Relevanz, Dokumente können z.B. als nicht relevant (0), teilweise relevant (1) oder relevant (2) bewer-

---

[50] vgl. Manning, Raghavan und Schütze: Introduction to Information Retrieval, 2008, S.140.

[51] vgl. Ellen M. Voorhees: Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1998, S. 315–323.

[52] vgl. Kenneth A. Kinney, Scott B. Huffman und Juting Zhai: How Evaluator Domain Expertise Affects Search Result Relevance Judgments. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management, 2008, S. 591–598.

[53] vgl. Clough und Sanderson: Evaluating the Performance of Information Retrieval Systems Using Test Collections, 2013.

tet werden. Die Entscheidung für ein Verfahren hängt von den Messwerten ab, mit denen die Effektivität des Retrieval ermittelt werden soll.[54]

3.4. Beispiele für Testkollektionen

Es gibt verschiedenste Testkollektionen mit unterschiedlichen Merkmalen und und verschiedenen Einsatzgebieten. Im Folgenden werden zwei bewährte Testkollektionen beispielhaft dargestellt.

Der wohl bekannteste Korpus für Testkollektionen, der aus Deutschland stammt, ist der GIRT-Korpus. Kurz vor der Jahrtausendwende beschloss das Informationszentrum Sozialwissenschaften in Bonn, Retrievaltests zur Überprüfung der Retrievalqualität verschiedener Systeme durchzuführen. Obwohl die Aufgabenstellung vergleichbar war mit der der TREC-Aufgaben, sollte kein TREC-Korpus genutzt, sondern ein neuer Korpus erstellt werden. Das Informationszentrum Sozialwissenschaften wollte vor allem die Retrievalqualität verschiedener Systeme, die mit deutschsprachiger Fachliteratur gespeist werden, prüfen. Dazu wurde ein Korpus konstruiert, der aus deutschsprachigen, fachwissenschaftlichen Dokumenten zum Fachbereich Sozialwissenschaften bestehen sollte. Die Dokumente wurden aus den Datenbanken des Informationszentrum Sozialwissenschaften entnommen und umfassten zu Beginn Autor, Titel, Dokumentensprache, Erscheinungsjahr, intellektuell vergebene Inhaltsmerkmale in Form von Schlagwörtern, Klassifizierungen und Abstracts.
Über die Jahre hat sich der GIRT-Korpus vergrößert und verändert, die Spezialisierung auf den Fachbereich Sozialwissenschaften ist nach wie vor gegeben. Die Version GIRT4-DE entspricht dem aktuellen Korpus und enthält 151.319 deutschsprachige Dokumente. Die Dokumente umfassen Titel, Autor, Dokumentensprache, Erscheinungsjahr, Abstracts, inhaltliche Deskriptoren, und weitere Informationen zum Literaturnachweis, jedoch keine Volltexte. Der Korpus ist daher nicht besonders groß und umfasst ca. 1,2 GB an Daten. Der gesamte Korpus liegt auch in englischer Sprache vor, er umfasst ebenfalls 151.319 Dokumente und heißt GIRT4-EN. Die englische Version ist das

[54] vgl. ebd.

Resultat einer Übersetzung von GIRT4-DE und hat insgesamt einen geringeren Textanteil, da nicht für alle Dokumente eine Abstractübersetzung erfolgte. Ein Korpus, der in zwei Sprachen vorliegt, ermöglicht bei Retrievaltests Aussagen darüber, wie sich die Retrievalqualität unterschiedlicher Systeme verhält, wenn diese mit Dokumenten in verschiedenen Sprachen konfrontiert werden. GIRT findet bei immer wieder Verwendung auch im internationalen Raum, z.B. für TREC-Aufgaben. [55]

Eine zweite, ebenfalls bekanntere Testkollektion ist die iSearch-Testkollektion. iSearch wurde von dänischen Informationswissenschaftlern für Retrievaltests konstruiert, die die Retrievalqualität von einer integrierte Suche beurteilen sollen. Die integrierte Suche umfasst die Suche in mehreren verschiedenen Datenbanken, die Ergebnisse werden alle zusammengefasst und als eine gerankte Ergebnisliste präsentiert. Eine große Herausforderung der integrierten Suche ist die Heterogenität in der Struktur der einzelnen Dokumente der verschiedenen Quellen. Einige Quellen beinhalten Dokumente im Volltext, andere die Titelaufnahme mit einem Abstract, wieder andere die Metadaten zu einem Dokument. iSearch sollte dementsprechend aus unterschiedlichen Dokumenten zusammengesetzt werden, um eine möglichst geeignete Repräsentation einer Datenbasis einer integrierten Suche zu erlangen.
Dazu wurden Fachgebiete Physik, Informatik und Mathematik ausgewählt und ein Korpus aus ca. 160.000 Volltext-Dokumenten und ca. 270.000 Dokumenten, zu denen größtenteils Abstracts vorliegen, konstruiert. Diese Daten wurden der Open-Access-Plattform www.arXiv.org entnommen. Etwa 18.000 bibliothekarische Metadaten wurden ebenfalls hinzugefügt. Die Testkollektion besteht insgesamt aus ca. 450.000 Dokumenten und ist umfasst etwa 56 GB an Daten. Die Titelaufnahmen bestehen mindestens aus Autor, Titel, und der Quelle, hinzu kommen für die meisten Dokumente Abstracts oder der Volltext. Für iSearch wurden 65 Topics entworfen, zu denen jeweils bis zu 200 Dokumente in ihrer

[55] vgl. Michael Kluck: Die GIRT-Testdatenbank als Gegenstand informationswissenschaftlicher Evaluation. In: Informationen zwischen Kultur und Marktwirtschaft. Proceedings des 9. Internationalen Symposiums für Informationswissenschaft ISI, 2004, S. 247–268.

Relevanz bewertet wurden. Auch die iSearch-Testkollektion wurde bereits von verschiedenen Konferenzen im internationalen Raum genutzt.[56]

[56] Marianne Lykke u. a.: Developing a Test Collection for the Evaluation of Integrated Search. In: Advances in Information Retrieval, 2010, S. 627–630.

4. Dokumentation der Kollektionskonstruktion

Im Folgenden werden die der Arbeit zu Grunde liegende Dokumentenkollektion, ihre Herkunft und ihre Eigenschaften skizziert. Unter Einbeziehung der Kapitel 3.1 bis 3.3 werden die Literaturgrundlage und die vorliegende Datenbasis zusammengeführt und daraus Richtlinien zur Erstellung der Testkollektion abgeleitet, die sich für diesen Fall als sachgerecht und anwendbar herausstellen. In den folgenden Teilkapiteln werden die Anwendung und die Durchführung der erarbeiteten Richtlinien dokumentiert.

Der Datensatz, der für die retrievaltestgeeignete Testkollektion benötigt wird, wurde von der DNB zusammen- und zur Verfügung gestellt. Anfänglich wurde der Datensatz für eine Projektgruppe von Herrn Prof. Dr. Klaus Lepsky und Herrn Prof. Dr. Philipp Schaer der TH Köln aus dem Wintersemester 2017/18 generiert. Das Projektziel war es, eine Testkollektion mit Daten der DNB zu erstellen und mit Hilfe dieser einen Retrievaltest zur Qualitätsüberprüfung der in Kapitel 2 erläuterten, sich in der Veränderung befindlichen Sacherschließungspraxis durchzuführen. Durch Komplikationen und zeitliche Verzögerungen ist diese Testkollektion nicht fertiggestellt worden und war bisher keine geeignete Grundlage für einen Retrievaltest. Mit Abschluss dieser Arbeit wird eine Testkollektion für eine nächste Projektgruppe aus dem Sommersemester 2018 zur Durchführung eines oder mehrerer Retrievaltests mit demselben Ziel zur Verfügung stehen.
An die DNB wurden folgende Anforderungen an die zu liefernden Daten gestellt:

- der Umfang soll ca. 200.000 Titelaufnahmen betragen,
- die Daten sollen möglichst aktuell sein,
- bestenfalls sollen alle Fachgebiete vertreten sein, wenn möglich, in vergleichbaren Größenanteilen,
- nicht enthalten sein sollen Belletristik, Kinder- und Jugendliteratur und Kalender,
- es sollen keine elektronischen Parallelausgaben enthalten sein,
- die Sacherschließungsmerkmale sollen in ihrer Herkunft erkennbar sein (intellektuell oder maschinell)

- optimal wären zwei Varianten der Kollektion: eine rein intellektuell und eine automatisch erschlossene Variante.

Von großem Vorteil für die Vergleichbarkeit der Ergebnisse der Retrievaltests wären zwei gleiche Kollektionen gewesen, einmal rein intellektuell verschlagwortet, einmal rein maschinell. Dies war jedoch nicht möglich. Die Sacherschließungsmerkmale liegen in getrennten Kategorien vor, was eine Suche nach rein maschinell oder rein intellektuell erschlossenen Dokumenten ermöglicht. Hieraus ergeben sich auch Optionen zur Vergleichbarkeit der Ergebnisse eines Retrievaltests, optimaler wären jedoch zwei identische Kollektionen gewesen, die unterschiedlich erschlossen wurden. Die Vorgabe, elektronische Parallelausgaben zu vermeiden, konnte ebenfalls nicht ganz umgesetzt werden: Während der Erstellung der Relevanzurteile tauchten einige wenige Titelaufnahmen auf, die dasselbe Dokument mit unterschiedlichem Ressourcentyp beinhalteten.
Nicht enthalten in den gelieferten Daten sind Belletristik, Kinder- und Jugendliteratur und Kalender. Für die Erstellung der Relevanzurteile sind Medienwerke dieser Gattungen nicht geeignet, da ihr Anteil im Katalog groß ist und sie wenig bis keinen Wert für ein Topic bzw. ein möglichst reelles Informationsbedürfnis haben. Diese Medienwerke stellen potentiellen Ballast für einen Retrievaltest dar, der nachträglich herausgefiltert werden müsste.

Von der DNB wurden wie gewünscht 200.000 Titelaufnahmen geliefert. Diese Größe ist handhab- und verarbeitbar, lässt sich in angemessener Zeit über technische Schnittstellen übertragen, und bietet zudem eine ausreichend große Anzahl an Titelaufnahmen, die ein breites inhaltliches Spektrum ermöglichen. Die inhaltliche Verteilung erstreckt sich über 100 Sachgruppen[57]. Der Katalog der DNB bietet das denkbar größte Sachgruppenspektrum, das auf eine Testkollektion übertragen werden muss. Die gelieferten 100 Sachgruppen sind ein Ansatz, alle möglichen Themengebiete abzudecken und können keinen Anspruch auf Vollständigkeit erheben. Einige Sachgruppen, beispielsweise Medizin, sind von einer höheren Quote an Veröffentlichungen geprägt als Sachgruppen, wie z.B. Paläontologie. Die Daten umfassen also in einigen

[57] das Aufteilungsverhältnis ist aufgeschlüsselt im Anhang 1.

Sachgruppen mehr, in anderen weniger Titelaufnahmen, in den meisten Fällen ist aber eine gleichmäßige Verteilung umgesetzt worden.
Die Daten der DNB bestehen aus Titelaufnahmen ohne Inhaltsverzeichnisse oder Volltexte, es sind nur Metadaten zu finden. Alle gelieferten Daten wurden in dem internen Format Pica+ übermittelt.

Wie in Kapitel 3, S.16, erläutert, besteht eine Testkollektion aus drei Bestandteilen: der Dokumentenmenge, den Topics und den Relevanzurteilen. Die Basis für den Korpus lieferte die DNB. Die Titelaufnahmen müssen jedoch in ein Format überführt werden, das zur Bearbeitung geeignet ist. Die Metadaten müssen zudem auf Kategorien überprüft werden, die für die Erstellung der Relevanzurteile keinen Wert haben und somit entfernt werden können. Festgelegt werden sollte auch, ob der sich der Korpus für eine wiederverwendbare Testkollektion eignet oder nicht.
Für die Erstellung der Topics muss überlegt werden, ob Statistiken oder Nutzungsdaten ausgewertet werden und daraus Topics abgeleitet werden sollen oder ob auf bereits vorhandene Topics zurückgegriffen werden kann. Die Anzahl und das Format der Topics muss ebenso festgelegt werden. Ferner muss für jedes Topic am Ende eine Beschreibung existieren, die es erleichtert, die Bewertung der Dokumente entsprechend nachzuvollziehen. Die Topics und ihre Bestandteile müssen zu diesem Zweck in tabellarischer Form abrufbar sein, um die Weiterverwendung der Testkollektion zu ermöglichen.
Die Erstellung der Relevanzurteile ist in diesem Fall aufgrund der großen Dokumentenmenge eine Herausforderung. Für die Erstellung dieser Testkollektion gibt es keine finanziellen Ressourcen, eine Auslagerung der Erstellung der Relevanzurteile kommt daher nicht in Frage. Der Mangel an personellen Ressourcen begünstigt umfangreiche Relevanzurteile nicht. Hier muss ein Verfahren gewählt werden, welches ausreichend Relevanzurteile für jedes Topic generiert, ohne dass allzu große Einbußen in Bezug auf die Vollständigkeit in Kauf genommen werden müssen. Das aus den Ad-Hoc-Aufgaben von TREC bekannte Pooling-Verfahren kann in abgeänderter Form eine geeignete Herangehensweise sein. Es empfiehlt sich, im Sinne der Stringenz je Topic nur eine Person die Relevanzurteile fällen zu lassen. Verständlich und sinnvoll formulierte Topics erleichtern die Entscheidung,

welches Dokument einer Ergebnismenge für das jeweilige Topic relevant ist. Des Weiteren ist die Verschriftlichung festzulegen: Soll die Relevanz binär mit 0 und 1 festgehalten werden oder wird eine abgestufte Relevanz mit mehr Werten genutzt? Ebenso wie die Topics müssen sowohl die Relevanzurteile als auch die Suchanfragen, die zur Generierung der bewerteten Ergebnismengen eingegeben wurden, festgehalten werden.

4.1. Dokumentation der Erstellung des Korpus

Der Korpus, den die DNB geliefert hat, besteht wie bereits erwähnt aus 200.000 Dokumenten und ist mit einer inhaltlichen Verteilung über 100 Sachgruppen breit gefächert. In der von der DNB übermittelten Version sind ca. 130.000 Dokumenteninhalte intellektuell und ca. 70.000 Dokumenteninhalte maschinell erschlossen worden. Die im vorangehenden Teilkapitel erwähnte Tabelle bietet einen groben Überblick über die Aufschlüsselung der Dokumente und ihrer Erschließungsform in allen Sachgebieten über den gesamten Korpus. Einzelne Zahlen mögen geringfügig von der Realität abweichen, die Verteilung ist aber gleich geblieben. Dieser Umstand entstand durch geringfügige Komplikationen bei der Zusammenstellung der Dokumente.
Der Gesamtkatalog der DNB umfasst mehr 32 Millionen Medienwerke[58], eine Testkollektion aus diesem Datenbestand kann bei einer solchen Gesamtgröße höchstens eine Teilkollektion sein. Allein aus technischen Gründen kann nicht der gesamte Datenbestand der DNB als Korpus für die Testkollektion in Frage kommen, allein die Datenübertragung würde sich als äußerst kompliziert erweisen, da kaum eine Schnittstelle im Alltag auf solchen Datenmengen ausgelegt ist.

Das gelieferte Format Pica+ ist ein internes bibliografisches Datenformat einer speziellen Software, welches in seiner Reinform eine sehr schlechte Lesbarkeit aufweist (s. Abbildung 2). Das Format musste in ein übersichtlicheres Format umgewandelt werden. Die Projektgruppe einigte sich auf die Nutzung der

[58] „Deutsche *Nationalbibliothek* - Wir über uns“, Homepage. http://www.dnb.de/DE/Wir/wir_node.html (15. Mai 2018).

Suchplattform Solr als Arbeitsgrundlage. Solr ist eine Open-Source-Plattform, die von der Apache Software Foundation angeboten wird.[59] Solr wird über die Eingabeaufforderung installiert, gestartet und mit Dokumenten gespeist. Die Plattform wird über den Browser bedient, die Suche ist feldbasiert. Die Daten der DNB wurden in das von Solr unterstütze Format XML überführt. Bei dieser Überführung wurden die „Ballast-Kategorien“ der Metadaten (Beispiele folgen) herausgefiltert, nicht zuletzt aus Gründen der Übersichtlichkeit und der Datensatz-Verkleinerung. In den herausgefilterten Kategorien befinden sich größtenteils Zahlenketten, die weder für die Bewertung der Relevanz noch für andere Zwecke von Nutzen sind. Die Kategorien haben intern in der DNB möglicherweise einen anderen Stellenwert, führen in diesem Fall aber beispielsweise zu einer sehr unübersichtlichen Ansicht der Dokumente über Solr und mussten daher entfernt werden.

```
001@ $01-2$a5
001A $01145:21-11-07
001B $09999:06-05-08$t23:52:21.000
001D $01140:02-04-08
001U $0utf8
001X $00
002@ $0Aa
003@ $0986480932
004A $0978-3-89717-528-0$fPp. : EUR 9.90 (DE), EUR 10.20 (AT), sfr 18.90 (freier Pr.)
004A $03-89717-528-2$fPp. : EUR 9.90 (DE), EUR 10.20 (AT), sfr 18.90 (freier Pr.)
004K $09783897175280
006T $007,N50,0991
006U $008,A20,1326
006V $03028379
007I $So$0227331167
010@ $ager$ceng
011@ $a2008
017A $ara
019@ $aXA-DE-NW
021A $aZimmerpflanzen$hDorte Nissen. [Übers. aus dem Engl.: Angela Kuhk]
022A/01 $aIndoor plants$rdt.
028C $9134129520$7Tn3$Agnd$0134129520$dDorte$aNissen
028C/01 $9120863537$7Tp1$Vpiz$Agnd$0120863537$E1961$dAngela$aKuhk$BÜbers.
033A $pKöln$nFleurus Idee$55107256
034D $a256 S.
034I $a21 cm, 580 gr.
034M $aüberw. Ill.
041A $90406781O5$7Ts1$Vsaz$Agnd$04067810-6$aZimmerpflanzen
041A/01 $af Wörterbuch
041A/08 $f12
041A/09 $eDE-101$rDE-101
044N $bVLB-FS$aPflanzen
044N $bVLB-FS$aZimmerpflanzen
044N $bVLB-PF$aBA: Buch
044N $bVLB-WN$a1421: HC/Ratgeber/Natur/Garten
045E $e630
045F $eDDC22ger$a635.96503
045F/01 $a635.965
045F/03 $f03
047A $SFE$aSm
047A $SERW$aAJu
047I $u$$c04$dDNB$e1
```

Abbildung 2: Beispielhafte Titelaufnahme im Format Pica+.

[59] „Apache Solr“, Homepage. http://lucene.apache.org/solr/ (15. Mai 2018).

Die Metadaten der DNB in Pica+ bestehen je Titelaufnahme aus vielen Kategorien, von denen die meisten für einen Retrievaltest keine Bedeutung haben. Die Daten wurden auf Nützlichkeit geprüft und dementsprechend gefiltert. Felder wie z.B. 001@-001X oder 047A und 047I (s. Abbildung 2) enthalten nur Zeichenketten, die zur inhaltlichen Beurteilung der Titelaufnahme irrelevant sind. Im Folgenden sind die Kategorien aufgeführt, deren Inhalt für die Erstellung der Relevanzurteile genutzt werden kann oder die aus anderweitigen, ebenfalls im Folgenden erläuterten Gründen unverzichtbar sind:
(Die entsprechenden Pica+-Felder sind jeweils in Klammern angegeben)

- Dokument-ID (003@): Dieser eindeutige Identifikator in Form einer Zeichenkette ist in jeder Titelaufnahme enthalten und unerlässlich für den Datenexport der Ergebnismengen, die die Grundlage für die Relevanzurteile darstellen. Die ID gewährleistet eine eindeutige Relevanzzuordnung von jedem Dokument zu dem entsprechenden Topic und ist entscheidend bei der Dublettenentfernung, die bei einem poolingähnlichen Verfahren eingesetzt wird.
- Titel (021A): Der Titel des Dokuments und sein Zusatztitel sind für die Relevanzurteile von großer Bedeutung.
- Autor(028A, 028B): Das Wissen um den Namen des Autors kann für eine tiefergehende Recherche zur Unterstützung der Bewertung der Relevanz von Vorteil sein.
- Bearbeiter (028C, 028D): Ebenso wie im Fall des Autors können die Namen der Bearbeiter des Dokuments entscheidend sein.
- Verlag und Verlagsort (033A), Erscheinungsjahr (011A), Auflage (032A), ISBN-Nummer (004A), Sprache des Dokuments (010@), Umfang des Dokuments (034D): All diese Informationen sind hilfreich für eine erweiterte Recherche, die der Relevanzbewertung dient.
- Gesamtheitenvermerk (036E): Handelt es sich bei dem Dokument z.B. um einen Artikel aus einer Zeitschrift, sind der Titel der Zeitschrift, der Band und das Erscheinungsjahr hilfreiche Informationen, die helfen können, das Dokument korrekt einem Sachgebiet zuzuordnen.

- DDC-Notationen (045F), Sachgruppen-Zuordnung (045E): DDC-Notationen sind von Bedeutung für eine thematische Einordnung des Dokuments, die für ein Relevanzurteil entscheidend sein kann.
- Intellektuell vergebene Schlagwörter (041A): Die aus dem Vokabular der GND stammenden Schlagwörter sind ein sehr wichtiger Faktor bei der Bewertung der Relevanz.
- Maschinell vergebene Schlagwörter (044H): Die Inhalte dieser Felder können ebenfalls die Erstellung von Relevanzurteilen erleichtern. Außerdem kann über dieses Feld geprüft werden, ob für eine bestimmte Suchanfrage Dokumente vorliegen, die maschinell erschlossen wurden. Dies ist vor allem für den Retrievaltest von Bedeutung, z.B. wenn intellektuell erschlossene Dokumente mit rein maschinell erschlossenen verglichen werden.
- Schlagworte der Verlage/Sonstige Informationen der Verlage, die übernommen wurden (044N): Verlage liefern häufig inhaltserschließende Merkmale zu ihren Medienwerken. Für die Bewertung der Relevanz können auch diese Informationen von Wert sein.
- Fußnote (037C): Eine Fußnote kann nützliche Anmerkungen oder Informationen enthalten, die zur Einordnung des Dokuments beitragen.

Der eingefügte Screenshot (Abbildung 3, S. 32) zeigt die Ansicht einer in Solr indexierten Titelaufnahme nach Filterung der überflüssigen Kategorien.

Anhand dieser typischen Titelaufnahme, bei der fast alle Kategorien besetzt sind, ist zu erkennen, dass die oben erläuterten Kategorien sich in dieser Felder-Zuordnung wiederfinden:

- Dokument-ID: „id“
- Titel: „title_txt_de“ bzw. „title_s“
- Bearbeiter: „editor_ss“
- Verlag, Verlagsort: „imprint_ss“
- Erscheinungsjahr: „year_s“
- ISBN-Nummer: „isbn_ss“
- Sprache des Dokuments: „lang_ss“
- Umfang des Dokuments: „size_s“
- DDC-Notationen: „class_ddc_ss“ bzw. „class_ddc_txt_de“

- Sachgruppen-Zuordnung: „class_dnb_ss“ bzw. „class_dnb_txt_de“
- Intellektuell vergebene Schlagworte: „subject_gnd_ss“ bzw. „subject_gnd_txt_de“
- Schlagworte der Verlage/Sonstige Informationen der Verlage: „subject_vlb_ss“ bzw. „subject_vlb_txt_de“


```
{
  "collection_s":"dnb",
  "id":"0986480932",
  "title_txt_de":["Zimmerpflanzen"],
  "title_s":"Zimmerpflanzen",
  "editor_ss":["Nissen",
    "Kuhk"],
  "imprint_ss":["Köln : Fleurus Idee"],
  "year_s":"2008",
  "size_s":"256 S.",
  "isbn_ss":["0978-3-89717-528-0",
    "03-89717-528-2"],
  "lang_ss":["ger"],
  "subject_gnd_ss":["Zimmerpflanzen",
    "f Wörterbuch"],
  "subject_gnd_txt_de":["Zimmerpflanzen",
    "f Wörterbuch"],
  "subject_vlb_ss":["Pflanzen",
    "Zimmerpflanzen",
    "BA: Buch",
    "1421: HC/Ratgeber/Natur/Garten"],
  "subject_vlb_txt_de":["Pflanzen",
    "Zimmerpflanzen",
    "BA: Buch",
    "1421: HC/Ratgeber/Natur/Garten"],
  "class_ddc_ss":["635.96503"],
  "class_ddc_txt_de":["635.96503"],
  "class_dnb_ss":["630"],
  "class_dnb_txt_de":["630"],
  "_version_":1596375809959919619,
  "score":16.940464},
```


Abbildung 3: Beispielhafte Titelaufnahme in Solr.

Kategorien, die in diesem Beispiel nicht besetzt sind:

- Autor: „author_ss“
- Auflage: „notes_ss“
- Gesamtheitenvermerk: „series_ss“

- Maschinell vergebene Schlagworte: „subject_auto_ss“ bzw. „subject_auto_txt_de“
- Fußnote: „footnote_ss“ bzw. „footnote_txt_de“

Kategorien, die in jeder Titelaufnahme enthalten sind:

- „collection_s“: Diese Kategorie enthält den Namen der Kollektion, in der gerade gesucht wird. Der Inhalt bleibt immer gleich.
- „_version_“: Diese Versionsnummer wird von Solr vergeben. Sie hat für die Erstellung der Testkollektion oder den Retrievaltest keine Bedeutung.
- „score“: Der Score wird von Solr errechnet und ist als eine Art Übereinstimmungswert zu werten. Je höher der Score ausfällt, desto ähnlicher sind die Worte der einzelnen Titelaufnahme den Worten der Suchanfrage. Die Dokumente werden mit abnehmendem Score angezeigt.

Die verschiedenen Endungen der Kategorien-Namen („_txt_de" und „_ss“ oder „_s“) bzw. die scheinbare Dopplung einiger Kategorien (z.B. „title_txt_de“ und „title_s") ergeben sich aus der angepassten Indexierung über Solr. In den Kategorien Titel, intellektuell vergebene Schlagworte, maschinell vergebene Schlagworte, Schlagworte der Verlage, DDC-Nationen, Sachgruppe der DNB und Fußnote wurde der Indexierungsstandard von Solr um einige Funktionen (u.a. Stemming, Tokenizer, Stopwords) erweitert, damit der deutsche Text dieser Kategorien besser über eine natürlichsprachige Suche zu finden ist. Die scheinbar doppelt vorhandenen Kategorien werden dementsprechend über die Kategorie mit der Endung „_txt_de“ genutzt und durchsucht. Diese Änderung wurde über eine leicht veränderte Schema-Datei vorgenommen, die vor der Indexierung des Korpus im Unterordner „solr-7.1.0/server/solr/dnb/conf“ abgelegt wird und die bereits enthaltene Datei „managed-schema“ ersetzt. Die abgeänderte Datei ist auf der beigelegten DC-ROM enthalten.

Die Reduzierung des Korpus auf relevante Kategorien und deren Darstellung bzw. Umsetzung in Solr ist somit abgeschlossen. Der Korpus liegt als Textdatei vor, er ist nicht veränderbar und somit statisch. Diese Tatsache ist vor allem für die Vergleichbarkeit unterschiedlicher Retrievaltests und eine eventuelle Wiederverwendbarkeit von Vorteil. Da der Korpus nur aus Metadaten und nicht

aus Volltexten besteht, entstehen auch keine urheberrechtlichen Probleme bei einer weiteren Nutzung für andere Tests.

4.2. Dokumentation der Erstellung der Topics

Für die Auswahl der Topics, die in der Testkollektion genutzt werden sollen, wurden keine Nutzungsstatistiken der DNB oder Suchanfrage-Mitschriften von realen Bibliotheksbenutzern genutzt. In der Projektgruppe wurde beschlossen, dass ein vorhandener Topic-Pool als Grundlage für die Bildung neuer Topics herangezogen werden soll. Hierfür kamen die Topics, die bereits im MILOS II-Projekt bei einem Retrievaltest genutzt wurden, in Frage. Das MILOS II-Projekt beschäftigt sich mit der automatischen Indexierung von Titeldaten und wurde 1998 mit Daten der DNB durchgeführt. Die Datenbasis für den MILOS II-Retrievaltest ist vergleichbar mit der Basis der Testkollektion, die Gegenstand dieser Arbeit ist. In den schriftlich festgehaltenen Ergebnissen für den MILOS II-Test werden 100 Topics aufgeführt.[60] Diese Topics wurden für einen Korpus geschaffen, der aus Titelaufnahmen der DNB besteht und sich über alle Sachgruppen erstreckt, ausgenommen der Belletristik, Kinder- und Jugendliteratur und Kalender, die Datenbasis basiert folglich auf ähnlichen Voraussetzungen.

Aufgrund der Größe der Teilnehmerzahl in der Projektgruppe wurde die Zahl der zu erstellenden Topics auf 50 festgelegt. Die 100 Topics aus MILOS II wurden gesichtet und gefiltert. Es wurden vor allem die Topics entfernt, die zu keinen oder sehr wenigen Treffern geführt hatten, da sie teilweise sehr spezifisch sind. Wegen der Vergleichbarkeit der Dokumentenkollektionen ist zu erwarten, dass diese Topics für den geplanten Retrievaltest nicht geeignet sind. Eine Auflistung aller ausgewählten Topics befindet sich im Anhang (vgl. Anhang 2). Die Topics bestehen, wie bei TREC, aus einer zweistelligen Topic-ID, dem Topic-Titel, einer Description, die in einem Satz zusammenfasst, was gesucht

[60] Elisabeth Sachse, Martina Liebig und Winfried Gödert: Automatische Indexierung unter Einbeziehung semantischer Relationen: Ergebnisse des Retrievaltests zum MILOS II-Projekt, 1998, S. 20ff.

wird, und dem Narrative, einer Erläuterung dessen, was für das Topic relevant, teilweise relevant und nicht relevant ist.
Die beschreibenden und erläuternden Teile Description und Narrative sind von den Projektteilnehmern entworfen worden und wurden während der Erstellung der Relevanzurteile teilweise geringfügig abgeändert. Bei einer genauen Betrachtung der ausgegebenen Ergebnisse kann der Gutachter beispielsweise die Erkenntnis erlangen, dass ein bestimmter Typ von Dokumenten nicht relevant ist, und muss ggf. im Narrative eine Formulierung ändern. Für das Topic 08 - Tierexperimente lassen sich beispielsweise sehr viele medizinische Studien in der Ergebnismenge finden, in denen ein Tierversuch durchgeführt wurde oder deren Grundlage Tierversuche sind. Im Narrative steht: „Relevante Dokumente thematisieren Tierversuche/Tierexperimente, vor allem kritische Auseinandersetzungen damit. Nicht relevant sind andere medizinische Studien, in denen Tierversuche gemacht wurden.“ Einige der gefundenen Dokumente sind für die Suche nach Argumenten für und gegen Tierversuche und deren ethische Vertretbarkeit nicht als Ergebnis geeignet und daher irrelevant. Description und Narrative sind auch von großer Bedeutung für die Nachnutzung der Testkollektion. Anhand dieser Topic-Bestandteile kann nachvollzogen werden, warum bestimmte Dokumente für ein Topic als relevant oder nicht relevant bewertet wurden.

Die Topics wurden für die Erstellung dieser Testkollektion in eine einheitliche Form gebracht und sind der Projektgruppe, die sich als nächstes mit dem Retrievaltest beschäftigen wird, zugänglich gemacht worden. Da sowohl alle Topics als auch alle Relevanzurteile von derselben Person final überarbeitet bzw. erstellt wurden, werden für die Testkollektion ähnliche Voraussetzungen erfüllt wie bei der Erstellung der Topics und Relevanzurteile bei TREC. So wird vermieden, dass unterschiedliche Relevanzurteile unterschiedlicher Personen die Ergebnisse möglicher Retrievaltests beeinflussen können. Die Subjektivität in den Relevanzurteilen gänzlich zu vermeiden ist unmöglich. Wenn die Urteile zu einem Topic aber von derselben Person stammen, sind sie in sich stringent und können als valide betrachtet werden. Wie in Kapitel 2.3 erläutert, ist die Bearbeitung eines Topics durch mehrere Personen weniger geeignet.

Zusätzlich zu der tabellarischen Ansicht sind die Topics in dem Standard-Format, das von TREC genutzt wird, auf der beiliegenden CD-ROM in einer Textdatei enthalten. Das Format sieht für jedes Topic eine XML-Struktur vor, das aus folgenden Tags besteht:

- <top>, dient der Zusammenfassung aller folgenden Tags,
- <title>, enthält den Titel des Topics, meist ein oder zwei Worte,
- <desc>, enthält die Description,
- <narr>, enthält das Narrative.

Alle Tags werden nacheinander geöffnet, mit Inhalt in Form einer Zeichenkette versehen, und geschlossen. <top> wird geschlossen, nach dem alle anderen Tags geschlossen worden.

4.3. Dokumentation der Erstellung der Relevanzurteile

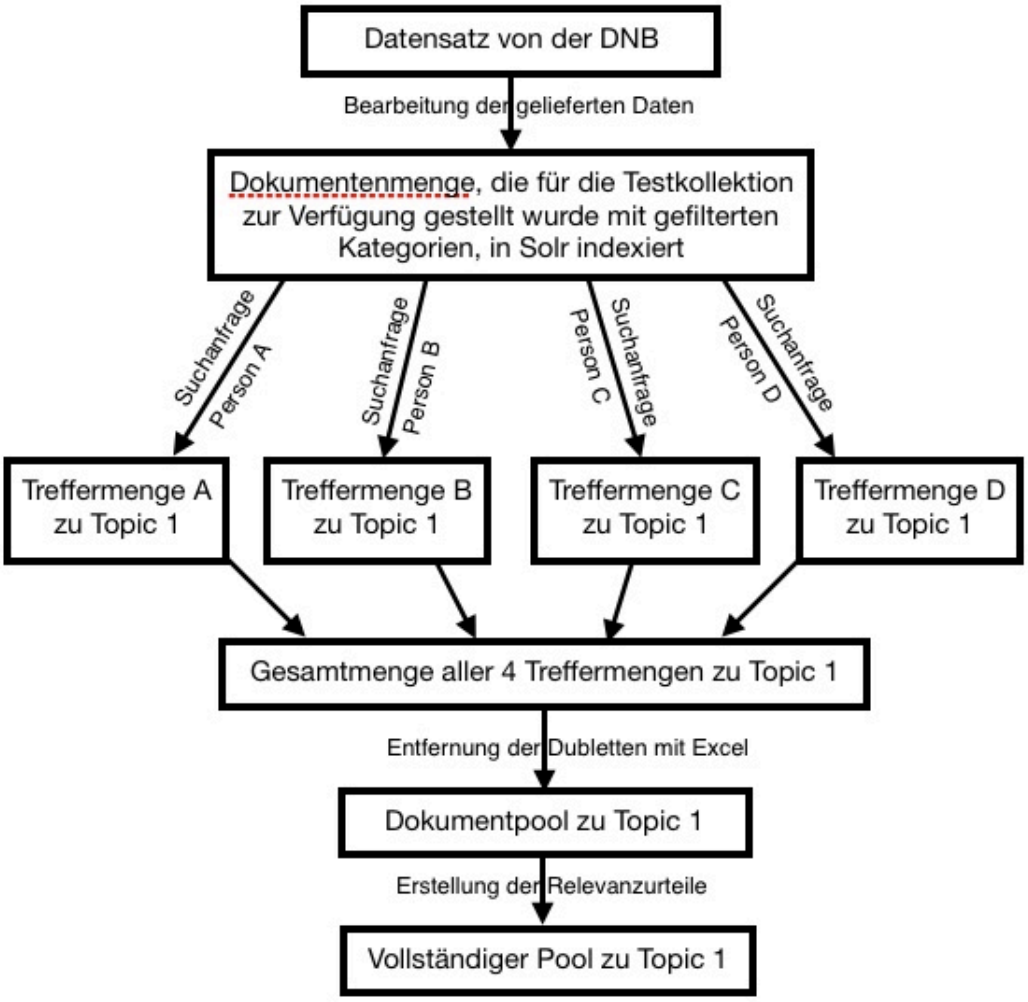

Abbildung 4: schematische Darstellung der Pool-Erstellung.

Der Korpus der Testkollektion besteht aus 200.000 Dokumenten und es wurden 50 Topics ausgewählt, die die Grundlage für die Suchanfragen bilden. Der ideale Fall, in dem für jedes einzelne Dokument-Topic-Paar ein Relevanzurteil vorliegt, würde die Erstellung von 10 Millionen Relevanzurteilen bedeuten, die von Menschen gefällt werden müssen. Um diesen enormen Arbeitsaufwand zu umgehen, wurde für jedes Topic ein Dokumentenpool gebildet, der für die Relevanzurteile genutzt wurde.

Das dazu genutzte Pooling-Verfahren (vgl. Abbildung 4, S. 36) ist angelehnt an das Verfahren, das bei TREC für die Pool-Erstellung im Rahmen der Ad-Hoc-Aufgabe betrieben wird. Bei TREC suchen die Systeme unterschiedlicher Teilnehmer im gleichen Korpus für das gleiche Topic. Da die Systeme meist eine andere Funktionsweise haben, unterscheiden sich die Ergebnismengen. Die ersten X der gerankten Ergebnisse werden jeweils an TREC übermittelt, daraus wird ein großer Pool gebildet und ebendieser wird bewertet.

Für die Erstellung dieser Testkollektionen konnte nicht auf unterschiedliche Systeme zurückgegriffen werden, die den Korpus durchsuchen. Daraus ergeben sich leichte Änderungen bei der Pool-Bildung. Vier Personen mit bibliothekarischen Fachkenntnissen und Vorwissen zum Korpus der DNB entwickelten zu jedem Topic eine Suchstrategie und formulierten eine Suchanfrage. Für alle 50 Topics liegen also je vier Suchanfragen vor, deren Ergebnismengen zu einem Pool zusammengefasst werden können. Die vier Suchanfragen für jedes Topic sind im Anhang 3 mit den jeweiligen Ergebnismengen dokumentiert. Jede Ergebnismenge ist in Solr mit Hilfe eines Stylesheets (Anhang 4, befindet sich ebenso auf der beigefügten CD-ROM) in das Standard-TREC-Format umgewandelt und als Textdokument gespeichert worden. Das Stylesheet wurde dazu in das zuvor in der solr-Ordnerstruktur angelegte Unterverzeichnis „solr-7.1.0/server/solr/dnb/conf/xslt" gelegt und für jedes Topic wurde die entsprechende Topic-ID in Zeile 10 eingetragen.

Das Standard-TREC-Format umfasst sechs Felder (vgl. Abbildung 5, S. 38):

- Topic-ID (im Beispiel die 22)
- eine Konstante 0, die trec_eval benötigt, jedoch unberücksichtigt bleibt
- Dokument-ID (Zeichenkette)

- non Solr ermitteltes Ranking (0-X)
- Score, der in Kapitel 4.1 erwähnte Übereinstimmungswert
- Name des Run (im Beispiel simpleRun)

```
22 0 01082131563 0 32.107918 simpleRun
22 0 01056961171 1 29.866627 simpleRun
22 0 01045552992 2 29.139214 simpleRun
22 0 01097484920 3 28.828114 simpleRun
```

Abbildung 5: Beispiel für das Standard-TREC-Format.

Durch das Einfügen der Anweisung

&fl=id&wt=xslt&tr=solr-rel.xsl&rows=1000

an das Ende der Zeile der Ergebnisausgabe von Solr (s. Abbildung 6, S. 39, grau unterlegtes Feld, durch Anklicken zu öffnen) wurde die Ergebnisliste mittels des verwiesen Stylesheets „solr-rel.xsl“ in das TREC-Standard-Format umgewandelt. Dieser Vorgang wurde für jede Suchanfrage zu jedem Topic wiederholt.

Insgesamt wurden so 200 Ergebnisliste im TREC-Standard-Format produziert und mit dem im Weiteren erläuterten Verfahren in 50 Pools, je Topic ein Pool aus vier Suchanfragen bzw. deren Ergebnislisten, zusammengeführt. Um Dubletten zu vermeiden, wurden alle vier für ein Topic als Textdokument vorliegenden Ergebnislisten in das Tabellenkalkulationsprogramm Excel in ein Tabellenblatt eingefügt und mittels einer dort vorhandenen Funktion auf Dubletten (über die Spalte, die die Dokument-ID enthält) geprüft. Die erkannten Dubletten wurden entfernt und die verbliebenen Dokumente in eine Textdatei konvertiert. Für jedes Topic wurde eine von diesen Textdateien erstellt. Einige Topics sind von einem höheren Medienwerkaufkommen geprägt als andere, die Pools haben daher unterschiedliche Größen. Die Anzahl der Ergebnisse pro Pool liegt mehrheitlich zwischen 50 und 150 Dokumenten. Der kleinste Pool besteht aus 13 Treffern (39 - Medizin im Dritten Reich), der größte Pool dagegen besteht aus 514 Treffern (14 - Alternative Energien), derartige Ausreißer sind Ausnahmen.

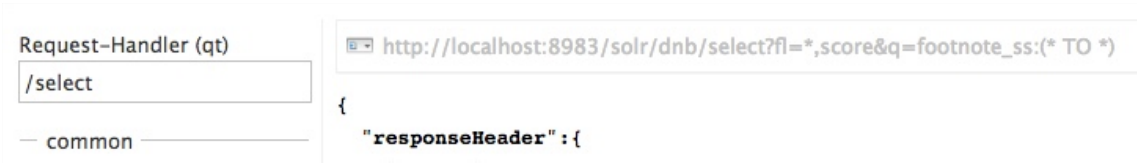

Abbildung 6: Ergebnisausgabenzeile in Solr.

Diese Pools sind die Grundlage der Erstellung der Relevanzurteile. Dazu wurden die Pools und eine Tabelle mit den 50 Topics und den dazugehörigen Topic-IDs in das Tool Relevation![61] eingespielt. Das Open-Source-Tool wird über die Eingabeaufforderung installiert, gestartet und mit einer Sammlung von Textdokumenten jeder einzelnen im Korpus vorhandenen Titelaufnahme gespeist. Bedient wird Relevation! über den Browser. Es erleichtert die Erstellung der Relevanzurteile deutlich, da für jedes Topic die Dokumente im zuvor festgelegten Format angezeigt werden und dann ausgewählt werden kann, ob das Dokument relevant, teilweise relevant oder nicht relevant ist. Die Nutzung des Tools beugt Fehlern vor, die entstehen, wenn die Relevanzurteile händisch in die Textdatei der Ergebnismengen eingefügt werden, und stellt vor allem für größere Ergebnismengen eine Arbeitserleichterung dar.

Die Relevanz wurde in abgestufter Form dokumentiert. Ein Treffer einer Ergebnismenge kann somit drei Werte annehmen: „0“ für nicht relevant, „1“ für teilweise relevant und „2“ für relevant. Es ist bei der Erstellung der Urteile darauf geachtet worden, dass diese sich an den Vorgaben aus den Topics und dem Narrative orientieren. Relevanzbewertungen können durch ihre Subjektivität jedoch nie den Anspruch auf Perfektion erheben. Auch die Vollständigkeit der Relevanzurteile ist durch die Nutzung eines Pooling-Verfahrens nicht gewährleistet. Das Pooling-Verfahren ist in diesem Fall ein Kompromiss zwischen Arbeitsaufwand, zur Verfügung stehender Zeit und den Ressourcen Mensch und Geld.

[61] “relevation: Information Retrieval Relevance Judging System" Hompage. https://github.com/ielab/relevation (15. Mai 2018).

5. Retrievaltest als Machbarkeitsnachweis

Die in Kapitel 4 dokumentierte Bearbeitung der Testkollektion ermöglicht die Durchführung von Retrievaltests. Die Testkollektion besteht nun aus den drei essenziellen Bestandteilen (s. Kapitel 3 S.16):

- eine überarbeitete Dokumentenkollektion, die alle relevanten Kategorien der Titelaufnahmen enthält,
- reelle Informationsbedürfnisse in Form von 50 Topics, die mit ID, Description und Narrative vorliegen (s. Anhang 2),
- Relevanzurteile, die für jedes Topic zu den mittels eines poolingähnlichen Verfahrens ermittelten Treffermengen gefällt wurden.

Folglich ist die Testkollektion vollständig. Zur Überprüfung dessen wird ein beispielhafter Retrievaltest durchgeführt. Dieser Test prüft zunächst keine zuvor aufgestellte These, sondern dient in erster Instanz dem Nachweis der Einsatzfähigkeit der erstellten Testkollektion für Tests dieser oder vergleichbarer Art.

Für den beispielhaften Retrievaltest wird das Programm trec_eval genutzt, das von der Konferenzreihe TREC für die Evaluation der Ergebnisse der Ad-Hoc-Aufgaben genutzt wird.[62] trec_eval wird über die Eingabeaufforderung genutzt und arbeitet mit zwei Textdateien, den Relevanzurteilen und der Treffermenge zu einer Suchanfrage für ein bestimmtes Topic. Die Treffermengen müssen im Standard-TREC-Format (s. Kapitel 4.3) vorliegen und sind pro Suchanfrage in der einen Textdatei gespeichert. Die Relevanzurteile werden aus der zweiten Textdatei, die alle Relevanzurteile zu der Kollektion enthält, abgerufen. Diese Datei kann über eine Funktion vom Tool Relevation! automatisch angelegt werden, wenn die Erstellung der Relevanzurteile beendet wurde. Die Urteile müssen ebenfalls in einem bestimmten Format vorliegen (vgl. Abbildung 7). Durch Leerzeichen (oder Tabstopp) getrennt sind pro Relevanzurteil vier Felder besetzt:

[62] „Text REtrieval Conference (TREC) trec_eval“, Homepage. https://trec.nist.gov/trec_eval/ (15. Mai 2018).

```
19 0 0986649368 2
19 0 0986656135 2
19 0 0987420135 0
19 0 0987455613 1
```

Abbildung 7: Beispiel von trec_eval benötigte Format der Relevanzurteile.

- Topic-ID (im Beispiel die „19“),
- eine Konstante „0", die trec_eval benötigt, jedoch unberücksichtigt bleibt,
- die Dokument-ID (Zeichenkette),
- das Relevanzurteil für das Dokument zu dem jeweiligen Topic, dargestellt durch „0", „1“ oder „2“.

Das Tool Relevation! gab jedoch eine Textdatei aus, in der die Spalte mit der Konstante „0“ an die Spalte, die die Dokument-ID enthält, geheftet wurde. Somit enthielt die ausgeworfene Datei nur drei Spalten. Über die Nutzung einer Excel-Funktion wurde das Format angepasst und die Dokument-ID korrigiert. Die korrigierte Version ist auf der beigelegten CD-ROM enthalten.

Die in der Treffermenge auftretenden Dokumente werden in trec_eval mit denen, die bewertet in der Datei mit den Relevanzurteilen vorliegen, anhand der Topic-ID und der Dokument-ID verglichen. Das Programm trec_eval gibt anschließend aus, wie viele Treffer gefunden wurden, wie viele relevante Dokumente in der gesamten Kollektion zu dem Topic vorliegen, und wie viele dieser relevanten Dokumente sich in der Treffermenge befinden. Dazu ermittelt das Programm einige Werte, die zur Beurteilung der Retrievalqualität genutzt werden.

Im Allgemeinen[63] gilt: Die Treffergenauigkeit wird über den Wert Precision angegeben, für den die Anzahl der in der Treffermenge gefundenen relevanten Dokumente durch die Anzahl aller in der Treffermenge gefundenen Dokumente dividiert wird. Eine Precision von 1,0 würde bedeuten, dass alle in der Treffermenge enthaltenen Dokumente relevant sind.

[63] vgl. Fachrichtung Informationswissenschaft Saarbrücken, „6. Recall und Precision | Informationswissenschaft Saarbrücken Archiv“, Homepage. https://saar.infowiss.net/projekte/ident/themen/info_aufbereitung/recall/ (15. Mai 2018).

Die Treffervollständigkeit wird über den Wert Recall angegeben. Dazu wird die Anzahl der in der Treffermenge gefundenen relevanten Dokumente durch die Anzahl aller in der Kollektion zu dem Topic enthaltenen relevanten Dokumente dividiert. Ein Recall von 1,0 würde bedeuten, dass alle in der Kollektion enthaltenen relevanten Dokumente in der Treffermenge zu finden sind und eventuell noch andere Treffer. Ein Recall von 1,0 wird also z.B. immer dann ausgegeben, wenn der gesamte Korpus in der Treffermenge enthalten ist. Wenn die Werte von Precision und Recall beide 1,0 wären, bestände die Treffermenge aus allen relevanten in der Kollektion enthaltenen Dokumenten. Diese Konstellation ist allerdings äußerst unrealistisch.
Anhand dieser beiden Werte können viele andere, zur Bewertung von Retrieval geeignete Werte errechnet werden. Die sogenannte R-Precision misst beispielsweise die Precision an der Stelle R in der Treffermenge, wobei R die Anzahl aller in der Kollektion zu dem Topic enthaltenen relevanten Dokumente ist. Auch die Precision der Treffermenge nach einer bestimmten Anzahl von Treffern – z.B. nach 5, 10, 20, 50 oder 100 Dokumenten – kann für die Evaluation von Interesse sein. So kann ermittelt werden, ob das Ranking innerhalb der Treffermenge angemessen ist oder nicht.

Das Programm trec_eval berechnet viele Werte, die nicht alle für jeden Retrievaltest benötigt werden. Die Abbildung 8 aus Seite 42 ist ein Beispiel für die Ausgabeform der von trec_eval berechneten Werte.

```
unknown48d705b3d465:trec_eval.9.0 JohannaMunkelt$ ./trec_eval qrels19test.txt select191.txt
runid                   all     simpleRun
num_q                   all     1
num_ret                 all     19
num_rel                 all     10
num_rel_ret             all     9
map                     all     0.6075
gm_map                  all     0.6075
Rprec                   all     0.7000
bpref                   all     0.7800
recip_rank              all     0.5000
iprec_at_recall_0.00    all     0.7500
iprec_at_recall_0.10    all     0.7500
iprec_at_recall_0.20    all     0.7500
iprec_at_recall_0.30    all     0.7500
iprec_at_recall_0.40    all     0.7500
iprec_at_recall_0.50    all     0.7500
iprec_at_recall_0.60    all     0.7500
iprec_at_recall_0.70    all     0.7273
iprec_at_recall_0.80    all     0.7273
iprec_at_recall_0.90    all     0.6000
iprec_at_recall_1.00    all     0.0000
P_5                     all     0.6000
P_10                    all     0.7000
P_15                    all     0.6000
P_20                    all     0.4500
P_30                    all     0.3000
P_100                   all     0.0900
P_200                   all     0.0450
P_500                   all     0.0180
P_1000                  all     0.0090
```

Abbildung 8: Beispiel für eine Ausgabe in trec_eval.

5.1. Durchführung des Retrievaltests

Zum Nachweis der Eignung der Testkollektion zur Durchführung eines Retrievaltests wurde ein beispielhafter Retrievaltest konzipiert. Da der Test nur der Überprüfung der Testkollektion dient, gibt es keine konkrete These, die überprüft werden soll. Angelehnt an den ursprünglichen Zweck der Testkollektion als Basis für einen Retrievaltest, der die Qualität der Inhaltserschließung in der DNB prüfen soll, wurden für diesen beispielhaften Test fünf Suchanfragen für ein Topic (Topic 13 – Bundestagswahl) formuliert. Die Suchanfragen sind jeweils identisch und werden für jeden Durchlauf in unterschiedlichen Kategorien oder unterschiedlichen Kategorienkombinationen ausgeführt. Die Suchanfrage besteht dabei immer aus dem Wort „Bundestagswahl“.

Die erste Suchanfrage fordert eine Ergebnismenge aus Treffern an, in der das Suchwort in der Titel-Kategorie oder der Kategorie für intellektuelle Schlagworte (GND-Schlagworte) zu finden ist. Für die zweite Suchanfrage werden Treffer aus der Titel-Kategorie mit Treffern aus der Kategorie für maschinelle Schlagworte als Ergebnismenge ausgegeben. Eine Suche in vier Kategorien ergab die Treffermenge für Suchanfrage 3. Hierzu wurden Titel, GND-Schlagworte, maschinelle Schlagworte und die Informationen der Verlage nach

dem Suchwort durchsucht. Suchanfrage 4 sucht allein in den GND-Schlagworten, Suchanfrage 5 analog dazu allein in den maschinellen Schlagworten. In Abbildung 9 ist für jeden Durchlauf die genaue Formulierung der Suchanfragen und die Größe der zugehörigen Treffermengen dokumentiert.

Jeder dieser Suchläufe ist mit Hilfe von trec_eval evaluiert worden. Abbildung 10, S. 45, dokumentiert für jeden Durchlauf folgende Werte:

- Anzahl der gefundenen Treffer (ret),
- Anzahl aller Korpus zu dem Topic enthaltenen relevanten Dokumente (rel),
- Anzahl der relevanten Dokumente, die in der Treffermenge enthalten sind,
- Recall (rel_ret dividiert durch ret),
- Precision (rel_ret dividiert durch rel),
- R-Precision (Rprec), Precision an der Stelle R (R=rel),
- Bpref, vergleicht, ob relevant bewertete Dokumente höher gerankt werden als nicht relevant bewertete Dokumente und schließt unbewertete Dokumente aus diesem Vergleich aus[64],
- Precision nach dem zehnten Dokument (P@10).

Durchlauf	Suchanfrage	Trefferanzahl oder: ret
1	title_txt_de:(Bundestagswahl) OR subject_gnd_txt_de:(Bundestagswahl)	40
2	title_txt_de:(Bundestagswahl) OR subject_auto_txt_de:(Bundestagswahl)	73
3	title_txt_de:(Bundestagswahl) OR subject_auto_txt_de:(Bundestagswahl) OR subject_gnd_txt_de:(Bundestagswahl) OR subject_vlb_txt_de:(Bundestagswahl)	84
4	subject_gnd_txt_de:(Bundestagswahl)	12
5	subject_auto_txt_de:(Bundestagswahl)	66

Abbildung 9: Tabelle der fünf Suchanfragen mit Ergebnismengen für den beispielhaften Retrievaltest.

64 Chris Buckley und Ellen M. Voorhees: Retrieval Evaluation with Incomplete Information. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2004, S. 25-32.

Durch-lauf	ret	rel	rel_ret	Recall	Precision	Rprec	Bpref	P@10
1	40	26	11	0,423	0,275	0,3462	0,2544	0,3000
2	73	26	16	0,615	0,219	0,3462	0,2500	0,4000
3	84	26	17	0,654	0,202	0,3462	0,2574	0,3000
4	12	26	3	0,115	0,25	0,1154	0,1036	0,3000
5	66	26	14	0,538	0,212	0,1923	0,1420	0,2000

Abbildung 10: Tabelle der fünf Suchanfragen mit von trec_eval berechneten Werten und Recall und Precision.

Recall und Precision werden von trec_eval nicht ausgegeben, diese Werte wurden ausgerechnet, auf die dritte Nachkommastelle gekürzt und eingefügt. Die Anzahl aller im Korpus zu dem Topic enthaltenen relevanten Dokumente (rel) beträgt 26 und ist in diesem Fall konstant, da nur innerhalb eines Topics gesucht wurde. Für einen ausführlichen Retrievaltest werden in der Regel viele Topics durchsucht. Dieser Umstand hat auch Auswirkungen auf die R-Precision, die in diesem Fall entsprechend immer nach dem 26. Dokument gemessen wird.
Alle von trec_eval ausgegebenen Werte sind im Anhang 5 zu finden.

Diese Werte können miteinander verglichen werden. Zu beachten ist hierbei, dass es sich ausschließlich um Werte zu einem von vorhandenen 50 Topics handelt und dass die Suchanfragen sich nur durch den Einsatz verschiedener Kategorien oder Kategorien-Kombinationen unterscheiden. Es wurden keine unterschiedlich formulierten Suchanfragen genutzt. Aus dem Vergleich dieser Werte der fünf Durchläufe können keine validen Aussagen über die Qualität der Inhaltserschließungsmethoden für die gesamte Kollektion getroffen werden. Dieser beispielhafte Retrievaltest bietet keine Anhaltspunkte für eine Bewertung der veränderten Sacherschließungspraxis der DNB, es können lediglich Aussagen zu der Erschließung dieses speziellen Topics getroffen werden.

Ein Vergleich würde in diesem Fall vor allem Aussagen über den Einsatz der verschiedenen Kategorien bei der Suche mit Solr zulassen. Durchlauf 3 durchsucht vier Kategorien und bildet aus den Treffermengen zu jeder Kategorie eine Gesamttreffermenge. Diese Gesamttreffermenge ist mit 84 Dokumenten die größte der fünf Durchläufe. Dies war zu erwarten, da in keinem anderen Durchlauf vier Kategorien durchsucht wurden. In Durchlauf 3 wurden auch die meisten relevanten Treffer ausgegeben (17), das führt zum höchsten Recall im Vergleich. Die Precision hingegen ist am niedrigsten. Viele Treffer führen meist zu einem hohen Recall, speisen aber auch nicht relevanten Ballast in die Treffermenge mit ein und haben so häufig einen negativen Einfluss auf die Precision. Auffällig ist, dass der Recall der Durchläufe 2, 3 und 5 um den Wert 0,6 gruppiert ist und die Precision bei circa. 0,21 liegt; während der Recall für 1 und 4 deutlich schlechter ist, die Precision aber nicht unter 0,25 fällt. Dies untermauert die These, dass sich ein hoher Recall häufig negativ auf die Precision auswirkt oder andersherum: dass sich eine höhere Precision scheinbar negativ auf den Recall auswirkt.

Die leicht erhöhte Precision der Durchläufe 1 und 4 lässt darauf schließen, dass die Kategorienkombination zwar zu weniger, aber etwas besser geeigneten Treffern führt. In den Durchläufen 1 und 4 wurde jeweils mit den GND-Schlagworten gesucht, in Durchlauf 1 wurde die Treffermenge zusätzlich mit der Treffermenge aus der Titelkategorie kombiniert.

Bei genauerer Betrachtung des Wertes R-Precision fällt auf, dass dieser für die ersten drei Durchläufe konstant ist. Die einzige Gemeinsamkeit dieser Durchläufe ist die Suche in der Titelkategorie. Daraus lässt sich folgern, dass die Ergebnisdokumente der Suche in der Titelkategorie höher gerankt werden als die Ergebnisdokumente aus den anderen Kategorien, weil dort die höchste Übereinstimmung von Suchanfrage und Kategorien-Inhalten zu finden ist. Auch die Werte für Bpref stützen diese Folgerung: Sie liegen für die ersten drei Durchläufe bei etwa 0,25, und nehmen in den Durchläufen 4 und 5 erkennbar ab. Die Suche in der Titelkategorie oder die Suche einer Kategorien-Kombination aus Titel und anderen Kategorien führt für dieses Topic zu etwas besseren Ergebnissen als die Suche in den Schlagwortkategorien allein.

Aussagen über die Qualität der maschinellen Inhaltserschließung lassen sich anhand dieses Beispiels kaum treffen, da zwar der Recall in den Durchläufen, in denen in maschinell erzeugten Schlagworten gesucht wurde, höher ist als in den anderen Durchläufen, die Precision gleichzeitig aber abgenommen hat. Letztlich muss immer entschieden werden, wie viele Ballastdokumente in der Treffermenge in Kauf genommen werden, wenn gleichzeitig möglichst viele relevante Dokumente in der Treffermenge zu finden sein sollen. Die hier ermittelten Werte beweisen in erster Linie, dass die erstellten Relevanzurteile mit den fünf Ergebnislisten für das Topic 13 - Bundestagswahl in der Evaluation mit einem international eingesetzten Tool zu untereinander vergleichbaren Werten führen. Somit ist die Testkollektion für Retrievaltests geeignet.

5.2. Evaluation

Der beispielhafte Retrievaltest wurde durchgeführt und die in Kapitel 5.1 dokumentierten Ergebnisse dieses Tests lassen darauf schließen, dass die erarbeitete Testkollektion einsatzfähig ist. Der Korpus konnte mit 200.000 Dokumenten vollständig in die Suchplattform Solr übertragen werden und mittels eines abgeänderten Pooling-Verfahrens wurden Ergebnispools für die zuvor ausgewählten 50 Topics kombiniert. Die Relevanzurteile wurden für jeden Pool und alle darin enthaltenen Dokumente gefällt.
Die daraus resultierende Ergebnisliste umfasst 6.984 Relevanzurteile. Zum Nachweis der Machbarkeit wurden die Relevanzurteile und fünf Ergebnislisten, die zuvor über ausgewählte Suchanfragen in Solr erstellt wurden, mit Hilfe des Programms trec_eval evaluiert. trec_eval berechnete Werte, die plausibel erscheinen (s. Kapitel 5.1).

Die Testkollektion kann den ihr ursprünglich zugedachten Retrievaltests als Grundlage dienen. Durch die Dokumentation aller Arbeitsschritte und Gedankengänge ist die Konstruktion der Testkollektion als solche nachvollziehbar gemacht worden. Alle Daten sind statisch, die Nachnutzung der Kollektion mit den enthaltenen Dokumenten, Topics und Relevanzurteilen ist

möglich. Auf der beigelegten DC-ROM sind zusätzlich zu der PDF-Version dieser Arbeit und einem Abstract folgende Bestandteile enthalten:

- der Dokumentenkorpus mit 200.000 Titelaufnahmen und gefilterten Kategorien im XML-Format,
- die managed-schema-Textdatei zur Anpassung der Indexierungsstrategie in Solr,
- die solr-rel.xsl-Textdatei zur Umwandlung der Solr-Ergebnisliste in das Standard-TREC-Format,
- eine Excel-Tabelle mit 50 Topics, Topic-ID, Titel, Description und Narrative,
- eine Textdatei, die alle 50 Topics im XML-Format enthält,
- eine Excel-Tabelle, die alle Suchanfragen für jedes Topic enthält, die zur Pool-Erstellung genutzt wurden,
- eine Textdatei mit allen 6.984 Relevanzbewertungen.

Sowohl Solr als auch trec_eval sind Open-Source-Programme, die frei genutzt werden können. Die Testkollektion kann wiederverwendet, in andere Systeme übertragen oder an die Anforderungen anderer Retrievaltests angepasst werden. Das breite Sachgruppenspektrum begünstigt die Eignung der Testkollektion als allgemeine Testkollektion mit bibliothekarischen Metadaten in deutscher Sprache, auch die Größe der Kollektion erleichtert die Wiederverwendung. Der Dokumentenkorpus ist ca. 400MB groß und kann angemessen über elektronische Schnittstellen übertragen werden.

Der primäre Einsatzbereich der Testkollektion wird der Retrievaltest sein, der die Qualität der inhaltserschließenden Merkmale der verschiedenen Erschließungsstrategien der DNB evaluiert, denn zu diesem Zweck wurde die Testkollektion entworfen. Der beispielhafte Retrievaltest beweist, dass die Testkollektion für diesen Zweck geeignet ist. Der Retrievaltest wird von einer Projektgruppe von Herrn Prof. Dr. Klaus Lepsky und Herrn Prof. Dr. Philipp Schaer im Sommersemester 2018 an der TH Köln entworfen und durchgeführt (Stand: 16. Mai 2018).

6. Fazit

Der in Kapitel 5 erfolgte Nachweis der Machbarkeit ermöglicht bezugnehmend auf die Forschungsfragen, die zu Beginn dieser Arbeit gestellt wurden, folgende Aussage: Es ist möglich, eine Testkollektion zu konstruieren, die der Überprüfung der Sacherschließungsqualität dient. Die bisherigen Resultate lassen die Annahme zu, dass die erstellte Kollektion für den geplanten Retrievaltest eingesetzt wird und für die Konzipierung des Tests eine solide Basis darstellt.

Die Konstruktion von geeigneten Testkollektion ist stets eine Herausforderung im Information Retrieval. Vor allem die Erstellung von Relevanzurteilen stellt einen großen Arbeitsaufwand dar. In dieser Disziplin ist ausreichend Raum für weitere Forschungen und Experimente. Bisher konnte noch kein Ansatz die bei TREC praktizierte Vorgehensweise durch eine ressourcenfreundlichere Strategie ersetzen, ohne dass ein allzu großer Qualitätsverlust in Kauf genommen werden muss.

Erst nach der Durchführung des Tests kann beurteilt werden, inwieweit die Testkollektion tatsächlich für ihren Einsatzzweck geeignet ist. Während eines Retrievaltests können auch bei sorgfältiger Planung Komplikationen auftreten, die Einfluss auf die Leistung der Kollektion und die Ergebnisse des Tests haben. Inwiefern die maschinelle Inhaltserschließung bibliothekarischen Ansprüchen bereits gewachsen ist, kann folglich erst nach der Evaluation einer intensiven Testphase beantwortet werden.

Die erarbeitete Testkollektion sowie der funktionsprüfende Retrievaltest haben nicht den Anspruch, eine Handlungsempfehlung dazu auszusprechen, ob Menschen bei der Inhaltserschließung von Medienwerken durch Maschinen ersetzt werden können oder sollen. Die Testkollektion legt lediglich den Grundstein dafür, das Leistungsvermögen menschlicher und maschineller Erschließung besser miteinander vergleichen zu können.

Die Autorin ist der Ansicht, dass Menschen absehbar im Prozess des Qualitätsmanagements als Arbeitskräfte erhalten bleiben und nicht vollständig einer maschineller Inhaltserschließung weichen werden. Nach aktuellem Stand liefern automatisierte Erschließungen offenbar nur teilweise zufriedenstellende Ergebnisse. Zu dieser Einsicht scheint auch die DNB gelangt zu sein, da sie die Erschließungspraxis bisher nur für die Reihen B und H geändert hat, nicht aber für Reihe A (Monografien und Periodika des Verlagsbuchhandels). Das maschinelle Verfahren scheint für diese Reihe noch keine angemessenen Inhaltserschließungsmerkmale zu produzieren oder sich aus weiteren Gründen nicht dafür zu eignen. Das Qualitätsmanagement und die Aktualisierung der Schlagwortnormdatei, aus der sich die maschinelle Erschließung bedient, unterliegen dort weiterhin der Pflege durch Mitarbeiter.

7. Literaturverzeichnis

Behnert, Christiane, und Dirk Lewandowski: A framework for designing retrieval effectiveness studies of library information systems using human relevance assessments. In: Journal of Documentation 73 (2017) 3, S. 509–527. https://doi.org/10.1108/JD-08-2016-0099 (15. Mai 2018).

Buckley, Chris, und Ellen M. Voorhees: Retrieval Evaluation with Incomplete Information. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2004),S. 25–32. https://doi.org/10.1145/1008992.1009000 (15. Mai 2018).

Carterette, Ben, James Allan, und Ramesh Sitaraman: Minimal Test Collections for Retrieval Evaluation. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2006), S. 268–275. https://doi.org/10.1145/1148170.1148219 (15. Mai 2018).

Ceynowa, Klaus: Deutsche Nationalbibliothek: In Frankfurt lesen jetzt zuerst Maschinen. In: Frankfurter Allgemeine Zeitung vom 31. Juli 2017. http://www.faz.net/1.5128954 (15. Mai 2018)

Clough, Paul, und Mark Sanderson: Evaluating the Performance of Information Retrieval Systems Using Test Collections. In: Information Research 18 (2013) 2. http://www.informationr.net/ir/18-2/paper582.html#.U2unLPldXTp (15. Mai 2018).

Cormack, Gordon V., Christopher R. Palmer, und Charles L. A. Clarke: Efficient Construction of Large Test Collections. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (1998), S. 282–289. https://doi.org/10.1145/290941.291009 (15. Mai 2018).

„Deutsche Nationalbibliothek - Inhaltserschließung - Änderung der Inhaltserschliessung in den Metadaten der Deutschen Nationalbibliografie ab 1. September 2017“. Webseite der DNB (2017) http://www.dnb.de/DE/Erwerbung/Inhaltserschliessung/aenderungInhaltserschliessungSeptember2017.html (15. Mai 2018).

„Deutsche Nationalbibliothek - Inhaltserschließung - Grundzüge und erste Schritte der künftigen inhaltlichen Erschliessung von Publikationen in der Deutschen Nationalbibliothek“. Webseite der DNB (Mai 2017). http://www.dnb.de/DE/Erwerbung/Inhaltserschliessung/grundzuegeInhaltserschliessungMai2017.html (15. Mai 2018).

Gömpel, Renate, Ulrike Junger, und Elisabeth Niggemann. Veränderungen im Erschließungskonzept der Deutschen Nationalbibliothek. In: Dialog mit Bibliotheken 2010 (2010), 1, S. 20–22.

Heck, Tamara, und Philipp Schaer: Performing Informetric Analysis on Information Retrieval Test Collections: Preliminary Experiments in the Physics Domain. In: 14th International Society of Scientometrics and Informetrics Conference ISSI (2013), S. 1392-1400. http://arxiv.org/abs/1306.1743 (15. Mai 2018).

Kinney, Kenneth A., Scott B. Huffman, und Juting Zhai: How Evaluator Domain Expertise Affects Search Result Relevance Judgments. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management (2008), S. 591–598. https://doi.org/10.1145/1458082.1458160 (15. Mai 2018).

Kluck, Michael: Die GIRT-Testdatenbank als Gegenstand informationswissenschaftlicher Evaluation. In: Informationen zwischen Kultur und Marktwirtschaft. Proceedings des 9. InternationalenSymposiums für Informationswissenschaft ISI (2004), S. 247–268.

Lykke, Marianne, Birger Larsen, Haakon Lund, und Peter Ingwersen: Developing a Test Collection for the Evaluation of Integrated Search. In: Advances in Information Retrieval (2010), S. 627–630. https://doi.org/10.1007/978-3-642-12275-0_63 (15. Mai 2018).

Manning, Christopher D, Prabhakar Raghavan, und Hinrich Schütze: Introduction to Information Retrieval. Cambridge: Cambridge University Press, 2008.

„NIST - TIPSTER Text Program - Overview“. Webseite NIST. https://www-nlpir.nist.gov/related_projects/tipster/ (15. Mai 2018).

Ritchie, Anna, Simone Teufel, und Stephen Robertson: Creating a Test Collection for Citation-based IR Experiments. In: Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (2006), S. 391–398. https://doi.org/10.3115/1220835.1220885 (15. Mai 2018).

„6. Recall und Precision | Informationswissenschaft Saarbrücken Archiv“. Webseite Saarbrücken, Fachrichtung Informationswissenschaft. Zugegriffen 12. Mai 2018. https://saar.infowiss.net/projekte/ident/themen/info_aufbereitung/recall/ (15. Mai 2018).

Sachse, Elisabeth, Martina Liebig, und Winfried Gödert: Automatische Indexierung unter Einbeziehung semantischer Relationen: Ergebnisse des Retrievaltests zum MILOS II-Projekt. Köln: Fachhochschule Köln, Fachbereich Bibliotheks- und Informationswesen, 1998.

Sanderson, Mark: Test Collection Based Evaluation of Information Retrieval Systems. In: Foundations and Trends® in Information Retrieval 4 (2010) 4, S. 247–375. https://doi.org/10.1561/1500000009 (15. Mai 2018).

Sanderson, Mark, und Hideo Joho: Forming Test Collections with No System Pooling. In: Proceedings of the 27th Annual International ACM SIGIR

Conference on Research and Development in Information Retrieval (2004), S. 33–40. https://doi.org/10.1145/1008992.1009001 (15. Mai 2018).

Scholer, Falk, Diane Kelly, und Ben Carterette: Information Retrieval Evaluation Using Test Collections. In: Information Retrieval Journal 19 (2016) 3, S. 225–29. https://doi.org/10.1007/s10791-016-9281-7 (15. Mai 2018).

Spärck Jones, Karen, und Cornelius J. van Rijsbergen: Report on the Need for and Provision of an „ideal“ Information Retrieval Test Collection. Cambridge: University Computer Laboratory, 1975.

„Text REtrieval Conference (TREC) Overview“. Website TREC. https://trec.nist.gov/overview.html (15. Mai 2018).

Uhlmann, Sandro: Automatische Beschlagwortung von deutschsprachigen Netzpublikationen mit dem Vokabular der Gemeinsamen Normdatei (GND). Dialog mit Bibliotheken 2013 (2013) 2, S. 26–36.

Voorhees, Ellen M.: TREC: Continuing Information Retrieval's Tradition of Experimentation. In: Communications of the ACM (2007), S. 51–54. https://doi.org/10.1145/1297797.1297822 (15. Mai 2018).

Ellen M. Voorhees: Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (1998), S.315-323. https://doi.org/10.1145/290941.291017 (15. Mai 2018).

Voorhees, Ellen M. und Donna K. Harman: TREC Experiment and Evaluation in Information Retrieval. Cambridge, Mass.: MIT Press, 2005.

Wiesenmüller, Heidrun: Das neue Sacherschließungskonzept der DNB in der FAZ. Basiswissen RDA (Blog), 2. August 2017. http://www.basiswissen-rda.de/neues-sacherschliessungskonzept-faz/ (15. Mai 2018).

8. Anhang

Anhang 1: Aufschlüsselung der Sachgruppenverteilung des Korpus' der DNB

Sachgruppen	Fachgebiete	Verteilung
0	Allgemeines, Wissenschaft	0.78%
4	Informatik	1.27%
10	Bibliografien	0.19%
20	Bibliotheks- und Informationswissenschaft	0.68%
30	Enzyklopädien	0.09%
50	Zeitschriften, fortlaufende Sammelwerke	0.08%
60	Organisationen, Museumswissenschaft	0.25%
70	Nachrichtenmedien, Journalismus, Verlagswesen	1.06%
80	Allgemeine Sammelwerke	0.06%
90	Handschriften, seltene Bücher	0.13%
100	Philosophie	4.41%
130	Parapsychologie, Okkultismus	1.70%
150	Psychologie	2.06%
200	Religion, Religionsphilosophie	1.09%
220	Bibel	0.90%
230	Theologie, Christentum	1.55%
290	Andere Religionen	1.19%
300	Sozialwissenschaften, Soziologie, Anthropologie	1.76%
310	Allgemeine Statistiken	0.05%
320	Politik	2.06%
330	Wirtschaft	1.17%
333.7	Natürliche Ressourcen, Energie und Umwelt	1.14%
340	Recht	1.05%
350	Öffentliche Verwaltung	0.59%
355	Militär	0.61%
360	Soziale Probleme, Sozialdienste, Versicherungen	1.58%
370	Erziehung, Schul- und Bildungswesen	2.04%
380	Handel, Kommunikation, Verkehr	0.96%
390	Bräuche, Etikette, Folklore	0.96%
400	Sprache, Linguistik	0.96%
420	Englisch	0.64%
430	Deutsch	1.31%
439	Andere germanische Sprachen	0.17%
440	Französisch, romanische Sprachen allgemein	0.42%
450	Italienisch, Rumänisch, Rätoromanisch	0.22%
460	Spanisch, Portugiesisch	0.41%
470	Latein	0.12%
480	Griechisch	0.07%
490	Andere Sprachen	1.11%

Sachgruppen	Fachgebiete	Verteilung
491.8	Slawische Sprachen	0.22%
500	Naturwissenschaften	0.46%
510	Mathematik	1.45%
520	Astronomie, Kartografie	0.52%
530	Physik	1.44%
540	Chemie	1.50%
550	Geowissenschaften	0.83%
560	Paläontologie	0.10%
570	Biowissenschaften, Biologie	1.64%
580	Pflanzen (Botanik)	0.39%
590	Tiere (Zoologie)	0.90%
600	Technik	0.29%
610	Medizin, Gesundheit	5.87%
620	Ingenieurwissenschaften und Maschinenbau	1.23%
621.3	Elektrotechnik, Elektronik	0.92%
624	Ingenieurbau und Umwelttechnik	0.90%
630	Landwirtschaft, Veterinärmedizin	1.10%
640	Hauswirtschaft und Familienleben	1.01%
650	Management	1.43%
660	Technische Chemie	0.93%
670	Industrielle und handwerkliche Fertigung	0.93%
690	Hausbau, Bauhandwerk	0.90%
700	Künste, Bildende Kunst allgemein	1.16%
710	Landschaftsgestaltung, Raumplanung	0.82%
720	Architektur	1.22%
730	Plastik, Numismatik, Keramik, Metallkunst	0.89%
740	Grafik, angewandte Kunst	1.07%
741.5	Comics, Cartoons, Karikaturen	0.46%
750	Malerei	1.33%
760	Druckgrafik, Drucke	0.44%
770	Fotografie, Video, Computerkunst	1.45%
780	Musik	1.00%
790	Freizeitgestaltung, Darstellende Kunst	0.08%
791	Öffentliche Darbietungen, Film, Rundfunk	1.76%
792	Theater, Tanz	0.87%
793	Spiel	0.92%
796	Sport	1.08%
800	Literatur, Rhetorik, Literaturwissenschaft	1.02%
810	Englische Literatur Amerikas	0.34%

Sachgruppen	Fachgebiete	Verteilung
820	Englische Literatur	0.57%
830	Deutsche Literatur	1.97%
839	Literatur in anderen germanischen Sprachen	0.13%
840	Französische Literatur	0.47%
850	Italienische, rumänische, rätoromanische Literatur	0.12%
860	Spanische und portugiesische Literatur	0.37%
870	Lateinische Literatur	0.22%
880	Griechische Literatur	0.18%
890	Literatur in anderen Sprachen	0.33%
891.8	Slawische Literatur	0.15%
900	Geschichte	0.76%
910	Geografie, Reisen	6.08%
914.3	Geografie, Reisen (Deutschland)	1.21%
920	Biografie, Genealogie, Heraldik	0.91%
930	Alte Geschichte, Archäologie	1.21%
940	Geschichte Europas	1.46%
943	Geschichte Deutschlands	5.03%
950	Geschichte Asiens	0.50%
960	Geschichte Afrikas	0.16%
970	Geschichte Nordamerikas	0.27%
980	Geschichte Südamerikas	0.09%
990	Geschichte der übrigen Welt	0.01%

Anhang 2: Tabelle mit Topic-ID, Titel, Description und Narrative für 50 Topics

Topic-ID	Topic	Beschreibung	Narrative
1	Kritische Theorie	Finde Dokumente, die die Kritische Theorie thematisieren.	Relevante Dokumente berichten über die kritische Analyse einer kapitalistischen Gesellschaft aus der Perspektive der Frankfurter Schule. Relevant sind Dokumente, die Herrschafts- und Unterdrückungsmechanismen und Ideologien in einer kapitalistischen Gesellschaft untersuchen und darauf abzielen, eine vernünftigen Gesellschaft mündiger Menschen zu bilden. Teilweise relevant sind Dokumente, die Kritische Theorie in Teilaspekten, im Ausland oder als Stufe einer Entwicklung behandeln.
2	Ökologie der Gewässer	Finde Dokumente, die die Ökologie von Gewässern diskutieren.	Relevante Dokumente berichten über den Einfluss des Menschen auf die Gewässerökologie von sämtlichen Gewässerarten in Deutschland und Möglichkeiten zur Renaturierung. Teilweise relevante Dokumente befassen sich mit einem bestimmten Gewässer, Gewässern in Europa oder einer sehr speziellen Entwicklung in einem Gewässer, z.B. die Entwicklung einer bestimmten Art.
3	Gerontologie	Finde Dokumente, die sich mit Altersforschung beschäftigen.	Relevante Dokumente berichten über die interdisziplinäre Forschung der mit dem Altern verbundenen Phänomene, Probleme und Ressourcen wie z.B. der demographische Wandel. Auch relevant sind Dokumente, die sich mit alternsrelevanten und alternskonstituierenden Umwelten und sozialen Umwelten im Alter befassen. Teilweise relevant sind Dokumente, die über sehr spezifische Bereiche berichten, die vom Alter beeinflusst werden.
4	Frau und Beruf	Finde Dokumente, die das Thema Frau und Beruf beleuchten.	Relevante Dokumente diskutieren die Probleme von Frauen in der Arbeitswelt und die Entwicklung der Thematik. Teilweise relevant sind Dokumente, die nur bestimmte Berufsgruppen thematisieren oder den Konflikt zwischen Arbeit und Privatleben/Familie für Frauen diskutieren. Nicht relevant sind Ratgeber jedweder Art.
5	Kunst in der Antike	Finde Dokumente, die über die Kunst in der Antike berichten.	Relevante Dokumente diskutieren die Entstehung und Erstellung von Kunst in der Antike (Epoche im Mittelmeerraum, etwa von 800 v. Chr. bis ca. 600 n Chr.). Auch relevant sind Dokumente die sich mit den Künstlern in der Antike auseinandersetzen. Nicht relevant sind Ratgeber für den Kauf/Verkauf antiker Kunst.
6	Angst, Angststörungen und Angsterkrankungen	Finde Dokumente, die sich mit Angst und Angsterkrankungen beschäftigen.	Relevante Dokumente diskutieren Angst und Angststörungen sowie deren Behandlung. Auch relevant sind Dokumente, die sich mit der Bildung von Phobien befassen. Teilweise relevant sind Dokumente, die nur auf einen sehr spezifischen medizinischen Aspekt eingehen. Nicht relevant sind Dokumente zu Essstörungen oder der Angst vor dem Tod.
7	Festkörperphysik	Finde Dokumente, die sich mit Festkörperphysik befassen.	Relevante Dokumente diskutieren die verschiedenen Eigenschaften von Festkörprern und deren Verhalten. Die Organisation und Beschreibung dieser ist ebenfalls relevant. Teilweise relevant sind Dokumente, in denen Festkörperphysik als eines der Grundlagenthemen erläutert wird.
8	Tierexperimente	Finde Dokumente, die Tierversuche thematisieren.	Relevante Dokumente thematisieren Tierversuche/Tierexperimente, vor allem kritische Auseinandersetzungen damit. Nicht relevant sind andere medizinische Studien, in denen Tierversuche gemacht wurden.
9	Widerstand im Nationalsozialismus	Finde Dokumente, die den Widerstand gegen den Nationalsozialismus thematisieren.	Relevante Dokumente berichten über den Widerstand von Personen, Gruppen und Einrichtungen im Gebiet der von der Wehrmacht besetzten Staaten und Deutschlands während der Zeit des Nationalsozialismus. Teilweise relevante Dokumente handeln vom Widerstand sehr kleiner Gruppierungen, bestimmten Ortskreisen oder sind die Biographien zu einzelnen Personen, die im Widerstand tätig waren. Nicht relevante Dokumente handeln vom Widerstand der Nazis in Kriegen.
10	Psychosomatische Krankheiten	Finde Dokumente über Psychosomatische Krankheiten.	Relevante Dokumente diskutieren körperliche Beschwerden/Krankheiten, die durch psychische Einflüsse/Belastung ausgelöst oder unterstützt werden. Teilweise relevant sind Dokumente zu psychsomatischen Essstörungen. Nicht relevant sind Dokumente, die sich mit der Somatophsychologie befassen, dem Elnfluss von körperlichen Beschwerden auf psychische Gesundheit.
11	Umweltökonomie/ -ökonomik	Finde Dokumente aus dem Themenbereich "Umweltökonomie".	Relevante Dokumente behandeln den Einfluss industriellen Wirtschaftens auf die Umwelt, setzen sich mit ressourcenorientierter Umwelts- oder Wirtschaftpolitik auseinander oder behandeln einzelne Umweltsysteme, die unter wirtschaftlichen Einfluss stehen (z.B. Photovoltaik). Teilweise relevante Dokumente berichten über eine größeren Zusammenhang, in dem Umweltökonomie nur einen Teilaspekt darstellt.

Topic-ID	Topic	Beschreibung	Narrative
12	Sprachwandel	Finde Dokumente, die die Entwicklung einer Sprache untersuchen.	Relevante Dokumente behandeln Sprachwandel/die Sprachdynamik in Deutschland, die Entwicklung und die Auswirkungen. Teilweise relevante Dokumente sind Dokumente, die sich mit einem der Faktoren für Sprachwandel befassen, nur ein sehr spezielles Teilgebiet behandeln oder Festschriften, die Sprachwandel thematisieren. Nicht relevant sind Dokumente, die sich mit dem Sprachwandel in anderen Ländern befassen.
13	Bundestagswahl	Finde Dokumente, die die Bundestagswahl Deutschlands betreffen.	Relevante Dokumente zeigen Abläufe, Prozesse und Gesetze einer Bundestagswahl in Detuschland auf. Teilweise relevante Dokumente diskutieren mögliche Einflüsse auf die Ergebnisse einer Bundestagswahl im Allgemeinen. Nicht relevante Dokumente beziehen sich auf ein bestimmtes Ereignis/Ergebnis einer Bundestagswahl, befassen sich mit der Wahlkampfstrategie einer/mehrerer Parteien für eine Bundestagswahl oder thematisieren Reaktionen aus dem Ausland auf Ergebnisse einer Bundestagswahl.
14	Alternative Energiequellen	Finde Dokumente, die die Nutzung alternativer Energiequellen behandeln.	Relevante Dokumente zeigen die Notwendigkeit des Nutzens erneuerbarer Energiequellen im Vergleich zu fossilen Energieträgern auf und befassen sich mit der Forschung dazu. Teilweise relevante Dokumente behandeln den Einfluss erneuerbarer Energie auf einen sehr spezifischen Bereich. Nicht relevante Dokumente thematisieren die Nutzung von Photovoltaik für Privatpersonen.
15	Wirtschaftsethik	Finde Dokumente, die das Thema Wirtschaftsethik beleuchten.	Relevante Dokumente handeln von Wirtschaftsethik im Allgemeinen, beschreiben die Methoden und Ausprägungen. Teilweise relevante Dokument thematisieren Wirtschafsethik in bestimmten Unternehmenskreisen oder sind allgemein gehalten, aber in englischer Sprache verfasst.
16	Hyperaktivität bei Kindern	Finde Dokumente, die das Krankheitsbild der Hyperaktivität bei Kindern untersuchen.	Relevante Dokumente beschreiben verschiedene Formen der Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung, kurz ADHS, und diskutieren verschiedene Ursachen, Ausprägungen und Therapie- und Behandlungsmöglichkeiten bei Kindern. Teilweise relevante Dokumente befassen sich mit einer medizinischen Erklärung für ADHS oder befassen sich mit ADHS im Schulalltag. Nicht relevante Dokumente befassen sich mit ADHS bei Erwachsenen oder anderen Krankheiten.
17	Direktmarketing	Finde Dokumente, die direkte Kundenansprache durch ein Unternehmen thematisieren.	Relevante Dokumente behandeln Taktiken und geben Beispiele für Unternehmen für das Direktmarketing. Nicht relevante Dokumente befassen sich mit anderen Marketingmethoden, beispielsweise Kundenakquise über Social-Media-Marketing.
18	Behandlung bei Schlaganfällen	Finde Dokumente, die Behandlungsmöglichkeiten bei Schlaganfällen aufzeigen.	Relevante Dokumente beschreiben die Behandlung eines Schlaganfallpatienten im Alltag nach seinem Krankenhausaufenthalt. Teilweise relevant sind medizinische Studien zur Behandlung von Schlaganfällen.
19	Zimmerpflanzen	Finde Dokumente, die sich mit Zimmerpflanzen befassen.	Relevante Dokumente beschreiben die Artenvielfalt, Aufzucht und Pflege verschiedener Zimmerpflanzen und bieten Anleitungen für ihr langes Bestehen. Teilweise relevante Dokumente handeln von einer einzelnen Zimmerpflanze (z.B. Orchidee). Nicht relevante Dokumente beschäftigen sich mit Gartengestaltung, Gartenpflanzen oder ähnlichem.
20	Reiseführer für die Toskana	Finde Dokumente, die als Reiseführer für die Region Toskana fungieren.	Relevante Dokumente zählen Restaurants, Wanderwege, Aussichtspunkte und öffentliche Einrichtungen auf und bewerten diese und sind allgemeine Reiseführer für die Toskana. Teilweise relevante Dokumente behandeln einen bestimmten Ort/eine bestimmte Insel der Toskana mit der Toskana gemeinsam oder sind allgemeine Reiseführer für ganz Italien. Nicht relevante Dokumente sind Bildbände oder Reiseführer nur für einzelne Städte/Inseln in der Toskana, z.B. Florenz oder Elba.
21	Lateinamerikanische Tänze	Finde Dokumente, die sich mit lateinamerikanischen Tänzen beschäftigen.	Relevante Dokumente behandeln die verschiedenen Lateinamerikanischen Tänze und die Abgrenzung zum Standardtanz. Relevant sind Dokumente, wenn sie allgemein über lateinamerikanische Tänze berichten. Relevante Dokumente können auch einen speziellen Tanz aus dem lateinamerikanischen Raum betrachten. Teilweise relevant sind Dokumente, die lateinamerikanische Tänze in anderem Zusammenhang behandeln (Musik, Dichtung..).
22	Gesprächsführung	Finde Dokumente, die sich mit Gesprächsführung beschäftigen.	Relevante Dokumente behandeln Methoden, Gespräche und Diskussionen anzuleiten und zu moderieren. Teilweise relevant sind Hilfen zur Gesprächsführung bei Patienten mit einem bestimmten Krankheitsbild oder Dokumente, die Rhetorik und Seelsorge thematisieren. Nicht relevant sind Wortschatzbücher, Wörterbücher, Sprachführer.

Topic-ID	Topic	Beschreibung	Narrative
23	Islamischer Fundamentalismus	Finde Dokumente, die sich mit islamischem Fundamentalismus beschäftigen.	Relevante Dokumente behandeln die Gründe, Ausprägungen und Auswirkungen des islamischen Fundamentalismus und die in diesem Kontext auftretende Radikalisierung von Menschen bis hin zur Terrorbereitschaft. Teilweise relevante Dokumente thematisieren den Islamischen Fundamentalismus im Hinblick auf spezielle Ereignisse oder der Islamische Fundamentalismus taucht in einer Übersicht auf.
24	Geschichte Israels	Finde Dokumente, die sich mit der Geschichte des Staates Israel beschäftigen.	Relevante Dokumente behandeln die Geschichte des Staates Israel und die territorialen Dispute, die damit einhergehen. Ebenfalls relevant sind Dokumente, die sich mit dem historischen, kulturellen und religiösen Staat Israel befassen und dessen Bedeutung für den gegenwärtigen Staat und die Konflikte. Teilweise relevante Dokumente behandeln die Geschichte Israels im Hinblick auf einen bestimmten Einzelaspekt. Nicht relevant sind Dokumente, in denen das Wort Geschichte im Sinne von "Erzählung" vorkommt.
25	Heilfasten	Finde Dokumente, die sich mit Heilfasten beschäftigen.	Relevante Dokumente behandeln unterschiedliche Methoden von Nulldiäten und Heilfasten und der Wirkung auf die körperliche und seelische Gesundheit sowie den Risiken, die damit verbunden sind. Relevant sind auch Dokumente, die sich mit dem esotherischen Aspekt von Fasten und Diät beschäftigen. Teilweise relevant sind Dokumente, die das religiöse Fasten zum Thema haben.
26	Elektroautos	Finde Dokumente, die sich mit elektrisch betriebenen Fahrzeugen befassen.	Relevante Dokumente behandeln die Entwicklung von Kraftfahrzeugen, die mit Elektromotoren betrieben werden und die Bedeutung für die Energiewirtschaft. Relevant sind auch Dokumente, die sich mit dem Umweltaspekt elektrischer Motoren beschäftigen. Teilweise relevant sind Dokumente, die Elektromobilität im Allgemeinen mit dem Thema Elektroautos verbinden.
27	Homöopathische Mittel	Finde Dokumente, die sich mit homöopathischen Mitteln und ihren Wirkungen beschäftigen.	Relevante Dokumente behandeln die verschiedenen homöopathischen Wirkstoffe sowie die Diskussion um Wirksamkeit homöopathischer Mittel allgemein. Auch tiermedizinische Dokumente sind relevant.
28	Künstliche Intelligenz	Finde Dokumente, die sich mit der Entwicklung künstlicher Intelligenz befassen.	Relevante Dokumente behandeln die Entwicklung von Technologien im Bereich künstlicher Intelligenz und die Diskussion über die gesellschaftliche Bedeutung intelligenter Maschinen. Teilweise relevant sind Dokumente, die den Einsatz von künstlicher Intelligenz in sehr spezifischen Bereichen thematiseren. Nicht relevant sind Dokumente, die eine Darstellung von oder eine Auseinandersetzung mit künstlicher Intelligenz in Filmen oder Romanen enthalten.
29	Außenpolitik der USA	Finde Dokumente, die sich mit der Außenpolitik der USA befassen.	Relevante Dokumente behandeln die Außenpolitik der USA im 20. oder 21. Jahrhundert, ihr Einfluss auf das Weltgeschehen und den Wandel in der Politik über die Jahre. Teilweise relevant sind Dokumente, die allgemein die Geschichte der USA thematisieren oder die Außenpolitik der Vereinigten Staaten und deren Einfluss bezogen auf einen sehr spezifischen Aspekt erwähnen. Nicht relevant sind Dokumente, die sich mit der Außenpolitik anderer Länder auseinandersetzen.
30	Internetzeitalter	Finde Dokumente, die das Thema "Internetzeitalter" beinhalten.	Relevante Dokumente behandeln den Vorschmarsch des Internets, die Integration in den Alltag und die gesellschaftliche Bedeutung, die damit einhergeht, insbesondere der Einfluss auf Informationsverhalten und Meinungsbildung. Teilweise relevante Dokumente behandeln das Internet als Einflussfaktor auf ein sehr spezielles Themengebiet und das Thema soziale Netzwerke und deren Einfluss auf das einzelne Themen.
31	Kurden	Finde Dokumente, die sich mit dem Volk der Kurden beschäftigen.	Relevante Dokumente behandeln die Volksgruppe der Kurden, ihren Wirkungsraum, ihre Geschichte und aktuelle Berichte über Kurden. Auch relevant sind Dokumente, die sich mit der Diskussion um die Gründung eines Kurdenstaates befassen. Teilweise relevant sind Dokumente, in denen die Kurdenfragen als Einflussnehmer auf andere (politische) Entscheidungen anderer Länder fungiert. Nicht relevant sind Dokumente, die den Nahen Osten im Allgemeinen thematisieren.
32	Chaostheorie	Finde Dokumente, die die Chaostheorie diskutieren.	Relevant sind alle Dokumente, die die Chaostheorie, ein vager Bereich der Nichtlinearen Dynamik, im Sinne der Mathematischen Physik oder der angewandten Mathematik behandeln. (Chaostheorie: Ordnung in speziellen dynamischen Systemen, deren zeitliche Entwicklung unvorhersehbar ist; die Systeme sind sehr empfindlich von den Anfangsbedingungen abhängig; Beispiel: magnetisches Pendel). Teilweise relevant sind Dokumente, die die Chaostheorie und ihr Einfluss auf andere Sachgebiete thematisieren. Nicht relevant sind geschichtliche Abhandlungen.

Topic-ID	Topic	Beschreibung	Narrative
33	Magersucht	Finde Dokumente, die sich mit der Anorexia nervosa, der sog. Magersucht, befassen.	Relevant sind alle Dokumente, die die psychische Störung aus dem Bereich der seelisch bedingten Essstörungen und ihre Folgen sowie Behandlungsmöglichkeiten diskutieren. Nicht relevant sind Dokumente, die sich mit Anorexie (=Appetitlosigkeit) beschäftigen.
34	Politischer Skandal	Finde Dokumente, die politische Skandale thematisieren.	Relevant sind Dokumente, die sich mit rein politischen Skandalen und ihren Auswirkungen befassen, auch wenn sie Folgen für Wirtschaft oder andere Bereiche haben. Relevant sind auch Dokumente, in denen das Wort Affäre für eine skandalös beurteilte Angelegenheit in der Politik benutzt wird.
35	Entstehung und Entwicklung der Schrift	Finde Dokumente, die sich mit der Entstehung und Entwicklung der Schrift im Allgemeinen befassen.	Relevant sind Dokumente, die sich mit der Entstehung der Schrift als Verständigungsebene befassen, angefangen bei den Knoten-Verfahren, die in Südamerika entwickelt wurden. Teilweise relevant sind Dokumente, die sich mit der Entwicklung der heutigen Schrift und diversen Schriftarten befassen. Nicht relevant sind Dokumente, die sich mit der Entstehung der "Schrift", also der Bibel oder anderen religiösen/theologischen Schriften beschäftigen.
36	Das Individuum in der Gesellschaft	Finde Dokumente, die sich mit dem Individuum in der Gesellschaft befassen.	Relevant sind alle Dokumente, die ein Individuum (eine Person) im Kontext der Gesellschaft diskutieren. Die Auswirkungen einer Gesellschaft auf ein Individuum, auf seine Eigenschaften, Interessen und Besonderheiten sind mögliche Themen. Teilweise relevant sind Dokumente, die die Sicht einer bestimmten Person/eines bestimmten Philosophen auf diesen Bereich thematisieren.
37	Straßenkinder	Finde Dokumente, die sich mit obdachlosen Kindern beschäftigen.	Relevante Dokumente behandeln das Leben von Kindern und Jugendlichen unter 18 Jahren, die obdachlos, von zu Hause weggelaufen oder ohne Angehörige sind. Relevant sind auch Dokumente, die mit den Zukunftsaussichten für Straßenkinder beschäftigen. Nicht relevant sind Dokumente über Straßengangs.
38	Kriegsberichterstattung	Finde Dokumente, die sich mit Kriegsberichterstattung befassen.	Relevant sind alle Dokumente, die sich mit der journalistischen Berichterstattung in Medien über Kriege und kriegsähnliche Auseinandersetzungen und Konflikte beschäftigen. Relevant sind nicht nur Dokumente, die die politischen und militärischen Ereignisse zum Thema haben, sondern auch Dokumente, die Hintergrundberichte zu diplomatischen, humanitären und wirtschaftlichen Themen behandeln. Teilweise relevant sind Dokumente, die Kriegsberichterstattung vor dem Jahr 2000 thematisieren. Nicht relevant sind Dokumente, die eine Kriegsberichterstattung an sich sind (z.B. Berichte von Soldaten o.Ä.).
39	Medizin im Dritten Reich	Finde Dokumente, die sich mit der Medizin im Dritten Reich befassen.	Relevante Dokumente berichten über die von der nationalsozialistischen Politik geprägte Richtung der Medizin und die sog. Rassenhygiene, die unter anderem mit Zwangssterilisationen, Experimenten (s. Dr. Mengele) und Massenmord gepflegt wurde. Dokumente, die die Verfolgung von jüdischen Medizinern thematisieren, sind nicht relevant.
40	Theatergeschichte	Finde Dokumente, die sich mit der Geschichte des Theaters im Allgemeinen befassen.	Relevant sind alle Dokumente, die sich mit der Geschichte der szenischen Aufführung dramatischer Texte in einem Theater oder einer ähnlichen Elnrichtung befassen. Teilweise relevant sind Dokumente, die die Geschichte eines bestimmten Theaters über einen sehr langen Zeitraum, mindestens 100 Jahre, thematisieren. Nicht relevant sind Dokumente, die Stücke behandeln, die im Theater aufgeführt werden und das Wort "Geschichte(n)" im Titel enthalten.
41	Politik und Massenmedien	Finde Dokumente, die das Verhältnis von Politik und Massenmedien zueinander diskutieren.	Relevante Dokumente diskutieren die politische Beteiligung in einer Massendemokratie durch Presse, Funk, Fernsehen und Internet sowie die Konkurrenz privater mit öffentlich-rechtlichen Anbietern und benennen die Aufgaben der Massenmedien sowie verfassungsrechtliche Regelungen. Nicht relevant sind Dokumente, die den Themenkomplex Mediennutzung als Suchtproblem behandeln.
42	Wirtschaftswunder in Deutschland	Finde Dokumente, die sich mit dem Wirtschaftswunder in Deutschland befassen.	Relevante Dokumente berichten über das schnelle und nachhaltige Wirtschaftswachstum in Deutschland nach dem Zweiten Weltkrieg, welches derart unerwartet erfolgte, dass das Schlagwort „Wirtschaftswunder“ geprägt wurde. Relevant sind auch Dokumente, die das deutsche Wirtschaftswachstum als natürlichen Prozess im Sinne des Solow-Modells diskutieren.Teilweise relevant sind Dokumente, die das Wirtschaftswunder und die resultierenden Folgen auf den Lebensalltag als Erfahrungsbericht einzelner Personen beschreiben. Nicht relevant sind Dokumente, die das Wort Wirtschaftswunder als Zeitabschnittsmarkierung nutzen und einen Alltagsaspekt in einer Zeitspanne vor oder nach dem Wirtschaftswunder beleuchten.

Topic-ID	Topic	Beschreibung	Narrative
43	Spurenelemente in der Ernährung des Menschen	Finde Dokumente, die die Spurenelemente in der Ernährung des Menschen behandeln.	Relevante Dokumente berichten über die biologische Bedeutung von Spurenelementen für den Menschen sowie über ihr das Vorkommen in Nahrungsmitteln. Relevant sind auch Dokumente, die im Allgemeinen Spurenelemente definieren sowie auflisten. Nicht relevant sind Dokumente, die die Bedeutung von Spurenelementen in der Ernährung anderer Spezies als den Menschen behandeln ohne dass ein direkter Einfluss auf den Menschen entsteht.
44	Shakespeares Dramen	Finde Dokumente, die sich mit Shakespeares Dramen beschäftigen.	Relevante Dokumente behandeln die verschiedenen Dramen Shakespeares im Allgemeinen oder die Tragödien Shakespeares.Teilweise relevant sind Dokumente, die sich mit bestimmten Merkmalen der Dramen Shakespeares beschäftigen, ihren Einfluss auf andere Themenbereiche thematisieren oder Shakespeares Leben und seine Werke im Allgemeinen behandeln. Ebenfalls relevant sind Dokumente, welche die Rezeption in unterschiedlichen Kulturkreisen betrachten. Nicht relevant sind Dokumente, die sich mit bestimmten Aufführungen befassen.
45	Betriebspsychologie	Finde Dokumente, die Betriebspsychologie behandeln.	Relevante Dokumente analysieren Betriebspsychologie als Teilbereich der Wirtschaftspsychologie. Relevante Dokumente setzen sich auch mit dem Erleben und Verhalten von Menschen in betrieblichen Prozessen auseinander, beispielsweise den Arbeitsbedingungen, den Ursachen von Veränderungen der Produktivität oder bestimmten psychologischen Problemen, die sich in Betrieben stellen. Nicht relevant sind Dokumente, die sich mit der Untersuchung der Arbeitstätigkeiten selbst befassen.
46	Marktanalyse	Finde Dokumente, die das Thema Marktanalyse behandeln.	Relevante Dokumente thematisieren die statistische, zeitpunktbezogene Analyse eines bestimmten Marktes und beschäftigen sich mit der Momentaufnahme dieses Marktes. Teilweise relevante Dokumente sind die Dokumentation einer Marktanalyse für ein Unternehmen oder einen bestimmten Teilbereich. Nicht relevant sind Dokumente, die, in Abgrenzung zur vergangenheits- oder gegenwartsbezogenen Reflexion eines Marktes, Marktprognosen über künftige Markentwicklungen thematisieren.
47	Selbstbewusstsein stärken	Finde Dokumente, die Möglichkeiten aufzeigen, das Selbstbewusstsein zu stärken.	Relevante Dokumente berichten über allgemeine Möglichkeiten/Techniken das Selbstbewusstsein zu stärken. Teilweise relevante Dokumente behandeln Selbstbewusstsein bei Kindern, sind in englischer Sprache verfasst oder setzen sich mit Selbstbewusstsein im Zusammenhang mit der Philosophie auseinander.Nicht relevant sind Dokumente, welche psychiatrische Erkrankung thematisieren, die mit einem verminderten Selbstbewusstsein assoziiert sind.
48	Umweltgifte	Finde Dokumente über Umweltgifte.	Relevante Dokumente behandeln die verschiedenen Umweltgifte sowie ihre Wirkung auf Wasser, Boden, Luft, Mikroorganismen, Pflanzen, Tiere und den Menschen. Relevant sind Dokumente, wenn sie allgemein Umweltgifte thematisieren oder ein spezielles Umweltgift behandeln. Teilweise relevant sind Dokumente, die die Auswirkungen von Umweltgiften auf ein sehr spezifisches Gebiet thematisieren. Nicht relevant sind Dokumente, die sich mit giftigen Pflanzen befassen.
49	Studium im Ausland	Finde Dokumente, die sich mit dem Studium im Ausland beschäftigen.	Relevant sind Dokumente, wenn sie allgemein über Studienmöglichkeiten im Ausland informieren. Teilweise relevante Dokumente sind allgemeine Ratgeber,die sich mit dem Leben und Arbeiten und der Bildung im Ausland befassen. Nicht relevant sind Dokumente, welche sich mit einem Studium in Deutschland oder ausländischen Studierenden an deutschen Hochschulen befassen.
50	Verkehrsgeographie	Finde Dokumente aus dem Themenbereich "Verkehrsgeographie".	Relevante Dokumente beschäftigen sich mit den Rahmenbedingungen, der Realisierung und den Konsequenzen der Raumüberwindung von Personen und Gütern; teilweise auch der von Informationen. Relevant sind auch solche Dokumente, die sich mit der Geschichte der Verkehrsgeographie beschäftigen, wie auch solche, die die Aufgabenfelder der Verkehrsgeographie in Hinblick auf die Einflüsse aus anderen Bereichen der Geographie und deren Nachbarwissenschaften betrachten. Teilweise relevant sind Dokumente, die sich mit Verkehrsgeographie im Ausland befassen.

Anhang 3: Tabelle den Suchanfragen der Personen A, B, C und C zu jedem der 50 Topics mit der jeweiligen Ergebnismenge

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
01	Kritische Theorie	A	8	(title_txt_de:(kritisch AND Theorie)) OR (subject_vlb_txt_de:(kritisch AND Theorie)) OR (subject_gnd_txt_de:(kritisch AND Theorie)) OR (subject_auto_txt_de:(kritisch AND Theorie))
		B	1	(title_txt_de: ("kritische Theorie" AND Kapitalismus)) OR (subject_vlb_txt_de: ("kritische Theorie" AND Kapitalismus)) OR (subject_gnd_txt_de: ("kritische Theorie" AND Kapitalismus))
		C	9	(title_txt_de:((kritisch AND Theorie) AND "Frankfurter Schule")) OR (subject_vlb_txt_de:((kritisch AND Theorie) AND "Frankfurter Schule")) OR (subject_gnd_txt_de:((kritisch AND Theorie) AND "Frankfurter Schule")) OR (subject_auto_txt_de:((kritisch AND Theorie) AND "Frankfurter Schule"))
		D	71	(subject_vlb_txt_de: ((kritische AND Theorie) OR (Kritik AND bürgerl* AND gesellsch* AND kapitalist*))) OR (subject_gnd_txt_de: ((kritische AND Theorie) OR (Kritik AND bürgerl* AND gesellsch* AND kapitalist*))) OR (subject_auto_txt_de: ((kritische AND Theorie) OR (Kritik AND bürgerl* AND gesellsch* AND kapitalist*)))
02	Ökologie der Gewässer	A	10	title_txt_de:(Ökologie AND Wasser) OR subject_vlb_txt_de:(Ökologie AND Wasser) OR subject_gnd_txt_de:(Ökologie AND Wasser) OR subject_auto_txt_de:(Ökologie AND Wasser)
		B	2	title_txt_de: (Ökologie AND *gewässer) OR subject_vlb_txt_de: (Ökologie AND *gewässer) OR subject_gnd_txt_de: (Ökologie AND *gewässer)
		C	35	(title_txt_de:(Ökolog* AND (Gewäss* OR Fluss OR See* OR Fließgewäss*)) OR subject_auto_txt_de:(Ökolog* AND (Gewäss* OR Fluss OR See* OR Fließgewäss*)) OR subject_vlb_txt_de:(Ökolog* AND (Gewäss* OR Fluss OR See* OR Fließgewäss*)) OR subject_gnd_txt_de:(Ökolog* AND (Gewäss* OR Fluss OR See* OR Fließgewäss*)))
		D	9	((title_txt_de: ((ökolog* AND *gewässer*)OR(selbstreinig* AND *gewässer*))) OR (subject_gnd_txt_de:((ökolog* AND *gewässer*)OR(selbstreinig* AND *gewässer*))) OR (subject_vlb_txt_de: ((ökolog* AND *gewässer*)OR(selbstreinig* AND *gewässer*))) OR (subject_auto_txt_de: ((ökolog* AND *gewässer*)OR(selbstreinig* AND *gewässer*))))
03	Gerontologie	A	31	title_txt_de:(Gerontolog*) OR subject_vlb_txt_de:(Gerontolog*) OR subject_gnd_txt_de:(Gerontolog*) OR subject_auto_txt_de:(Gerontolog*)
		B	90	title_txt_de:(Gerontologie OR (Wissenschaft AND Alter)) OR subject_vlb_txt_de: (Gerontologie OR (Wissenschaft AND Alter))OR subject_gnd_txt_de: (Gerontologie OR (Wissenschaft AND Alter))
		C	179	title_txt_de:(Gerontolog* OR Gerontologie OR Alterswissenschaft OR Alternswissenschaft OR ((Alter* OR Altern*) AND Wissenschaft)) OR subject_vlb_txt_de:(Gerontolog* OR Gerontologie OR Alterswissenschaft OR Alternswissenschaft OR ((Alter* OR Altern*) AND Wissenschaft)) OR subject_gnd_txt_de:(Gerontolog* OR Gerontologie OR Alterswissenschaft OR Alternswissenschaft OR ((Alter* OR Altern*) AND Wissenschaft)) OR subject_auto_txt_de:(Gerontolog* OR Gerontologie OR Alterswissenschaft OR Alternswissenschaft OR ((Alter* OR Altern*) AND Wissenschaft))
		D	89	((title_txt_de:(Gerontolog* OR (Alter* AND Forsch*) OR (Alter* AND Gesellschaf*))) OR (subject_gnd_txt_de:(Gerontolog* OR (Alter* AND Forsch*) OR (Alter* AND Gesellschaf*))) OR (subject_vlb_txt_de: (Gerontolog* OR (Alter* AND Forsch*) OR (Alter* AND Gesellschaf*))) OR (subject_auto_txt_de:(Gerontolog* OR (Alter* AND Forsch*) OR (Alter* AND Gesellschaf*))))
04	Frau und Beruf	A	32	title_txt_de: (Frau AND Beruf) OR subject_vlb_txt_de: (Frau AND Beruf) OR subject_gnd_txt_de: (Frau AND Beruf) OR subject_auto_txt_de:(Frau AND Beruf)
		B	26	title_txt_de: (Frau AND Beruf) OR subject_vlb_txt_de: (Frau AND Beruf) OR subject_gnd_txt_de: (Frau AND Beruf)
		C	112	(title_txt_de:(Frau* AND (Beruf* OR Arbeit* OR Job*)) OR subject_auto_txt_de:(Frau* AND (Beruf* OR Arbeit* OR Job*)) OR subject_vlb_txt_de:(Frau* AND (Beruf* OR Arbeit* OR Job*)) OR subject_gnd_txt_de:(Frau* AND (Beruf* OR Arbeit* OR Job*)))

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		D	111	((title_txt_de:((arbeit* OR beruf*) AND frau*)) OR (subject_gnd_txt_de:((arbeit* OR beruf*) AND frau*)) OR (subject_vlb_txt_de: ((arbeit* OR beruf*) AND frau*)) OR (subject_auto_txt_de: ((arbeit* OR beruf*) AND frau*)))
05	Kunst in der Antike	A	72	title_txt_de:(Kunst AND Antike) OR subject_vlb_txt_de:(Kunst AND Antike) OR subject_gnd_txt_de:(Kunst AND Antike) OR subject_auto_txt_de:(Kunst AND Antike)
		B	119	title_txt_de: (Kunst AND (Antike OR Altertum)) OR subject_vlb_txt_de: (Kunst AND (Antike OR Altertum)) OR subject_gnd_txt_de: (Kunst AND (Antike OR Altertum))
		C	77	title_txt_de:((Kunst OR Künste) AND Antik*) OR subject_vlb_txt_de:((Kunst OR Künste) AND Antik*) OR subject_gnd_txt_de:((Kunst OR Künste) AND Antik*) OR subject_auto_txt_de:((Kunst OR Künste) AND Antik*)
		D	180	((title_txt_de:((antik* OR altertum* OR röm* OR griech*) AND (künstl* OR kunst*))) OR (subject_gnd_txt_de:((antik* OR altertum* OR röm* OR griech*) AND (künstl* OR kunst*))) OR (subject_vlb_txt_de: ((antik* OR altertum* OR röm* OR griech*) AND (künstl* OR kunst*))) OR (subject_auto_txt_de: ((antik* OR altertum* OR röm* OR griech*) AND (künstl* OR kunst*))))
06	Angst, Angststörungen und Angsterkrankungen	A	13	title_txt_de:(Angst AND (Angststörung OR angsterkrankung)) OR subject_vlb_txt_de:(Angst AND (Angststörung OR angsterkrankung)) OR subject_gnd_txt_de:(Angst AND (Angststörung OR angsterkrankung)) OR subject_auto_txt_de:(Angst AND (Angststörung OR angsterkrankung))
		B	1	title_txt_de: (Psychologie AND Phobie) OR subject_vlb_txt_de: (Psychologie AND Phobie) OR subject_gnd_txt_de: (Psychologie AND Phobie)
		C	377	title_txt_de:(Angst OR Angststör* OR Angsterkrank* OR (Angst* AND Krank*)) OR subject_vlb_txt_de:(Angst OR Angststör* OR Angsterkrank* OR (Angst* AND Krank*)) OR subject_gnd_txt_de:(Angst OR Angststör* OR Angsterkrank* OR (Angst* AND Krank*)) OR subject_auto_txt_de:(Angst OR Angststör* OR Angsterkrank* OR (Angst* AND Krank*))
		D	180	((title_txt_de:(*Phobie* OR *Angst* OR (Psych* AND *stör* AND *angst*))) OR (subject_gnd_txt_de: (*Phobie* OR *Angst* OR (Psych* AND *stör* AND *angst*))) OR (subject_vlb_txt_de: (*Phobie* OR *Angst* OR (Psych* AND *stör* AND *angst*))) OR (subject_auto_txt_de:(*Phobie* OR *Angst* OR (Psych* AND *stör* AND *angst*))))
07	Festkörperphysik	A	57	title_txt_de:(Festkörperphysik) OR subject_vlb_txt_de:(Festkörperphysik) OR subject_gnd_txt_de:(Festkörperphysik) OR subject_auto_txt_de:(Festkörperphysik)
		B	55	title_txt_de: (Festkörperphysik) OR subject_vlb_txt_de: (Festkörperphysik) OR subject_gnd_txt_de: (Festkörperphysik)
		C	57	title_txt_de:((Materie AND Körper AND Physik) OR "Festkörperphysik") OR subject_vlb_txt_de:((Materie AND Körper AND Physik) OR "Festkörperphysik") OR subject_gnd_txt_de:((Materie AND Körper AND Physik) OR "Festkörperphysik") OR subject_auto_txt_de:((Materie AND Körper AND Physik) OR "Festkörperphysik")
		D	10	title_txt_de:((Festkörper AND Physik) OR (Materie AND Fest AND Aggregatzustand)) OR subject_vlb_txt_de:((Festkörper AND Physik) OR (Materie AND Fest AND Aggregatzustand)) OR subject_gnd_txt_de:((Festkörper AND Physik) OR (Materie AND Fest AND Aggregatzustand)) OR subject_auto_txt_de:((Festkörper AND Physik) OR (Materie AND Fest AND Aggregatzustand))
08	Tierexperimente	A	35	title_txt_de:(Tierexperiment* OR (Tier* AND *experiment)) OR subject_vlb_txt_de:(Tierexperiment* OR (Tier* AND *experiment)) OR subject_gnd_txt_de:(Tierexperiment* OR (Tier* AND *experiment)) OR subject_auto_txt_de:(Tierexperiment* OR (Tier* AND *experiment))
		B	3	title_txt_de:(Tierversuch* AND (Medizin OR Studi*)) OR subject_vlb_txt_de:(Tierversuch* AND (Medizin OR Studi*)) OR subject_gnd_txt_de:(Tierversuch* AND (Medizin OR Studi*))

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		C	69	title_txt_de:(tierversuch* OR tierexperiment* OR versuchstier*) OR subject_auto_txt_de:(tierversuch* OR tierexperiment* OR versuchstier*) OR subject_vlb_txt_de:(tierversuch* OR tierexperiment* OR versuchstier*) OR subject_gnd_txt_de:(tierversuch* OR tierexperiment* OR versuchstier*)
		D	9	((title_txt_de:(Tier* AND (*versuch* OR *experiment*) AND (*forsch* OR *wissenschaft* OR *medizin* OR *universit*))) OR (subject_gnd_txt_de: (Tier* AND (*versuch* OR *experiment*) AND (*forsch* OR *wissenschaft* OR *medizin* OR *universit*))) OR (subject_vlb_txt_de: (Tier* AND (*versuch* OR *experiment*) AND (*forsch* OR *wissenschaft* OR *medizin* OR *universit*))) OR (subject_auto_txt_de: (Tier* AND (*versuch* OR *experiment*) AND (*forsch* OR *wissenschaft* OR *medizin* OR *universit*))))
09	Widerstand im Nationalsozialismus	A	109	title_txt_de:(Widerstand AND Nationalsozia*) OR subject_vlb_txt_de:(Widerstand AND Nationalsozia*) OR subject_gnd_txt_de:(Widerstand AND Nationalsozia*) OR subject_auto_txt_de:(Widerstand AND Nationalsozia*)
		B	107	title_txt_de: (Widerstand AND Nationalsozialismus) OR subject_vlb_txt_de: (Widerstand AND Nationalsozialismus) OR subject_gnd_txt_de: (Widerstand AND Nationalsozialismus)
		C	131	title_txt_de:(Widerstan* AND (Nationalsozial* OR (Dritt* AND reich*))) OR subject_vlb_txt_de:(Widerstan* AND (Nationalsozial* OR (Dritt* AND reich*))) OR subject_gnd_txt_de:(Widerstan* AND (Nationalsozial* OR (Dritt* AND reich*))) OR subject_auto_txt_de:(Widerstan* AND (Nationalsozial* OR (Dritt* AND reich*)))
		D	132	((title_txt_de: ((Nazi* OR Nationalsozi* OR Wehrmacht) AND (Wider* OR "Kampf gegen"))) OR (subject_gnd_txt_de: ((Nazi* OR Nationalsozi* OR Wehrmacht) AND (Wider* OR "Kampf gegen"))) OR (subject_vlb_txt_de: ((Nazi* OR Nationalsozi* OR Wehrmacht) AND (Wider* OR "Kampf gegen"))) OR (subject_auto_txt_de: ((Nazi* OR Nationalsozi* OR Wehrmacht) AND (Wider* OR "Kampf gegen"))))
10	Psychosomatische Krankheiten	A	10	title_txt_de:(Psychosomatisch* AND Krankheit*) OR subject_vlb_txt_de:(Psychosomatisch* AND Krankheit*) OR subject_gnd_txt_de:(Psychosomatisch* AND Krankheit*) OR subject_auto_txt_de:(Psychosomatisch* AND Krankheit*)
		B	15	title_txt_de: (Somatoform* AND Störung) OR subject_vlb_txt_de: (Somatoform* AND Störung) OR subject_gnd_txt_de: (Somatoform* AND Störung)
		C	20	title_txt_de:(Krank* AND (psychosom* OR (Ursach* AND Psych*))) OR subject_vlb_txt_de:(Krank* AND (psychosom* OR (Ursach* AND Psych*))) OR subject_gnd_txt_de:(Krank* AND (psychosom* OR (Ursach* AND Psych*))) OR subject_auto_txt_de:(Krank* AND (psychosom* OR (Ursach* AND Psych*)))
		D	100	((title_txt_de:(psychosomat* OR (essstör* AND psych*))) OR (subject_gnd_txt_de:(psychosomat* OR (essstör* AND psych*))) OR (subject_vlb_txt_de: (psychosomat* OR (essstör* AND psych*))) OR (subject_auto_txt_de: (psychosomat* OR (essstör* AND psych*))))
11	Umweltökonomie/ -ökonomik	A	27	title_txt_de:(Umweltökonomik OR Umweltökonomie) OR subject_vlb_txt_de:(Umweltökonomik OR Umweltökonomie) OR subject_gnd_txt_de:(Umweltökonomik OR Umweltökonomie) OR subject_auto_txt_de:(Umweltökonomik OR Umweltökonomie)
		B	28	title_txt_de: (Umweltökonom*) OR subject_vlb_txt_de: (Umweltökonom*) OR subject_gnd_txt_de: (Umweltökonom*)
		C	37	title_txt_de:(Umweltökonomik OR Umweltökonomi* OR (umwelt* AND (ökonomie OR ökonomik))) OR subject_vlb_txt_de:(Umweltökonomik OR Umweltökonomi* OR (umwelt* AND (ökonomie OR ökonomik))) OR subject_gnd_txt_de:(Umweltökonomik OR Umweltökonomi* OR (umwelt* AND (ökonomie OR ökonomik))) OR subject_auto_txt_de:(Umweltökonomik OR Umweltökonomi* OR (umwelt* AND (ökonomie OR ökonomik)))
		D	279	((title_txt_de:(Umwelt* AND (*wirtschaft OR *ökonomie))) OR (subject_gnd_txt_de: (Umwelt* AND (*wirtschaft OR *ökonomie))) OR (subject_vlb_txt_de: (Umwelt* AND (*wirtschaft OR *ökonomie))) OR (subject_auto_txt_de: (Umwelt* AND (*wirtschaft OR *ökonomie))))
12	Sprachwandel	A	170	title_txt_de:(Sprachwandel) OR subject_vlb_txt_de:(Sprachwandel) OR subject_gnd_txt_de:(Sprachwandel) OR subject_auto_txt_de:(Sprachwandel)

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		B	2	title_txt_de: (Sprachwandel AND Migration) OR subject_vlb_txt_de: (Sprachwandel AND Migration) OR subject_gnd_txt_de: (Sprachwandel AND Migration)
		C	212	title_txt_de:(((Veränderung OR Entwicklung) AND Sprache) OR Sprachwandel OR Sprachdynamik) OR subject_vlb_txt_de: (((Veränderung OR Entwicklung) AND Sprache) OR Sprachwandel OR Sprachdynamik) OR subject_gnd_txt_de:(((Veränderung OR Entwicklung) AND Sprache) OR Sprachwandel OR Sprachdynamik) OR subject_auto_txt_de:(((Veränderung OR Entwicklung) AND Sprache) OR Sprachwandel OR Sprachdynamik)
		D	172	(title_txt_de: (sprachwandel OR (Sprache AND Einfluss AND(migration technisch sozial politisch wandel)))) OR (subject_gnd_txt_de: (sprachwandel OR (Sprache AND Einfluss AND (migration technisch sozial politisch wandel)))) OR (subject_vlb_txt_de:(sprachwandel OR (Sprache AND Einfluss AND(migration technisch sozial politisch wandel)))) OR (subject_auto_txt_de:(sprachwandel OR (sprache AND Einfluss AND (migration technisch sozial politisch wandel))))
13	Bundestagswahl	A	84	title_txt_de: (Bundestagswahl) OR subject_vlb_txt_de: (Bundestagswahl) OR subject_gnd_txt_de: (Bundestagswahl) OR subject_auto_txt_de:(Bundestagswahl)
		B	52	title_txt_de: (Bundestagswahl) OR subject_vlb_txt_de: (Bundestagswahl) OR subject_gnd_txt_de: (Bundestagswahl)
		C	91	title_txt_de:(Bundestagswahl* OR (Bundestag* AND wahl*)) OR subject_vlb_txt_de:(Bundestagswahl* OR (Bundestag* AND wahl*)) OR subject_gnd_txt_de:(Bundestagswahl* OR (Bundestag* AND wahl*)) OR subject_auto_txt_de:(Bundestagswahl* OR (Bundestag* AND wahl*))
		D	94	((title_txt_de:(*wahl* AND Bundestag*)) OR (subject_gnd_txt_de: (*wahl* AND Bundestag*)) OR (subject_vlb_txt_de: (*wahl* AND Bundestag*)) OR (subject_auto_txt_de: (*wahl* AND Bundestag*)))
14	Alternative Energiequellen	A	32	title_txt_de:(Alternativ* AND Energie*) OR subject_vlb_txt_de:(Alternativ* AND Energie*) OR subject_gnd_txt_de:(Alternativ* AND Energie*) OR subject_auto_txt_de:(Alternativ* AND Energie*)
		B	425	title_txt_de: ("Erneuerbare Energie") OR subject_vlb_txt_de: ("Erneuerbare Energie") OR subject_gnd_txt_de: ("Erneuerbare Energie")
		C	365	title_txt_de:((alternativ* OR regenrativ* OR erneuerbar*) AND (energie*)) OR subject_vlb_txt_de:((alternativ* OR regenrativ* OR erneuerbar*) AND (energie*)) OR subject_gnd_txt_de:((alternativ* OR regenrativ* OR erneuerbar*) AND (energie*)) OR subject_auto_txt_de:((alternativ* OR regenrativ* OR erneuerbar*) AND (energie*))
		D	49	((title_txt_de:(Energie* AND (erneuerbar* OR alternativ* OR nachhaltig*) AND (*quelle* OR *erzeug?* OR *gewinnung))) OR (subject_gnd_txt_de: (Energie* AND (erneuerbar* OR alternativ* OR nachhaltig*) AND (*quelle* OR *erzeug?* OR *gewinnung))) OR (subject_vlb_txt_de: (Energie* AND (erneuerbar* OR alternativ* OR nachhaltig*) AND (*quelle* OR *erzeug?* OR *gewinnung))) OR (subject_auto_txt_de: (Energie* AND (erneuerbar* OR alternativ* OR nachhaltig*) AND (*quelle* OR *erzeug?* OR *gewinnung))))
15	Wirtschaftsethik	A	191	title_txt_de:(Wirtschaftsethik) OR subject_vlb_txt_de:(Wirtschaftsethik) OR subject_gnd_txt_de:(Wirtschaftsethik) OR subject_auto_txt_de:(Wirtschaftsethik)
		B	28	title_txt_de:(Verantwortung* AND Wirtschaft*) OR subject_vlb_txt_de:(Verantwortung* AND Wirtschaft*) OR subject_gnd_txt_de: (Verantwortung* AND Wirtschaft*)
		C	14	title_txt_de:(Wirtschaft* AND *ethik AND (Verantwortun* OR Humantität OR Solidarität)) OR subject_vlb_txt_de:(Wirtschaft* AND *ethik AND (Verantwortun* OR Humantität OR Solidarität)) OR subject_gnd_txt_de:(Wirtschaft* AND *ethik AND (Verantwortun* OR Humantität OR Solidarität)) OR subject_auto_txt_de:(Wirtschaft* AND *ethik AND (Verantwortun* OR Humantität OR Solidarität))
		D	132	(title_txt_de: ((ethische AND Wirtschaft) OR (wirtschaft AND (sozial gerechtigkeit Nachhaltigkeit)))) OR (subject_gnd_txt_de: ((ethische AND Wirtschaft) OR (wirtschaft AND (sozial gerechtigkeit Nachhaltigkeit)))) OR (subject_vlb_txt_de: ((ethische AND Wirtschaft) OR (wirtschaft AND (sozial gerechtigkeit Nachhaltigkeit)))) OR (subject_auto_txt_de: ((ethische AND Wirtschaft) OR (wirtschaft AND (sozial gerechtigkeit Nachhaltigkeit))))

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
16	Hyperaktivität bei Kindern	A	3	title_txt_de:(Hyperaktivität AND Kind) OR subject_vlb_txt_de:(Hyperaktivität AND Kind) OR subject_gnd_txt_de:(Hyperaktivität AND Kind) OR subject_auto_txt_de:(Hyperaktivität AND Kind)
		B	2	title_txt_de: (ADHS AND Behandlung) OR subject_vlb_txt_de: (ADHS AND Behandlung) OR subject_gnd_txt_de: (ADHS AND Behandlung)
		C	38	title_txt_de:((Hyperaktiv* OR "ADHS") AND Kind*) OR subject_vlb_txt_de:((Hyperaktiv* OR "ADHS") AND Kind*) OR subject_gnd_txt_de:((Hyperaktiv* OR "ADHS") AND Kind*) OR subject_auto_txt_de:((Hyperaktiv* OR "ADHS") AND Kind*)
		D	41	((title_txt_de:(kind* AND (hyperaktiv* OR adhs OR (ads AND stör*) OR (Aufmerksam* AND stör*)))) OR (subject_gnd_txt_de:(kind* AND (hyperaktiv* OR adhs OR (ads AND stör*) OR (Aufmerksam* AND stör*)))) OR (subject_vlb_txt_de: (kind* AND (hyperaktiv* OR adhs OR (ads AND stör*) OR (Aufmerksam* AND stör*)))) OR (subject_auto_txt_de:(kind* AND (hyperaktiv* OR adhs OR (ads AND stör*) OR (Aufmerksam* AND stör*)))))
17	Direktmarketing	A	20	title_txt_de: (Direkt* AND *marketing) OR subject_vlb_txt_de: (Direkt* AND *marketing) OR subject_gnd_txt_de: (Direkt* AND *marketing) OR subject_auto_txt_de:(Direkt* AND *marketing)
		B	7	title_txt_de: (Direktmarketing) OR subject_vlb_txt_de: (Direktmarketing) OR subject_gnd_txt_de: (Direktmarketing)
		C	13	(title_txt_de:(*direktmarketing*)) OR subject_auto_txt_de:(*direktmarketing*) OR subject_vlb_txt_de:(*direktmarketing*) OR subject_gnd_txt_de:(*direktmarketing*)
		D	89	((title_txt_de:(Direktmarketing OR ((*Marketing* OR *Werbung*) AND Kunden*))) OR (subject_gnd_txt_de:(Direktmarketing OR ((*Marketing* OR *Werbung*) AND Kunden*))) OR (subject_vlb_txt_de: (Direktmarketing OR ((*Marketing* OR *Werbung*) AND Kunden*))) OR (subject_auto_txt_de: (Direktmarketing OR ((*Marketing* OR *Werbung*) AND Kunden*))))
18	Behandlung bei Schlaganfällen	A	8	title_txt_de:(Behandlung* AND Schlaganfall) OR subject_vlb_txt_de:(Behandlung* AND Schlaganfall) OR subject_gnd_txt_de:(Behandlung* AND Schlaganfall) OR subject_auto_txt_de:(Behandlung* AND Schlaganfall)
		B	7	title_txt_de: (Schlaganfall AND Rehabilitation) OR subject_vlb_txt_de: (Schlaganfall AND Rehabilitation) OR subject_gnd_txt_de:(Schlaganfall AND Rehabilitation)
		C	121	title_txt_de:((*Schlaganfall* AND *Behandlung*) OR (stroke unit)) OR subject_auto_txt_de:((*Schlaganfall* AND *Behandlung*) OR (stroke unit)) OR subject_vlb_txt_de:((*Schlaganfall* AND *Behandlung*) OR (stroke unit)) OR subject_gnd_txt_de:((*Schlaganfall* AND *Behandlung*) OR (stroke unit))
		D	28	((title_txt_de:((Schlaganfall* OR Apoplex* OR *Hirnschlag*) AND (Behandlung* OR Rekonvales* OR Therap*))) OR (subject_gnd_txt_de:((Schlaganfall* OR Apoplex* OR *Hirnschlag*) AND (Behandlung* OR Rekonvales* OR Therap*))) OR (subject_vlb_txt_de: ((Schlaganfall* OR Apoplex* OR *Hirnschlag*) AND (Behandlung* OR Rekonvales* OR Therap*))) OR (subject_auto_txt_de: ((Schlaganfall* OR Apoplex* OR *Hirnschlag*) AND (Behandlung* OR Rekonvales* OR Therap*))))
19	Zimmerpflanzen	A	18	title_txt_de:(Zimmerpflanz*) OR subject_vlb_txt_de:(Zimmerpflanz*) OR subject_gnd_txt_de:(Zimmerpflanz*) OR subject_auto_txt_de:(Zimmerpflanz*)
		B	16	title_txt_de: (*Zimmerpflanz*) OR subject_vlb_txt_de: (*Zimmerpflanz*) OR subject_gnd_txt_de: (*Zimmerpflanz*)
		C	21	(title_txt_de:(*Zimmerpflanz* OR (Pflanz* AND raum*)) OR subject_vlb_txt_de:(*Zimmerpflanz* OR (Pflanz* AND raum*)) OR subject_gnd_txt_de:(*Zimmerpflanz* OR (Pflanz* AND raum*)))

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		D	31	((title_txt_de:(((Topf* OR Zimmer* OR Raum* OR Indoor*) AND (*pflanz* OR *garten* OR *begrün* OR *gewächs*)) AND (*pfleg* OR *behandl* OR *zucht* OR *pflanz*))) OR (subject_gnd_txt_de:(((Topf* OR Zimmer* OR Raum* OR Indoor*) AND (*pflanz* OR *garten* OR *begrün* OR *gewächs*)) AND (*pfleg* OR *behandl* OR *zucht* OR *pflanz*))) OR (subject_vlb_txt_de:(((Topf* OR Zimmer* OR Raum* OR Indoor*) AND (*pflanz* OR *garten* OR *begrün* OR *gewächs*)) AND (*pfleg* OR *behandl* OR *zucht* OR *pflanz*))) OR (subject_auto_txt_de:(((Topf* OR Zimmer* OR Raum* OR Indoor*) AND (*pflanz* OR *garten* OR *begrün* OR *gewächs*)) AND (*pfleg* OR *behandl* OR *zucht* OR *pflanz*))))
20	Reiseführer für die Toskana	A	67	title_txt_de:(Reiseführer AND Toskana) OR subject_vlb_txt_de:(Reiseführer AND Toskana) OR subject_gnd_txt_de:(Reiseführer AND Toskana) OR subject_auto_txt_de:(Reiseführer AND Toskana)
		B	43	title_txt_de: (Toskana AND Reise) OR subject_vlb_txt_de: (Toskana AND Reise) OR subject_gnd_txt_de: (Toskana AND Reise)
		C	96	title_txt_de:(Reise* AND (Toskan* OR Toscan*)) OR subject_vlb_txt_de:(Reise* AND (Toskan* OR Toscan*)) OR subject_gnd_txt_de:(Reise* AND (Toskan* OR Toscan*)) OR subject_auto_txt_de:(Reise* AND (Toskan* OR Toscan*))
		D	51	((title_txt_de:((Reiseführer OR Sehenswürdigkeiten OR "schönsten Orte") AND (Toscana OR Pisa OR Siena OR Lucca OR Florenz OR "San Giminiano"))) OR (subject_gnd_txt_de:((Reiseführer OR Sehenswürdigkeiten OR "schönsten Orte") AND (Toscana OR Pisa OR Siena OR Lucca OR Florenz OR "San Giminiano"))) OR (subject_vlb_txt_de: ((Reiseführer OR Sehenswürdigkeiten OR "schönsten Orte") AND (Toscana OR Pisa OR Siena OR Lucca OR Florenz OR "San Giminiano"))) OR (subject_auto_txt_de: ((Reiseführer OR Sehenswürdigkeiten OR "schönsten Orte") AND (Toscana OR Pisa OR Siena OR Lucca OR Florenz OR "San Giminiano"))))
21	Lateinamerikanische Tänze	A	2	title_txt_de:(Lateinamerikanisch* AND (Tanz OR Tänze)) OR subject_vlb_txt_de:(Lateinamerikanisch* AND (Tanz OR Tänze)) OR subject_gnd_txt_de:(Lateinamerikanisch* AND (Tanz OR Tänze)) OR subject_auto_txt_de:(Lateinamerikanisch* AND (Tanz OR Tänze))
		B	2	title_txt_de: ("Lateinamerikanische Tänze") OR subject_vlb_txt_de: ("Lateinamerikanische Tänze") OR subject_gnd_txt_de: ("Lateinamerikanische Tänze")
		C	32	title_txt_de:((Lateinamerika AND (Tanz OR Tänze)) OR jive OR rumba OR salsa OR (cha-cha-cha) OR (paso doble)) OR subject_vlb_txt_de:((Lateinamerika AND (Tanz OR Tänze)) OR jive OR rumba OR salsa OR (cha-cha-cha) OR (paso doble)) OR subject_gnd_txt_de:((Lateinamerika AND (Tanz OR Tänze)) OR jive OR rumba OR salsa OR (cha-cha-cha) OR (paso doble)) OR subject_auto_txt_de:((Lateinamerika AND (Tanz OR Tänze)) OR jive OR rumba OR salsa OR (cha-cha-cha) OR (paso doble))
		D	14	(title_txt_de: ((Tango Rumba Argentino Salsa Samba lateinamerika) AND Tanz)) OR (subject_gnd_txt_de: ((Tango Rumba Argentino Salsa Samba lateinamerika) AND Tanz)) OR (subject_vlb_txt_de: ((Tango Rumba Argentino Salsa Samba lateinamerika) AND Tanz)) OR (subject_auto_txt_de: ((Tango Rumba Argentino Salsa Samba lateinamerika) AND Tanz))
22	Gesprächsführung	A	89	title_txt_de:(Gesprächsführung) OR subject_vlb_txt_de:(Gesprächsführung) OR subject_gnd_txt_de:(Gesprächsführung) OR subject_auto_txt_de:(Gesprächsführung)
		B	26	title_txt_de: (Moderation) OR subject_vlb_txt_de: (Moderation) OR subject_gnd_txt_de: (Moderation)
		C	101	title_txt_de:((Gespräch* AND *Führung*) OR (Gespräch* AND Kommunikation*)) OR ((Gespräch* AND *Führung*) OR (Gespräch* AND Kommunikation*)) OR subject_gnd_txt_de:((Gespräch* AND *Führung*) OR (Gespräch* AND Kommunikation*)) OR subject_auto_txt_de:((Gespräch* AND *Führung*) OR (Gespräch* AND Kommunikation*))

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		D	93	(title_txt_de: (gesprächsführung* *moderieren* (*gespräch* AND anleiten*) (diskussion* AND anleiten*))) OR (subject_gnd_txt_de: (gesprächsführung* *moderieren* (*gespräch* AND anleiten*) (diskussion* AND anleiten*))) OR (subject_vlb_txt_de: (gesprächsführung* *moderieren* (*gespräch* AND anleiten*) (diskussion* AND anleiten*))) OR (subject_auto_txt_de: (gesprächsführung* *moderieren* (*gespräch* AND anleiten*) (diskussion* AND anleiten*)))
23	Islamischer Fundamentalismus	A	52	title_txt_de:(Islam AND Fundamentalismus) OR subject_vlb_txt_de:(Islam AND Fundamentalismus) OR subject_gnd_txt_de:(Islam AND Fundamentalismus) OR subject_auto_txt_de:(Islam AND Fundamentalismus)
		B	42	title_txt_de: (*fundamentalis* AND Islam*) OR subject_vlb_txt_de: (*fundamentalis* AND Islam*) OR subject_gnd_txt_de: (*fundamentalis* AND Islam*)
		C	116	title_txt_de:((Islam* AND Fundamentalism*) OR Islamismu*) OR subject_vlb_txt_de:((Islam* AND Fundamentalism*) OR Islamismu*) OR subject_gnd_txt_de:((Islam* AND Fundamentalism*) OR Islamismu*) OR subject_auto_txt_de:((Islam* AND Fundamentalism*) OR Islamismu*)
		D	120	(title_txt_de: (((islam* AND fundamentalis*) (radikal* AND islam*) (muslim* AND fundamentalis*) (islam* AND extrem*)) OR (islam* AND terror*))) OR (subject_gnd_txt_de: (((islam* AND fundamentalis*) (radikal* AND islam*) (muslim* AND fundamentalis*) (islam* AND extrem*)) OR (islam* AND terror*))) OR (subject_vlb_txt_de: (((islam* AND fundamentalis*) (radikal* AND islam*) (muslim* AND fundamentalis*) (islam* AND extrem*)) OR (islam* AND terror*))) OR (subject_auto_txt_de: (((islam* AND fundamentalis*) (radikal* AND islam*) (muslim* AND fundamentalis*) (islam* AND extrem*)) OR (islam* AND terror*)))
24	Geschichte Israels	A	134	title_txt_de:(Geschichte AND Israel) OR subject_vlb_txt_de:(Geschichte AND Israel) OR subject_gnd_txt_de:(Geschichte AND Israel) OR subject_auto_txt_de:(Geschichte AND Israel)
		B	134	title_txt_de: (Geschichte AND Israel) OR subject_vlb_txt_de: (Geschichte AND Israel) OR subject_gnd_txt_de: (Geschichte AND Israel)
		C	20	title_txt_de:((Israel* AND Staat* AND Geschicht*) OR (Geschicht* AND Israel* AND Juden*)) OR subject_vlb_txt_de:((Israel* AND Staat* AND Geschicht*) OR (Geschicht* AND Israel* AND Juden*)) OR subject_gnd_txt_de:((Israel* AND Staat* AND Geschicht*) OR (Geschicht* AND Israel* AND Juden*)) OR subject_auto_txt_de:((Israel* AND Staat* AND Geschicht*) OR (Geschicht* AND Israel* AND Juden*))
		D	184	((title_txt_de:(Israel AND (*staat* OR *geschicht* OR historisch* OR politikgesch* OR *gründung))) OR (subject_gnd_txt_de:(Israel AND (*staat* OR *geschicht* OR historisch* OR politikgesch* OR *gründung))) OR (subject_vlb_txt_de: (Israel AND (*staat* OR *geschicht* OR historisch* OR politikgesch* OR *gründung))) OR (subject_auto_txt_de: (Israel AND (*staat* OR *geschicht* OR historisch* OR politikgesch* OR *gründung))))
25	Heilfasten	A	4	title_txt_de:(Heilfasten) OR subject_vlb_txt_de:(Heilfasten) OR subject_gnd_txt_de:(Heilfasten) OR subject_auto_txt_de:(Heilfasten)
		B	6	title_txt_de: (Heilfasten OR Entschlackung) OR subject_vlb_txt_de: (Heilfasten OR Entschlackung) OR subject_gnd_txt_de: (Heilfasten OR Entschlackung)
		C	4	("Heilfasten" OR (seelisch AND Reinigung)) OR subject_vlb_txt_de:("Heilfasten" OR (seelisch AND Reinigung)) OR subject_gnd_txt_de:("Heilfasten" OR (seelisch AND Reinigung)) OR subject_auto_txt_de:("Heilfasten" OR (seelisch AND Reinigung))
		D	138	title_txt_de: (Heilfasten Nulldiät Fasten *Entschlack* Entgiften)
26	Elektroautos	A	12	title_txt_de:(Elektro* AND *Auto) OR subject_vlb_txt_de:(Elektro* AND *Auto) OR subject_gnd_txt_de:(Elektro* AND *Auto) OR subject_auto_txt_de:(Elektro* AND *Auto)
		B	2	title_txt_de: (Elektromobilität AND Kraftfahrzeug) OR subject_vlb_txt_de: (Elektromobilität AND Kraftfahrzeug) OR subject_gnd_txt_de: (Elektromobilität AND Kraftfahrzeug)

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		C	197	title_txt_de:((elektr* OR e-motor OR e-Mobilität) AND (Auto* OR Fahrzeug)) OR subject_vlb_txt_de:((elektr* OR e-motor OR e-Mobilität) AND (Auto* OR Fahrzeug)) OR subject_gnd_txt_de:((elektr* OR e-motor OR e-Mobilität) AND (Auto* OR Fahrzeug)) OR subject_auto_txt_de:((elektr* OR e-motor OR e-Mobilität) AND (Auto* OR Fahrzeug))
		D	36	(title_txt_de: ((Elektroantrieb E-Motor Elektromotor E-Mobilität) AND (Auto OR Fahrzeug))) OR (subject_gnd_txt_de: ((Elektroantrieb E-Motor Elektromotor E-Mobilität) AND (Auto OR Fahrzeug))) OR (subject_vlb_txt_de: ((Elektroantrieb E-Motor Elektromotor E-Mobilität) AND (Auto OR Fahrzeug))) OR (subject_auto_txt_de:((Elektroantrieb E-Motor Elektromotor E-Mobilität) AND (Auto OR Fahrzeug)))
27	Homöopathische Mittel	A	15	title_txt_de:(Homöopathi* AND *Mittel) OR subject_vlb_txt_de:(Homöopathi* AND *Mittel) OR subject_gnd_txt_de:(Homöopathi* AND *Mittel) OR subject_auto_txt_de:(Homöopathi* AND *Mittel)
		B	8	title_txt_de: (Homöopathie AND Medizin) OR subject_vlb_txt_de: (Homöopathie AND Medizin) OR subject_gnd_txt_de: (Homöopathie AND Medizin)
		C	29	title_txt_de:((Homöopathi* AND (*Mittel OR *medizin*)) OR Homöopathik* OR "homöopathische zubereitung") OR subject_vlb_txt_de:((Homöopathi* AND (*Mittel OR *medizin*)) OR Homöopathik* OR "homöopathische zubereitung") OR subject_gnd_txt_de:((Homöopathi* AND (*Mittel OR *medizin*)) OR Homöopathik* OR "homöopathische zubereitung") OR subject_auto_txt_de:((Homöopathi* AND (*Mittel OR *medizin*)) OR Homöopathik* OR "homöopathische zubereitung")
		D	44	(title_txt_de: (homöopathi* OR (samuel AND hahnemann) OR (homöopath* AND arzneimittel) OR globoli)) OR (subject_gnd_txt_de: (homöopathi* OR (samuel AND hahnemann) OR (homöopath* AND arzneimittel) OR globoli)) OR (subject_vlb_txt_de: (homöopathi* OR (samuel AND hahnemann) OR (homöopath* AND arzneimittel) OR globoli)) OR (subject_auto_txt_de: (homöopathi* OR (samuel AND hahnemann) OR (homöopath* AND arzneimittel) OR globoli))
28	Künstliche Intelligenz	A	105	title_txt_de:(Künstlich AND Intelligenz) OR subject_vlb_txt_de:(Künstlich AND Intelligenz) OR subject_gnd_txt_de:(Künstlich AND Intelligenz) OR subject_auto_txt_de:(Künstlich AND Intelligenz)
		B	58	title_txt_de: ("Künstliche Intelligenz") OR subject_vlb_txt_de: ("Künstliche Intelligenz") OR subject_gnd_txt_de: ("Künstliche Intelligenz")
		C	186	title_txt_de:((Künstlich* AND Intelligenz*) OR "KI" OR (artificial* AND intelligenc*)) OR subject_vlb_txt_de:((Künstlich* AND Intelligenz*) OR "KI" OR (artificial* AND intelligenc*)) OR subject_gnd_txt_de:((Künstlich* AND Intelligenz*) OR "KI" OR (artificial* AND intelligenc*)) OR subject_auto_txt_de:((Künstlich* AND Intelligenz*) OR "KI" OR (artificial* AND intelligenc*))
		D	151	((title_txt_de:((KI OR künstl* OR intelligen* OR maschin*) AND (*gesellschaft* OR technolog* OR bedeutung* OR entwickl*))) OR (subject_gnd_txt_de: ((KI OR künstl* OR intelligen* OR maschin*) AND (*gesellschaft* OR technolog* OR bedeutung* OR entwickl*))) OR (subject_vlb_txt_de: ((KI OR künstl* OR intelligen* OR maschin*) AND (*gesellschaft* OR technolog* OR bedeutung* OR entwickl*))) OR (subject_auto_txt_de: ((KI OR künstl* OR intelligen* OR maschin*) AND (*gesellschaft* OR technolog* OR bedeutung* OR entwickl*))))
29	Außenpolitik der USA	A	57	title_txt_de:(Außenpolitik AND USA) OR subject_vlb_txt_de:(Außenpolitik AND USA) OR subject_gnd_txt_de:(Außenpolitik AND USA) OR subject_auto_txt_de:(Außenpolitik AND USA)
		B	50	title_txt_de: ((Außenpolitik OR Diplomatie) AND USA) OR subject_vlb_txt_de: ((Außenpolitik OR Diplomatie) AND USA) OR subject_gnd_txt_de: ((Außenpolitik OR Diplomatie) AND USA)
		C	58	title_txt_de:((Außenpolit* OR (Außen* AND Politi*)) AND (USA or (Vereinigt* AND Staat*))) OR subject_vlb_txt_de:((Außenpolit* OR (Außen* AND Politi*)) AND (USA or (Vereinigt* AND Staat*))) OR subject_gnd_txt_de:((Außenpolit* OR (Außen* AND Politi*)) AND (USA or (Vereinigt* AND Staat*))) OR subject_auto_txt_de:((Außenpolit* OR (Außen* AND Politi*)) AND (USA or (Vereinigt* AND Staat*)))

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		D	163	((title_txt_de:((USA OR (vereinigt* AND staat*) OR US-amerik*) AND (außenpoliti* OR international* OR welt*))) OR (subject_gnd_txt_de:((USA OR (vereinigt* AND staat*) OR US-amerik*) AND (außenpoliti* OR international* OR welt*))) OR (subject_vlb_txt_de: ((USA OR (vereinigt* AND staat*) OR US-amerik*) AND (außenpoliti* OR international* OR welt*))) OR (subject_auto_txt_de: ((USA OR (vereinigt* AND staat*) OR US-amerik*) AND (außenpoliti* OR international* OR welt*))))
30	Internetzeitalter	A	24	title_txt_de:((Internet AND Zeitalter) OR Internetzeitalter) OR subject_vlb_txt_de:((Internet AND Zeitalter) OR Internetzeitalter) OR subject_gnd_txt_de:((Internet AND Zeitalter) OR Internetzeitalter) OR subject_auto_txt_de:((Internet AND Zeitalter) OR Internetzeitalter)
		B	5	title_txt_de: ("Digitale Revolution") OR subject_vlb_txt_de: ("Digitale Revolution") OR subject_gnd_txt_de: ("Digitale Revolution")
		C	28	title_txt_de:(((Internet* OR Netzwerk* OR www) AND Zeitalter) OR Internetzeitalter) OR subject_vlb_txt_de:(((Internet* OR Netzwerk* OR www) AND Zeitalter) OR Internetzeitalter) OR subject_gnd_txt_de:(((Internet* OR Netzwerk* OR www) AND Zeitalter) OR Internetzeitalter) OR subject_auto_txt_de:(((Internet* OR Netzwerk* OR www) AND Zeitalter) OR Internetzeitalter)
		D	98	(title_txt_de: ((Internet* (world AND wide AND web) online (social AND media) OR (soziale AND medien)) AND (Alltag* Gesellschaft* politi* Meinungsbildung* informationsverhalten*))) OR (subject_gnd_txt_de: ((Internet* (world AND wide AND web) online (social AND media) OR (soziale AND medien)) AND (Alltag* Gesellschaft* politi* Meinungsbildung* informationsverhalten*))) OR (subject_vlb_txt_de: ((Internet* (world AND wide AND web) online (social AND media) OR (soziale AND medien)) AND (Alltag* Gesellschaft* politi* Meinungsbildung* informationsverhalten*))) OR (subject_auto_txt_de: ((Internet* (world AND wide AND web) online (social AND media) OR (soziale AND medien)) AND (Alltag* Gesellschaft* politi* Meinungsbildung* informationsverhalten*)))
31	Kurden	A	35	title_txt_de:(Kurden) OR subject_vlb_txt_de:(Kurden) OR subject_gnd_txt_de:(Kurden) OR subject_auto_txt_de:(Kurden)
		B	32	title_txt_de: (Kurden) OR subject_vlb_txt_de: (Kurden) OR subject_gnd_txt_de: (Kurden)
		C	40	title_txt_de:(Kurden OR Kurde* OR Kudistan*) OR subject_vlb_txt_de:(Kurden OR Kurde* OR Kudistan*) OR subject_gnd_txt_de: (Kurden OR Kurde* OR Kudistan*) OR subject_auto_txt_de:(Kurden OR Kurde* OR Kudistan*)
		D	26	((title_txt_de:(kurd* AND (ethni* OR volk* OR geschicht* OR lebensr* OR staat*))) OR (subject_gnd_txt_de:(kurd* AND (ethni* OR volk* OR geschicht* OR lebensr* OR staat*))) OR (subject_vlb_txt_de: (kurd* AND (ethni* OR volk* OR geschicht* OR lebensr* OR staat*))) OR (subject_auto_txt_de: (kurd* AND (ethni* OR volk* OR geschicht* OR lebensr* OR staat*))))
32	Chaostheorie	A	22	title_txt_de:(Chaostheorie) OR subject_vlb_txt_de:(Chaostheorie) OR subject_gnd_txt_de:(Chaostheorie) OR subject_auto_txt_de: (Chaostheorie)
		B	3	title_txt_de:(Chaos AND Mathe*) OR subject_vlb_txt_de:(Chaos AND Mathe*) OR subject_gnd_txt_de:(Chaos AND Mathem*)
		C	22	(title_txt_de:(Chaostheorie OR (Chaos AND Theorie)) OR subject_auto_txt_de:(Chaostheorie OR (Chaos AND Theorie)) OR subject_vlb_txt_de:(Chaostheorie OR (Chaos AND Theorie)) OR subject_gnd_txt_de:(Chaostheorie OR (Chaos AND Theorie)))
		D	6	((title_txt_de:(Chaos* AND (*theorie* OR *forsch*))) OR (subject_gnd_txt_de: (Chaos* AND (*theorie* OR *forsch*))) OR (subject_vlb_txt_de: (Chaos* AND (*theorie* OR *forsch*))) OR (subject_auto_txt_de: (Chaos* AND (*theorie* OR *forsch*))))
33	Magersucht	A	27	title_txt_de:(*Magersucht*) OR subject_vlb_txt_de:(*Magersucht*) OR subject_gnd_txt_de:(*Magersucht*) OR subject_auto_txt_de: (*Magersucht*)
		B	7	title_txt_de:(Magersucht* AND (Therapie OR Psycholog*)) OR subject_vlb_txt_de:(Magersucht* AND (Therapie OR Psycholog*)) OR subject_gnd_txt_de:(Magersucht* AND (Therapie OR Psycholog*))

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		C	92	(subject_vlb_txt_de:(Magersucht OR *Essstörung* OR anorexia) OR subject_auto_txt_de:(Magersucht OR *Essstörung* OR anorexia) OR title_txt_de:(Magersucht OR *Essstörung* OR anorexia) OR subject_gnd_txt_de:(Magersucht OR *Essstörung* OR anorexia))
		D	108	((title_txt_de:((essstör* OR bulim* OR magersücht* OR magersucht* OR ((psych* OR seeli*) AND ess*)))) OR (subject_gnd_txt_de:((essstör* OR bulim* OR magersücht* OR magersucht* OR ((psych* OR seeli*) AND ess*)))) OR (subject_vlb_txt_de: ((essstör* OR bulim* OR magersücht* OR magersucht* OR ((psych* OR seeli*) AND ess*)))) OR (subject_auto_txt_de: ((essstör* OR bulim* OR magersücht* OR magersucht* OR ((psych* OR seeli*) AND ess*)))))
34	Politischer Skandal	A	5	title_txt_de:(Politisch AND Skandal) OR subject_vlb_txt_de:(Politisch AND Skandal) OR subject_gnd_txt_de:(Politisch AND Skandal) OR subject_auto_txt_de:(Politisch AND Skandal)
		B	4	title_txt_de: ("Politischer Skandal" OR "Politische Affaire") OR subject_vlb_txt_de: ("Politischer Skandal" OR "Politische Affaire") OR subject_gnd_txt_de: ("Politischer Skandal" OR "Politische Affaire")
		C	13	title_txt_de:((polit* AND skandal*) OR (Demokrati* AND Skandal*)) OR subject_vlb_txt_de:((polit* AND skandal*) OR (Demokrati* AND Skandal*)) OR subject_gnd_txt_de:((polit* AND skandal*) OR (Demokrati* AND Skandal*)) OR subject_auto_txt_de:((polit* AND skandal*) OR (Demokrati* AND Skandal*))
		D	20	((title_txt_de:((Politi* OR staat*) AND (*skandal* OR *affäre* OR *affair*))) OR (subject_gnd_txt_de:((Politi* OR staat*) AND (*skandal* OR *affäre* OR *affair*))) OR (subject_vlb_txt_de:((Politi* OR staat*) AND (*skandal* OR *affäre* OR *affair*))) OR (subject_auto_txt_de:((Politi* OR staat*) AND (*skandal* OR *affäre* OR *affair*))))
35	Entstehung und Entwicklung der Schrift	A	14	title_txt_de:((Entstehung OR Entwicklung) AND Schrift) OR subject_vlb_txt_de:((Entstehung OR Entwicklung) AND Schrift) OR subject_gnd_txt_de:((Entstehung OR Entwicklung) AND Schrift) OR subject_auto_txt_de:((Entstehung OR Entwicklung) AND Schrift)
		B	73	title_txt_de: (Schrift AND Geschichte) OR subject_vlb_txt_de: (Schrift AND Geschichte) OR subject_gnd_txt_de: (Schrift AND Geschichte)
		C	24	title_txt_de:((Entstehung OR Entwicklung) AND Schrift*) OR subject_auto_txt_de:((Entstehung OR Entwicklung) AND Schrift*) OR subject_vlb_txt_de:((Entstehung OR Entwicklung) AND Schrift*) OR subject_gnd_txt_de:((Entstehung OR Entwicklung) AND Schrift*)
		D	173	((title_txt_de:(((Schrift* OR typogr*) AND (*entsteh* OR *entwickl* OR *kommunik* OR *histor* OR *verständ*)))) OR (subject_gnd_txt_de:(((Schrift* OR typogr*) AND (*entsteh* OR *entwickl* OR *kommunik* OR *histor* OR *verständ*)))) OR (subject_vlb_txt_de: (((Schrift* OR typogr*) AND (*entsteh* OR *entwickl* OR *kommunik* OR *histor* OR *verständ*)))) OR (subject_auto_txt_de: (((Schrift* OR typogr*) AND (*entsteh* OR *entwickl* OR *kommunik* OR *histor* OR *verständ*)))))
36	Das Individuum in der Gesellschaft	A	37	title_txt_de:(Individuum AND Gesellschaft) OR subject_vlb_txt_de:(Individuum AND Gesellschaft) OR subject_gnd_txt_de:(Individuum AND Gesellschaft) OR subject_auto_txt_de:(Individuum AND Gesellschaft)
		B	1	title_txt_de: (Individuum AND Gesellschaft AND Wechselwirkung) OR subject_vlb_txt_de: (Individuum AND Gesellschaft AND Wechselwirkung) OR subject_gnd_txt_de: (Individuum AND Gesellschaft AND Wechselwirkung)
		C	69	title_txt_de:(Individu* AND Gesellschaft*) OR subject_vlb_txt_de:(Individu* AND Gesellschaft*) OR subject_gnd_txt_de:(Individu* AND Gesellschaft*) OR subject_auto_txt_de:(Individu* AND Gesellschaft*)
		D	184	((title_txt_de:((Individu* OR einzelpers* OR einzel*) AND (gesellsch* OR *sozial*))) OR (subject_gnd_txt_de:((Individu* OR einzelpers* OR einzel*) AND (gesellsch* OR *sozial*))) OR (subject_vlb_txt_de: ((Individu* OR einzelpers* OR einzel*) AND (gesellsch* OR *sozial*))) OR (subject_auto_txt_de: ((Individu* OR einzelpers* OR einzel*) AND (gesellsch* OR *sozial*))))
37	Straßenkinder	A	19	title_txt_de:(Straße* AND *Kind) OR subject_vlb_txt_de:(Straße* AND *Kind) OR subject_gnd_txt_de:(Straße* AND *Kind) OR subject_auto_txt_de:(Straße* AND *Kind)
		B	0	title_txt_de: (Straßenkinder AND Perspektive) OR subject_vlb_txt_de: (Straßenkinder AND Perspektive) OR subject_gnd_txt_de: (Straßenkinder AND Perspektive)

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		C	19	(title_txt_de:((*Kind* OR *Jugendlich*) AND Straße* OR *armut*) OR subject_auto_txt_de:((*Kind* OR *Jugendlich*) AND Straße* OR *armut*) OR subject_vlb_txt_de:((*Kind* OR *Jugendlich*) AND Straße* OR *armut*) OR subject_gnd_txt_de:((*Kind* OR *Jugendlich*) AND Straße* OR *armut*))
		D	18	((title_txt_de:(straßenkind* OR (obdachlos* AND (minderjähr* OR *kind*) OR (ausreiß* AND (minderjähr* OR *kind*))))) OR (subject_gnd_txt_de:(straßenkind* OR (obdachlos* AND (minderjähr* OR *kind*) OR (ausreiß* AND (minderjähr* OR *kind*))))) OR (subject_vlb_txt_de: (straßenkind* OR (obdachlos* AND (minderjähr* OR *kind*) OR (ausreiß* AND (minderjähr* OR *kind*))))) OR (subject_auto_txt_de: (straßenkind* OR (obdachlos* AND (minderjähr* OR *kind*) OR (ausreiß* AND (minderjähr* OR *kind*))))))
38	Kriegsberichterstattung	A	63	title_txt_de:(Kriegsberichterstattung) OR subject_vlb_txt_de:(Kriegsberichterstattung) OR subject_gnd_txt_de: (Kriegsberichterstattung) OR subject_auto_txt_de:(Kriegsberichterstattung)
		B	60	title_txt_de: (Kriegsberichterstattung OR Kriegsjournalismus) OR subject_vlb_txt_de: (Kriegsberichterstattung OR Kriegsjournalismus) OR subject_gnd_txt_de: (Kriegsberichterstattung OR Kriegsjournalismus)
		C	76	title_txt_de:(((Krieg* OR Krisengebi*) AND Berichterstatt*) OR Kriegsberichterstattung OR Kriegsjournalismus OR (Krieg* AND Journalismus*)) OR subject_vlb_txt_de:(((Krieg* OR Krisengebi*) AND Berichterstatt*) OR Kriegsberichterstattung OR Kriegsjournalismus OR (Krieg* AND Journalismus*)) OR subject_gnd_txt_de:(((Krieg* OR Krisengebi*) AND Berichterstatt*) OR Kriegsberichterstattung OR Kriegsjournalismus OR (Krieg* AND Journalismus*)) OR subject_auto_txt_de:(((Krieg* OR Krisengebi*) AND Berichterstatt*) OR Kriegsberichterstattung OR Kriegsjournalismus OR (Krieg* AND Journalismus*))
		D	66	((title_txt_de:((berichterstatt* OR (*journalist* OR *nachrichten*)) AND (krisengeb* OR *krieg* OR (*konflikt* AND (befaffnet* OR *internation*))))) OR (subject_gnd_txt_de:((berichterstatt* OR (*journalist* OR *nachrichten*)) AND (krisengeb* OR *krieg* OR (*konflikt* AND (befaffnet* OR *internation*))))) OR (subject_vlb_txt_de: ((berichterstatt* OR (*journalist* OR *nachrichten*)) AND (krisengeb* OR *krieg* OR (*konflikt* AND (befaffnet* OR *internation*))))) OR (subject_auto_txt_de: ((berichterstatt* OR (*journalist* OR *nachrichten*)) AND (krisengeb* OR *krieg* OR (*konflikt* AND (befaffnet* OR *internation*))))))
39	Medizin im Dritten Reich	A	1	title_txt_de:(Medizin AND Drittes AND Reich) OR subject_vlb_txt_de:(Medizin AND Drittes AND Reich) OR subject_gnd_txt_de: (Medizin AND Drittes AND Reich) OR subject_auto_txt_de:(Medizin AND Drittes AND Reich)
		B	4	title_txt_de: (Nationalsozialismus AND Medizin) OR subject_vlb_txt_de: (Nationalsozialismus AND Medizin) OR subject_gnd_txt_de: (Nationalsozialismus AND Medizin)
		C	7	title_txt_de:((Medizi* AND ((Dritt* AND Reich*) OR Nationalsozialism*)) OR ((Nationalsozialism* OR (Dritt* AND Reich*)) AND (Leistungsmedizi* OR Gesundheitsführun* OR Rassenhygien*))) OR subject_vlb_txt_de:((Medizi* AND ((Dritt* AND Reich*) OR Nationalsozialism*)) OR ((Nationalsozialism* OR (Dritt* AND Reich*)) AND (Leistungsmedizi* OR Gesundheitsführun* OR Rassenhygien*))) OR subject_gnd_txt_de:((Medizi* AND ((Dritt* AND Reich*) OR Nationalsozialism*)) OR ((Nationalsozialism* OR (Dritt* AND Reich*)) AND (Leistungsmedizi* OR Gesundheitsführun* OR Rassenhygien*))) OR subject_auto_txt_de:((Medizi* AND ((Dritt* AND Reich*) OR Nationalsozialism*)) OR ((Nationalsozialism* OR (Dritt* AND Reich*)) AND (Leistungsmedizi* OR Gesundheitsführun* OR Rassenhygien*)))
		D	12	((title_txt_de:((Nationalsoz* OR Nazi OR (ditt* AND Reich*) OR (Dr. AND Mengele)) AND (medizin* OR rassenhygien* OR experiment* OR Zwangssteril* OR *arzt*))) OR (subject_gnd_txt_de:((Nationalsoz* OR Nazi OR (ditt* AND Reich*) OR (Dr. AND Mengele)) AND (medizin* OR rassenhygien* OR experiment* OR Zwangssteril* OR *arzt*))) OR (subject_vlb_txt_de: ((Nationalsoz* OR Nazi OR (ditt* AND Reich*) OR (Dr. AND Mengele)) AND (medizin* OR rassenhygien* OR experiment* OR Zwangssteril* OR *arzt*))) OR (subject_auto_txt_de:((Nationalsoz* OR Nazi OR (ditt* AND Reich*) OR (Dr. AND Mengele)) AND (medizin* OR rassenhygien* OR experiment* OR Zwangssteril* OR *arzt*))))

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
40	Theatergeschichte	A	74	title_txt_de:(Theatergeschichte) OR subject_vlb_txt_de:(Theatergeschichte) OR subject_gnd_txt_de:(Theatergeschichte) OR subject_auto_txt_de:(Theatergeschichte)
		B	265	title_txt_de: (Theater AND Geschichte) OR subject_vlb_txt_de: (Theater AND Geschichte) OR subject_gnd_txt_de: (Theater AND Geschichte)
		C	77	(title_txt_de:(Theater* AND Geschichte) OR subject_auto_txt_de:(Theater* AND Geschichte) OR subject_vlb_txt_de:(Theater* AND Geschichte) OR subject_gnd_txt_de:(Theater* AND Geschichte))
		D	186	((title_txt_de:(*geschicht* AND (*theater* OR *bühne* OR (szenisch* AND *darst* AND *literat*)))) OR (subject_gnd_txt_de:(*geschicht* AND (*theater* OR *bühne* OR (szenisch* AND *darst* AND *literat*)))) OR (subject_vlb_txt_de: (*geschicht* AND (*theater* OR *bühne* OR (szenisch* AND *darst* AND *literat*)))) OR (subject_auto_txt_de: (*geschicht* AND (*theater* OR *bühne* OR (szenisch* AND *darst* AND *literat*))))
41	Politik und Massenmedien	A	6	title_txt_de:(Politik AND Masse* AND Medi*) OR subject_vlb_txt_de:(Politik AND Masse* AND Medi*) OR subject_gnd_txt_de:(Politik AND Masse* AND Medi*) OR subject_auto_txt_de:(Politik AND Masse* AND Medi*)
		B	13	title_txt_de: (Politik AND massenmed*) OR subject_vlb_txt_de: (Politik AND massenmed*) OR subject_gnd_txt_de: (Politik AND Massenmed*)
		C	15	title_txt_de:(Polit* AND Mass* AND Medi*) OR subject_vlb_txt_de:(Polit* AND Mass* AND Medi*) OR subject_gnd_txt_de:(Polit* AND Mass* AND Medi*) OR subject_auto_txt_de:(Polit* AND Mass* AND Medi*)
		D	6	((title_txt_de:(*politi* AND (*medien OR *presse*) AND (*masse* OR *demokrat* OR *öffentlich-rechtlich*))) OR (subject_gnd_txt_de:(*politi* AND (*medien OR *presse*) AND (*masse* OR *demokrat* OR *öffentlich-rechtlich*))) OR (subject_vlb_txt_de: (*politi* AND (*medien OR *presse*) AND (*masse* OR *demokrat* OR *öffentlich-rechtlich*))) OR (subject_auto_txt_de: (*politi* AND (*medien OR *presse*) AND (*masse* OR *demokrat* OR *öffentlich-rechtlich*))))
42	Wirtschaftswunder in Deutschland	A	7	title_txt_de:(Wirtschaftswunder AND Deutschland) OR subject_vlb_txt_de:(Wirtschaftswunder AND Deutschland) OR subject_gnd_txt_de:(Wirtschaftswunder AND Deutschland) OR subject_auto_txt_de:(Wirtschaftswunder AND Deutschland)
		B	29	title_txt_de: (Wirtschaftswunder) OR subject_vlb_txt_de: (Wirtschaftswunder) OR subject_gnd_txt_de: (Wirtschaftswunder)
		C	44	title_txt_de:(Wirtschaftswund* OR (unerwarte* AND wachstum)) OR subject_vlb_txt_de:(Wirtschaftswund* OR (unerwarte* AND wachstum)) OR subject_gnd_txt_de:(Wirtschaftswund* OR (unerwarte* AND wachstum)) OR subject_auto_txt_de:(Wirtschaftswund* OR (unerwarte* AND wachstum))
		D	31	((title_txt_de:(*wirtschaftswunder* OR (*nachkriegszeit* AND *wirtschaft*) OR Solow*)) OR (subject_gnd_txt_de:(*wirtschaftswunder* OR (*nachkriegszeit* AND *wirtschaft*) OR Solow*)) OR (subject_vlb_txt_de: (*wirtschaftswunder* OR (*nachkriegszeit* AND *wirtschaft*) OR Solow*)) OR (subject_auto_txt_de: (*wirtschaftswunder* OR (*nachkriegszeit* AND *wirtschaft*) OR Solow*)))
43	Spurenelemente in der Ernährung des Menschen	A	6	title_txt_de:(Spurenelemente AND Ernährung*) OR subject_vlb_txt_de:(Spurenelemente AND Ernährung*) OR subject_gnd_txt_de:(Spurenelemente AND Ernährung*) OR subject_auto_txt_de:(Spurenelemente AND Ernährung*)
		B	18	title_txt_de: (Spurenelemente OR (Spurenelemente AND Mensch)) OR subject_vlb_txt_de: (Spurenelemente OR (Spurenelemente AND Mensch)) OR subject_gnd_txt_de: (Spurenelemente OR (Spurenelemente AND Mensch))
		C	22	(title_txt_de:((*spurenelement* AND *ernährung*) OR (*ernährung* AND *mensch*)) OR subject_auto_txt_de:((*spurenelement* AND *ernährung*) OR (*ernährung* AND *mensch*)) OR subject_vlb_txt_de:((*spurenelement* AND *ernährung*) OR (*ernährung* AND *mensch*)) OR subject_gnd_txt_de:((*spurenelement* AND *ernährung*) OR (*ernährung* AND *mensch*)))
		D	9	((title_txt_de:((Spurenelement* OR *mikorplasti* OR *schwermetall*) AND (*organismus* OR *gesundheit* OR *ernähr*))) OR (subject_gnd_txt_de:((Spurenelement* OR *mikorplasti* OR *schwermetall*) AND (*organismus* OR *gesundheit* OR *ernähr*))) OR (subject_vlb_txt_de: ((Spurenelement* OR *mikorplasti* OR *schwermetall*) AND (*organismus* OR *gesundheit* OR *ernähr*))) OR (subject_auto_txt_de:((Spurenelement* OR *mikorplasti* OR *schwermetall*) AND (*organismus* OR *gesundheit* OR *ernähr*))))

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
44	Shakespeares Dramen	A	47	title_txt_de:(Shakespeare AND Drama) OR subject_vlb_txt_de:(Shakespeare AND Drama) OR subject_gnd_txt_de:(Shakespeare AND Drama) OR subject_auto_txt_de:(Shakespeare AND Drama)
		B	46	title_txt_de: (Shakespeare AND Drama) OR subject_vlb_txt_de: (Shakespeare AND Drama) OR subject_gnd_txt_de: (Shakespeare AND Drama)
		C	56	title_txt_de:((Dram* OR Dramen) AND Shakespear*) OR subject_vlb_txt_de:((Dram* OR Dramen) AND Shakespear*) OR subject_gnd_txt_de:((Dram* OR Dramen) AND Shakespear*) OR subject_auto_txt_de:((Dram* OR Dramen) AND Shakespear*)
		D	89	((title_txt_de:(Shakespeare AND (dram* OR *stück* OR *theater* OR komöd* OR tragöd* OR Werk* OR Rezeption*))) OR (subject_gnd_txt_de:(Shakespeare AND (dram* OR *stück* OR *theater* OR komöd* OR tragöd* OR Werk* OR Rezeption*))) OR (subject_vlb_txt_de: (Shakespeare AND (dram* OR *stück* OR *theater* OR komöd* OR tragöd* OR Werk* OR Rezeption*))) OR (subject_auto_txt_de: (Shakespeare AND (dram* OR *stück* OR *theater* OR komöd* OR tragöd* OR Werk* OR Rezeption*))))
45	Betriebspsychologie	A	3	title_txt_de:(Betriebspsychologie) OR subject_vlb_txt_de:(Betriebspsychologie) OR subject_gnd_txt_de:(Betriebspsychologie) OR subject_auto_txt_de:(Betriebspsychologie)
		B	2	title_txt_de: (Betriebspsychologie) OR subject_vlb_txt_de: (Betriebspsychologie) OR subject_gnd_txt_de: (Betriebspsychologie)
		C	13	title_txt_de:(Betriebspsycholog* OR (Betrieb* AND (Psycholog* OR Wirtschaftspsycholog*))) OR subject_vlb_txt_de: (Betriebspsycholog* OR (Betrieb* AND (Psycholog* OR Wirtschaftspsycholog*))) OR subject_gnd_txt_de:(Betriebspsycholog* OR (Betrieb* AND (Psycholog* OR Wirtschaftspsycholog*))) OR subject_auto_txt_de:(Betriebspsycholog* OR (Betrieb* AND (Psycholog* OR Wirtschaftspsycholog*)))
		D	23	((title_txt_de:(Betriebspsycholog* OR (Verhalten* AND (betrieblich* OR arbeits* OR Organisations*)))) OR (subject_gnd_txt_de:(Betriebspsycholog* OR (Verhalten* AND (betrieblich* OR arbeits* OR Organisations*)))) OR (subject_vlb_txt_de: (Betriebspsycholog* OR (Verhalten* AND (betrieblich* OR arbeits* OR Organisations*)))) OR (subject_auto_txt_de: (Betriebspsycholog* OR (Verhalten* AND (betrieblich* OR arbeits* OR Organisations*)))))
46	Marktanalyse	A	43	title_txt_de:(Marktanalyse) OR subject_vlb_txt_de:(Marktanalyse) OR subject_gnd_txt_de:(Marktanalyse) OR subject_auto_txt_de: (Marktanalyse)
		B	36	title_txt_de: (Marktanalyse OR Marktpsychologie) OR subject_vlb_txt_de: (Marktanalyse OR Marktpsychologie) OR subject_gnd_txt_de: (Marktanalyse OR Marktpsychologie)
		C	111	title_txt_de:(Marktanalyse OR (Markt* AND (Analys* OR analyt*))) OR subject_vlb_txt_de:(Marktanalyse OR (Markt* AND (Analys* OR analyt*))) OR subject_gnd_txt_de:(Marktanalyse OR (Markt* AND (Analys* OR analyt*))) OR subject_auto_txt_de:(Marktanalyse OR (Markt* AND (Analys* OR analyt*)))
		D	43	((title_txt_de:(Marktanaly*)) OR (subject_gnd_txt_de:(Marktanaly*)) OR (subject_vlb_txt_de: (Marktanaly*)) OR (subject_auto_txt_de: (Marktanaly*)))
47	Selbstbewusstsein Starken	A	6	title_txt_de:(Selbstbewusst* AND stark*) OR subject_vlb_txt_de:(Selbstbewusst* AND stark*) OR subject_gnd_txt_de: (Selbstbewusst* AND stark*) OR subject_auto_txt_de:(Selbstbewusst* AND stark*)
		B	7	title_txt_de: ((Selbstbewusstsein OR "Selbstbewusstsein stärken") AND Psychologie) OR subject_vlb_txt_de: ((Selbstbewusstsein OR "Selbstbewusstsein stärken") AND Psychologie) OR subject_gnd_txt_de: ((Selbstbewusstsein OR "Selbstbewusstsein stärken") AND Psychologie)
		C	80	title_txt_de:(Selbstbewusstsein*) OR subject_vlb_txt_de:(Selbstbewusstsein*) OR subject_gnd_txt_de:(Selbstbewusstsein*) OR subject_auto_txt_de:(Selbstbewusstsein*)

Topic	Topic	Person	Anzahl Treffer	Suchanfrage
		D	26	((title_txt_de:((selbstbewusst* OR selbstvertrau*) AND (stärk* OR entwickl* OR methoden* OR bildung* OR reifung*))) OR (subject_gnd_txt_de:((selbstbewusst* OR selbstvertrau*) AND (stärk* OR entwickl* OR methoden* OR bildung* OR reifung*))) OR (subject_vlb_txt_de: ((selbstbewusst* OR selbstvertrau*) AND (stärk* OR entwickl* OR methoden* OR bildung* OR reifung*))) OR (subject_auto_txt_de: ((selbstbewusst* OR selbstvertrau*) AND (stärk* OR entwickl* OR methoden* OR bildung* OR reifung*))))
48	Umweltgifte	A	4	title_txt_de:(Umweltgifte) OR subject_vlb_txt_de:(Umweltgifte) OR subject_gnd_txt_de:(Umweltgifte) OR subject_auto_txt_de: (Umweltgifte)
		B	2	title_txt_de: (*Umweltgift*) OR subject_vlb_txt_de: (*Umweltgift*) OR subject_gnd_txt_de: (*Umweltgift*)
		C	7	(Umweltgift* OR (Stoff* AND Umwelt* AND Gefahr*) OR (umweltgefährli* AND Stoff*) OR (Umwelt* AND (Gift* OR Vergift*)))
		D	89	((title_txt_de:((Natur* OR Umwelt*) AND (*gift* OR *toxi*))) OR (subject_gnd_txt_de:((Natur* OR Umwelt*) AND (*gift* OR *toxi*))) OR (subject_vlb_txt_de: ((Natur* OR Umwelt*) AND (*gift* OR *toxi*))) OR (subject_auto_txt_de: ((Natur* OR Umwelt*) AND (*gift* OR *toxi*))))
49	Studium im Ausland	A	3	title_txt_de:(Studium AND Ausland) OR subject_vlb_txt_de:(Studium AND Ausland) OR subject_gnd_txt_de:(Studium AND Ausland) OR subject_auto_txt_de:(Studium AND Ausland)
		B	6	title_txt_de: (Auslandsstudium OR (Studium AND Ausland)) OR subject_vlb_txt_de: (Auslandsstudium OR (Studium AND Ausland)) OR subject_gnd_txt_de: (Auslandsstudium OR (Studium AND Ausland))
		C	7	title_txt_de:((Stud* AND Ausland) OR (Ausland AND Stipendium)) OR subject_vlb_txt_de:((Stud* AND Ausland) OR (Ausland AND Stipendium)) OR subject_gnd_txt_de:((Studi* AND Ausland) OR (Ausland AND Stipendium)) OR subject_auto_txt_de:((Stud* AND Ausland) OR (Ausland AND Stipendium))
		D	61	((title_txt_de:((Ausland* OR (internat* AND Austausch*)) AND (studi* OR Erasmus* OR DAAD OR *ausbild* OR *bildung*))) OR (subject_gnd_txt_de:((Ausland* OR (internat* AND Austausch*)) AND (studi* OR Erasmus* OR DAAD OR *ausbild* OR *bildung*))) OR (subject_vlb_txt_de: ((Ausland* OR (internat* AND Austausch*)) AND (studi* OR Erasmus* OR DAAD OR *ausbild* OR *bildung*))) OR (subject_auto_txt_de: ((Ausland* OR (internat* AND Austausch*)) AND (studi* OR Erasmus* OR DAAD OR *ausbild* OR *bildung*))))
50	Verkehrsgeographie	A	3	title_txt_de:(Studium AND Ausland) OR subject_vlb_txt_de:(Studium AND Ausland) OR subject_gnd_txt_de:(Studium AND Ausland) OR subject_auto_txt_de:(Studium AND Ausland)
		B	3	title_txt_de: (Verkehrsgeographie) OR subject_vlb_txt_de: (Verkehrsgeographie) OR subject_gnd_txt_de: (Verkehrsgeographie)
		C	48	title_txt_de:((Verkehr* AND (*Geographie OR *Nachfrage OR *Infrastruktur)) OR Verkehrswissenschaft*) OR subject_vlb_txt_de: ((Verkehr* AND (*Geographie OR *Nachfrage OR *Infrastruktur)) OR Verkehrswissenschaft*) OR subject_gnd_txt_de:((Verkehr* AND (*Geographie OR *Nachfrage OR *Infrastruktur)) OR Verkehrswissenschaft*) OR subject_auto_txt_de:((Verkehr* AND (*Geographie OR *Nachfrage OR *Infrastruktur)) OR Verkehrswissenschaft*)
		D	56	(title_txt_de: ((logisti* geograph* geograf*) AND (güter* person* verkehr*))) OR (subject_gnd_txt_de: ((logisti* geograph* geograf*) AND (güter* person* verkehr*))) OR (subject_vlb_txt_de: ((logisti* geograph* geograf*) AND (güter* person* verkehr*))) OR (subject_auto_txt_de: ((logisti* geograph* geograf*) AND (güter* person* verkehr*)))

Anhang 4: solr-rel.xsl Datei, Stylesheet zur Umformung der Solr-Ergebnisliste in das Standard-TREC-Format

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
    <xsl:output method="text" indent="yes"
media-type="text/text;charset=utf-8"/>

    <xsl:template match="/">
            <xsl:apply-templates select="//result/doc"/>
    </xsl:template>

    <xsl:template match="doc">
<xsl:text> 01 </xsl:text>
<xsl:text>0 </xsl:text>
<xsl:value-of select="str[@name='id']/text()"/>
<xsl:text> </xsl:text>
<xsl:value-of select="position()-1"/>
<xsl:text> </xsl:text>
<xsl:value-of select="float[@name='score']/text()"/>
<xsl:text> simpleRun</xsl:text>
<xsl:text>
</xsl:text>
    </xsl:template>
</xsl:stylesheet>
```

Anhang 5: Ausgegebene Werte von trec_eval zu den fünf Suchläufen des beispielhaften Retrievaltests

```
Johannas-Air:trec_eval JohannaMunkelt$ ./trec_eval greldnb.txt select131.txt
runid                   all simpleRun
num_q                   all 1
num_ret                 all 40
num_rel                 all 26
num_rel_ret             all 11
map                     all 0.1882
gm_map                  all 0.1882
Rprec                   all 0.3462
bpref                   all 0.2544
recip_rank              all 1.0000
iprec_at_recall_0.00    all 1.0000
iprec_at_recall_0.10    all 0.6000
iprec_at_recall_0.20    all 0.4091
iprec_at_recall_0.30    all 0.4091
iprec_at_recall_0.40    all 0.3333
iprec_at_recall_0.50    all 0.0000
iprec_at_recall_0.60    all 0.0000
iprec_at_recall_0.70    all 0.0000
iprec_at_recall_0.80    all 0.0000
iprec_at_recall_0.90    all 0.0000
iprec_at_recall_1.00    all 0.0000
P_5                     all 0.6000
P_10                    all 0.3000
P_15                    all 0.2667
P_20                    all 0.3500
P_30                    all 0.3333
P_100                   all 0.1100
P_200                   all 0.0550
P_500                   all 0.0220
P_1000                  all 0.0110
```

```
Johannas-Air:trec_eval JohannaMunkelt$ ./trec_eval greldnb.txt select132.txt
runid                   all simpleRun
num_q                   all 1
num_ret                 all 73
num_rel                 all 26
num_rel_ret             all 16
map                     all 0.2096
gm_map                  all 0.2096
Rprec                   all 0.3462
bpref                   all 0.2500
recip_rank              all 0.5000
iprec_at_recall_0.00    all 0.5000
iprec_at_recall_0.10    all 0.4545
iprec_at_recall_0.20    all 0.3750
iprec_at_recall_0.30    all 0.3750
iprec_at_recall_0.40    all 0.2821
iprec_at_recall_0.50    all 0.2653
iprec_at_recall_0.60    all 0.2286
iprec_at_recall_0.70    all 0.0000
iprec_at_recall_0.80    all 0.0000
iprec_at_recall_0.90    all 0.0000
iprec_at_recall_1.00    all 0.0000
P_5                     all 0.2000
P_10                    all 0.4000
P_15                    all 0.3333
P_20                    all 0.3500
P_30                    all 0.3333
P_100                   all 0.1600
P_200                   all 0.0800
P_500                   all 0.0320
P_1000                  all 0.0160
```

```
Johannas-Air:trec_eval JohannaMunkelt$ ./trec_eval greldnb.txt select133.txt
runid                   all simpleRun
num_q                   all 1
num_ret                 all 84
num_rel                 all 26
num_rel_ret             all 17
map                     all 0.2356
gm_map                  all 0.2356
Rprec                   all 0.3462
bpref                   all 0.2574
recip_rank              all 0.5000
iprec_at_recall_0.00    all 0.6667
iprec_at_recall_0.10    all 0.4286
iprec_at_recall_0.20    all 0.3793
iprec_at_recall_0.30    all 0.3793
iprec_at_recall_0.40    all 0.3793
iprec_at_recall_0.50    all 0.3333
iprec_at_recall_0.60    all 0.2667
iprec_at_recall_0.70    all 0.0000
iprec_at_recall_0.80    all 0.0000
iprec_at_recall_0.90    all 0.0000
iprec_at_recall_1.00    all 0.0000
P_5                     all 0.4000
P_10                    all 0.3000
P_15                    all 0.3333
P_20                    all 0.3500
P_30                    all 0.3667
P_100                   all 0.1700
P_200                   all 0.0850
P_500                   all 0.0340
P_1000                  all 0.0170
```

```
Johannas-Air:trec_eval JohannaMunkelt$ ./trec_eval qreldnb.txt select134.txt
runid                   all simpleRun
num_q                   all 1
num_ret                 all 12
num_rel                 all 26
num_rel_ret             all 3
map                     all 0.0529
gm_map                  all 0.0529
Rprec                   all 0.1154
bpref                   all 0.1036
recip_rank              all 0.5000
iprec_at_recall_0.00    all 0.5000
iprec_at_recall_0.10    all 0.3750
iprec_at_recall_0.20    all 0.0000
iprec_at_recall_0.30    all 0.0000
iprec_at_recall_0.40    all 0.0000
iprec_at_recall_0.50    all 0.0000
iprec_at_recall_0.60    all 0.0000
iprec_at_recall_0.70    all 0.0000
iprec_at_recall_0.80    all 0.0000
iprec_at_recall_0.90    all 0.0000
iprec_at_recall_1.00    all 0.0000
P_5                     all 0.4000
P_10                    all 0.3000
P_15                    all 0.2000
P_20                    all 0.1500
P_30                    all 0.1000
P_100                   all 0.0300
P_200                   all 0.0150
P_500                   all 0.0060
P_1000                  all 0.0030
```

```
Johannas-Air:trec_eval JohannaMunkelt$ ./trec_eval qreldnb.txt select135.txt
runid                   all simpleRun
num_q                   all 1
num_ret                 all 66
num_rel                 all 26
num_rel_ret             all 14
map                     all 0.1539
gm_map                  all 0.1539
Rprec                   all 0.1923
bpref                   all 0.1420
recip_rank              all 1.0000
iprec_at_recall_0.00    all 1.0000
iprec_at_recall_0.10    all 0.2667
iprec_at_recall_0.20    all 0.2281
iprec_at_recall_0.30    all 0.2281
iprec_at_recall_0.40    all 0.2281
iprec_at_recall_0.50    all 0.2281
iprec_at_recall_0.60    all 0.0000
iprec_at_recall_0.70    all 0.0000
iprec_at_recall_0.80    all 0.0000
iprec_at_recall_0.90    all 0.0000
iprec_at_recall_1.00    all 0.0000
P_5                     all 0.2000
P_10                    all 0.2000
P_15                    all 0.2667
P_20                    all 0.2500
P_30                    all 0.2000
P_100                   all 0.1400
P_200                   all 0.0700
P_500                   all 0.0280
P_1000                  all 0.0140
```

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde.

Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Dies gilt auch für Quellen aus eigenen Arbeiten.

Ich versichere, dass ich diese Arbeit oder nicht zitierte Teile daraus vorher nicht in einem anderen Prüfungsverfahren eingereicht habe.

Mir ist bekannt, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs mittels einer Plagiatserkennungssoftware auf ungekennzeichnete Übernahme von fremdem geistigem Eigentum überprüft werden kann.

Hagen, den 18.05.2018

_____

pers. Unterschrift