

Maschinelle Indexierung am Beispiel der DNB

Analyse und Entwicklungsmöglichkeiten

Vorab: Was will dieser Vortrag?

- **Analyse von maschinell indexierten Beispielen bei DNB** (v.a. Reihe B)
Ziel: typische Probleme erkennen und daraus Schlussfolgerungen ziehen
 - ➔ *die Überlegungen beziehen sich zunächst nur auf eine bestimmte Methode, zeigen aber auch grundsätzliche Schwierigkeiten auf, die bei der Weiterentwicklung der maschinellen Indexierung zu beachten sind*
- **Maschinelle Indexierung bei der DNB**
beruht auf Software der Fa. Averbis, seit 2014 im Einsatz
 - *basiert auf Titeldaten, Inhaltsverzeichnissen und ggf. Teilen von Volltexten*
 - *computerlinguistische Behandlung der Texte (insbes. Identifikation sinntragender Wörter, Kompositazerlegung und Rückführung auf Stammformen)*
 - *Abgleich mit dem Teilbestand „s“ der GND als Wörterbuch (ebenfalls „vorbehandelt“) und Auswahl von maximal sieben Schlagwörtern*
 - *kein „lernendes“ Verfahren*

Agenda

1. Nicht alle Wörter stehen für Themen
2. Nicht alle Themen kann man an Wörtern ablesen
3. Falsche oder fehlende Erkennung von Konzepten
4. Gewichtung und Auswahl von Schlagwörtern
5. Überlegungen und Anregungen

Agenda

1. Nicht alle Wörter stehen für Themen
2. Nicht alle Themen kann man an Wörtern ablesen
3. Falsche oder fehlende Erkennung von Konzepten
4. Gewichtung und Auswahl von Schlagwörtern
5. Überlegungen und Anregungen

Beispiele für andere Funktionen von Wörtern

Unternehmen* ; Guben* ; Land* ; **Dokument*** ; **Reprivatisierung*** ; Spree* ; Brandenburg* (*maschinell ermittelt)

<http://d-nb.info/1136451161>

200 Jahre Wilhelm Wilke Guben : 1817-2017 : Bilder & **Dokumente** einer wechselvollen Firmengeschichte / Herausgeber Wilhelm Wilke Spedition GmbH

Titelwort bezeichnet Formaspekt

Grußwort vom Ministerpräsidenten des Landes Brandenburg	9
Grußwort vom Landrat des Landkreises Spree-Neiße	10

Die Jahre von 1903 bis zur Reprivatisierung 1990	31
Zur Entwicklung der Spedition seit der Reprivatisierung 1990	52

dient der zeitlichen Untergliederung der Firmengeschichte

Unternehmen* ; Guben* ; Land* ; Dokument* ; Reprivatisierung* ; Spree* ;
 Brandenburg* (*maschinell ermittelt)

<http://d-nb.info/1136451161>

200 Jahre Wilhelm Wilke Guben : 1817-2017 : Bilder & Dokumente einer
 wechselvollen Firmengeschichte / Herausgeber Wilhelm Wilke Spedition
 GmbH

Grußwort vom Ministerpräsidenten des Landes Brandenburg 9

Grußwort vom Landrat des Landkreises Spree-Neiße

*formale Angaben im
 Zusammenhang mit
 Grußwörtern*

Die Jahre von 1903 bis zur Reprivatisierung 1990 31

Zur Entwicklung der Spedition seit der Reprivatisierung 1990 52

1

Etage:

Altdeutsche, Italienische, Französische und Spanische Meister

Führer durch Kasseler Gemäldegalerie unter Aspekt der Provenienzforschung

Provenienzg Geschichten : Gemäldegalerie Alte Meister Schloss Wilhelmshöhe / Texte: Justus Lange, Günther Kuss, Stefanie Rehm

Geschoss <Bauwesen>* ; Meister* ; Gemäldegalerie* ; Spanisch*
(*maschinell ermittelt)

<http://d-nb.info/1136337318>

2

Etage:

Altniederländische, Flämische und Holländische Meister

*150 Geschoss \$g Bauwesen
450 Etage*

3

Etage:

Flämische und Holländische Meister

*„Etage“ hier kein Thema:
Gemäldeführer ist nach den Stockwerken in der Galerie geordnet*

Deutsch-polnische Erinnerungsorte : 10. Deutsch-Polnischer
 Kommunalpolitischer Kongress der Landsmannschaft Ostpreußen : 10.-11.
 Oktober 2015 **Hotel** Warminski **Allenstein** = Nimiecko-polskie miejsca pamięci
 / Herausgeber: Landsmannschaft Ostpreußen e.V. ; Übersetzungen: Kordula
 Gryglak, Antoni Niemczura, Edyta Gładkowska

<http://d-nb.info/1128671891>

Kongress* ; **Allenstein*** ; **Hotel*** (*maschinell ermittelt)

formale Angaben zur Tagung

Voller Esprit und Wissensdurst : Herzogin
 Luise Dorothea von Sachsen-Gotha-
 Altenburg (1710-1767) : mit einer
 kommentierten Edition ihres
 Nachlassinventars / herausgegeben von
 der Stiftung Schloss Friedenstein Gotha

Luise Dorothea, Sachsen-Gotha-Altenburg,
 Herzogin* ; Herzogin* ; **Esprit*** ;
 Inventar* ; Nachlass* ; **Wissbegier*** ;
 Eberle, Martin* ; Netzwerk* (*maschinell
 ermittelt)

<http://d-nb.info/1137880813>

*Die Titelwörter charakterisieren die thematisierte Person,
 sie sind nicht selbst als Thema gedacht*

Namen von VerfasserInnen etc.

- Werden häufig als Thema interpretiert
ergibt falsche Personenschlagwörter

Grußwort: Das Unrecht anerkennen und an die Opfer erinnern

7

Olaf Scholz, Erster Bürgermeister der Freien und Hansestadt Hamburg

Landes-Pflegeanstalt Brandenburg a. H.* ; Hamburg-Langenhorn* ; Patientin* ; Rheinische Kliniken Langenfeld* ; Scholz, Olaf* ; Euthanasie <Nationalsozialismus>* ; Hamburg*
(*maschinell ermittelt)

<http://d-nb.info/1147658048>

GERHART V. GRAEVENITZ

DAS ICH AM ENDE
Strukturen der Ich-Erzählung in Apuleius' *Goldenem Esel*
und Grimmelshausens *Simplicissimus Teutsch*

Ich-Erzählung* ; Apuleius, Madaurensis* ; Das Andere* ; Esel* ; Grimmelshausen* ; Rede* ; Autobiografische Erzählung* ; Graevenitz, Gerhart von* ; Ich-Form* ; Autobiografie* ; Sprecher* ; Dialog <Literaturgattung>* (*maschinell ermittelt)

<http://d-nb.info/1080824693>

- **Dabei häufig Zuordnung zur falschen Person**
da nur Teilbestand „s“ berücksichtigt wird

Die Jahre der Wiedergründung (1999–2003).....
Christian Attig

60 Jahre VDSt Karlsruhe : die Geschichte des Vereins
Deutscher Studenten Karlsruhe von 1957 bis 2017 /
Wolfhart Grote (Hrsg.)

Die neue Blütephase (2003–2007).....72
Christian Roth

Immer weiter, ganz nach vorn – die zweite Vorortszeit (2006/07)88
Konrad Thalheim

Inhaltsverzeichnis

*insgesamt 12 Namen von Beiträ-
gern genannt, einer davon wird
zum Schlagwort*

Geschichtsverein* ; Karlsruhe* ; **Roth, Christian*** ;
Heim* ; Vorstand* (*maschinell ermittelt)

<http://d-nb.info/1131098242>

Roth, Christian, 1873-1934
bayerischer Staatsminister

- Aus Nachnamen werden falsche Schlagwörter extrahiert

Carl Pause	Arma et impedimenta: Die Ausrüstung der römischen Soldaten	5
Norbert Mersch	Überlegungen zur Rekonstruktion einer Militärtunika nach römischen Soldatengrabsteinen des 1. und 2. Jahrhunderts nach Christus	19
Andreas Heinen und Sascha Severin	Belastungsprofil eines Legionärs: Eine sportwissenschaftliche Annäherung	27
Norbert Mersch	Märsche als Selbsterfahrung	

Der Legionär als Leistungssportler : die Leistung römischer Soldaten auf dem Prüfstand / herausgegeben von Carl Pause ; Clemens-Sels-Museum Neuss

Leistung <Recht>* ; Soldat* ; Leistungssportler* ; Lastganglinie* ; Prüfstand* ; **Mersch*** (*maschinell ermittelt)

<http://d-nb.info/1049547829>

Ort in Luxemburg

Agenda

1. Nicht alle Wörter stehen für Themen
2. Nicht alle Themen kann man an Wörtern ablesen
3. Falsche oder fehlende Erkennung von Konzepten
4. Gewichtung und Auswahl von Schlagwörtern
5. Überlegungen und Anregungen

Der Maschine fehlt das „Einordnungswissen“

Carl Pause	Arma et impedimenta: Die Ausrüstung der römischen Soldaten	5
Norbert Mersch	Überlegungen zur Rekonstruktion einer Militärtunika nach römischen Soldatengrabsteinen des 1. und 2. Jahrhunderts nach Christus	19
Andreas Heinen und Sascha Severin	Belastungsprofil eines Legionärs: Eine sportwissenschaftliche Annäherung	27
Norbert Mersch	Märsche als Selbsterfahrung	

Der Legionär als Leistungssportler : die Leistung römischer Soldaten auf dem Prüfstand / herausgegeben von Carl Pause ; Clemens-Sels-Museum Neuss

- **Zahlreiche Indizien, die ein Mensch wahrnimmt**
machen Kontext „Römisches Reich“ und „Experimentelle Archäologie“ deutlich
(Basis Kombination aus Weltwissen und Sprachwissen)

Vergleich maschinelle und intellektuelle Erschließung dieses Titels:

DNB:

Leistung <Recht>* ; Soldat* ; Leistungssportler* ; Lastganglinie* ;
Prüfstand* ; Mersch* (*maschinell ermittelt)

<http://d-nb.info/1049547829>

GBV:

*Römisches Reich ☘ ; Soldat ☘ ; Experiment ☘ ; Archäologie ☘

<https://gso.gbv.de/DB=2.1/PPNSET?PPN=78266167X>

hbz:

- [Römisches Reich \[\(DE-588\)4076778-4\]](#)
- [Soldat \[\(DE-588\)4055409-0\]](#)
- [Körperliche Belastung \[\(DE-588\)4031594-0\]](#)

Optimal wäre Kombination:

Römisches Reich ; Soldat ; Körperliche Belastung ; Experiment ; Archäologie

1

Etage:
Altdeutsche, Italienische, Französische und
Spanische Meister

Provenienzg Geschichten : Gemäldegalerie Alte Meister Schloss
Wilhelmshöhe / Texte: Justus Lange, Günther Kuss, Stefanie Rehm

Geschoss <Bauwesen>* ; **Meister*** ; Gemäldegalerie* ; **Spanisch***
(*maschinell ermittelt)

<http://d-nb.info/1136337318>

2

Etage:
Altniederländische, Flämische und Holländische
Meister

„Einordnungswissen“:
*Formulierungen wie „Spanische Meister“
stehen hier für Gemälde bedeutender Maler*

3

Etage:
Flämische und Holländische Meister

Vorschlag für intellektuelle Erschließung:
Gemäldegalerie Alte Meister Kassel ; Gemälde ; Provenienzforschung

Agenda

1. Nicht alle Wörter stehen für Themen
2. Nicht alle Themen kann man an Wörtern ablesen
3. Falsche oder fehlende Erkennung von Konzepten
4. Gewichtung und Auswahl von Schlagwörtern
5. Überlegungen und Anregungen

Probleme bei der Disambiguierung

Leitfaden für die **Markierung** von Wanderwegen : für alle ehrenamtlich Tätigen im Fachbereich Wege des Schwäbischen Albvereins

Wanderweg* ; Beschilderung* ; Unfallversicherung* ; Rundwanderweg* ;
Markierung <Chemie>* (*maschinell ermittelt)

<http://d-nb.info/1151148180>

005 Tsv1
006 <http://d-nb.info/gnd/4443932-5>
011 s
012 w;o
035 [gnd/4443932-5](http://d-nb.info/gnd/4443932-5)
039 [swd/4443932-5](http://d-nb.info/swd/4443932-5)**\$vzg**
040 ger
065 1
150 Markierung
450 Markieren**\$vB** 1986 1.
670 B 1986 als Markieren
677 durch ein Zeichen kenntlich machen
680 Verknüpfe z.B. mit Wanderweg

Falsche Zuordnung, obwohl im korrekten Datensatz sogar das Wort „Wanderweg“ auftaucht.

Probleme durch Wortstämme und Zerlegung

Kläranlagen in der Energiewende: **Faulung** optimieren & Flexibilität wagen

Energienetz ; Flexibilität ; Schlammstabilisierung ; Faulgas ;
Energiespeicher ; Power-to-gas
Kläranlage* ; **Müßiggang*** ; Biogas* ; Erneuerbare
Energien* ; Energiepolitik* ; Klärschlammstabilisierung* ;
Schlammbehandlung* (*maschinell ermittelt)

<http://d-nb.info/1149512849>

150 Müßiggang
450 Faulenzen

faul₁ vs. faul₂

„Faulung“ ist eben nicht dasselbe wie „Faulenzen“ (trotz gleichen Wortstamms)

60 Jahre VDSt Karlsruhe : die **Geschichte des Vereins**
Deutscher Studenten Karlsruhe von 1957 bis 2017 /
Wolfhart Grote (Hrsg.)

Geschichtsverein* ; Karlsruhe* ; Roth, Christian* ;
Heim* ; Vorstand* (*maschinell ermittelt)

<http://d-nb.info/1131098242>

geschicht* verein*

reicht nicht zur Differenzierung „Geschichte des Vereins“ vs. „Geschichtsverein“

Nichtererkennung von „Named entities“

- **Namen von Körperschaften und Personen sowie Werktitel** werden *meistens nicht erkannt, auch wenn sie in der GND enthalten sind*
 ➔ führt zu unbefriedigenden, manchmal völlig irreführenden Schlagwörtern

Provenienzgeschichten : Gemäldegalerie Alte Meister Schloss Wilhelmshöhe / Texte: Justus Lange, Günther Kuss, Stefanie Rehm

Geschoss <Bauwesen>* ; Meister* ; Gemäldegalerie* ; Spanisch*
 (*maschinell ermittelt)

<http://d-nb.info/1136337318>

80. Tagung der Arbeitsgemeinschaft Norddeutscher Geologen, 6.-9. Juni 2017 in Rendsburg : Tagungsband und Exkursionsführer / Geologischer Dienst des Landes Schleswig-Holstein

Geologe* ; Norden* ; Salzstock* ; Hydrogeologie* ; Grube* ; Bauer, Sebastian* ; Sediment* ; Exkursion* ; Jasmund* ; Sandstein* (*maschinell ermittelt)

<http://d-nb.info/113649944X>

'68 : Stichworte Marburg A - Z / Christoph Becker, Kathrin Bonacker

Marburg* ; **Luther, Martin*** ; Prager Frühling* ; Maiunruhen <1968>* ;
Verführer* (*maschinell ermittelt)

<http://d-nb.info/1156209951>

Frage:

*Wie könnte man der
Maschine denn noch
genauer sagen, um
welche Person es
geht?*

Lexikon zu 1968

*(teilweise aus spezifisch
Marburger Sicht)*

Indien	51
Institutsbesetzungen	51
„Joli“ oder: Gotteslästerung in Marburg	53
Kennedy, Robert Francis (1925–1968) und King, Martin Luther (1929–1968)	55
Klarsfeld, Beate (*1939)	56
„Klasse Kampf – '68 erinnern“	57
Kolle, Oswald (1928–2010)	58
Kommune 1	58
Kuppelei	61
Lederjacken	63

Landgraf Carl (1654-1730) : fürstliches Planen und Handeln zwischen Innovation und Tradition / herausgegeben von Holger Th. Gräf, Christoph Kampmann und Bernd Küster

Landgraf* (*maschinell ermittelt)

<http://d-nb.info/1147882894>

Karl, Hessen-Kassel, Landgraf von, 1654-1730 sollte aus Informationen in Titel und Inhaltsverzeichnis abzuleiten sein

GERHART V. GRAEVENITZ

DAS ICH AM ENDE
Strukturen der Ich-Erzählung in Apuleius' Goldenem Esel
und Grimmelshausens Simplicissimus Teutsch

<http://d-nb.info/1080824693>

Ich-Erzählung* ; **Apuleius, Madaurensis*** ; Das Andere* ; **Esel*** ; **Grimmelshausen*** ;
Rede* ; Autobiografische Erzählung* ; Graevenitz, Gerhart von* ; Ich-Form* ;
Autobiografie* ; Sprecher* ; Dialog <Literaturgattung>* (*maschinell ermittelt)

Keins der Werke wird erkannt. Nur einer der Autoren wird erkannt („Grimmelshausen“ ist keine Person, sondern ein Ort)

- **Namen von „Named entities“, die nicht in der GND sind**

Besonderheit wird vom Algorithmus nicht erkannt, ergibt unsinnige Schlagwörter

Der Haigst - ein Spaziergang durch Geschichte und Gegenwart /
Stefan Clarenbach, Ulrich Schülke

Spaziergang* ; Lehen* ; Westen* ; Berg* ; **Santiago de Chile*** ;
Meistersang* ; Grenzstein* ; Maler* (*maschinell ermittelt)

<http://d-nb.info/1119549752>

*Es handelt sich um einen
Führer durch den „Haigst“,
einen Teil von Stuttgart-
Degerloch*

1	Die Haigstkirche	36
2	Im oberen Lehen	42
3	Der Santiago-de-Chile-Platz	46
4	Der Grenzstein	
	Ob der Steig – Unter der Steig	
5	Tell Geck – der Maler vom Haigst	
6	Die Wielandshöhe	
7	Die kleine Kuppe Haigst	

8	Am oberen Berg – Alte Wege im Westen	66
9	Der Gressly vom Haigst ?	70
10	Schimmelhütte und Scharrenberg	76
11	Die Meistersingerstraße	80
12	Stolpersteine zum Gedenken...	92
13	Der Steigensee	98
14	Die Kauzenhecke	106

Uneigentlicher Wortgebrauch

- Wird von der Maschine nicht erkannt
der Algorithmus nimmt alles wörtlich

Mein **Kompass**: Wegweiser zur Gesundheit für Dich / Peter Brill

Gesundheit* ; **Kompass*** ; Therapeut* ; Tastsinn* ; Humor* ;
Geborgenheit* ; Gesundheitsstörung* (*maschinell ermittelt)

<http://d-nb.info/1049547829>

Geldanlage für **Faule** / Sina Groß

Kapitalanlage* ; **Müßiggang*** ;
Altersversorgung* ; Fonds* ; Eigenheim*
(*maschinell ermittelt)

<http://d-nb.info/1070846937>

Agenda

1. Nicht alle Wörter stehen für Themen
2. Nicht alle Themen kann man an Wörtern ablesen
3. Falsche oder fehlende Erkennung von Konzepten
4. Gewichtung und Auswahl von Schlagwörtern
5. Überlegungen und Anregungen

Häufigkeit des Vorkommens als wichtiges Kriterium

Akupunktur als Behandlungsmöglichkeit bei Depression in der Schwangerschaft : systematische Übersichtsarbeit / von Julia Schmidt

Akupunktur* ; Depression* ; Schwangerschaft* (*maschinell ermittelt)

<http://d-nb.info/1140020501>

Beispiel für eine gut gelungene Indexierung:
die zentralen Begriffe kommen häufig im Inhaltsverzeichnis vor (13mal „Depression“, 9mal „Akupunktur“, 6mal „Schwanger“); diese Situation ist jedoch eher selten*

3	Akupunktur	53
3.1	Lehre und Vorstellung der Akupunktur	53
3.2	Geschichte der Akupunktur	60
3.3	Wirkweise der Akupunktur	61
3.4	Depression aus Sicht der TCM	62
3.5	Die Energiebilanz der Schwangerschaft aus Sicht der TCM	64
3.6	Mögliche Punktekombinationen bei Depression in der Schwangerschaft und zu unterlassende Punkte	64
4	Ziel der Arbeit	69
5	Methoden	70
6	Ergebnisse	72

Spezifika von Inhaltsverzeichnissen

- **Viele Wörter dienen der Strukturierung des Textes**
bezeichnen also keine Themen
 - **Zentrale Themen kommen in Überschriften oft nicht mehr vor**
weil sie schon im Titel benannt wurden und nun vorausgesetzt werden; häufig sind deshalb nur Unteraspekte genannt
 - **Gewichtung nach Häufigkeit des Vorkommens ist problematisch**
 - *oft kommen alle Wörter nur ein- bis zweimal vor*
 - *auch unwichtige Wörter können mehrfach vorkommen, z.B. in Anhängen*
 - *eigentlich müsste der Umfang der Kapitel mit einbezogen werden*
- ➔** *vielfach werden nicht die substantiell behandelten Themen abgebildet, sondern eine willkürlich erscheinende Auswahl von Einzelaspekten (dies führt häufig zu Ballast bei Recherchen)*

'68 : Stichworte Marburg A - Z / Christoph Becker, Kathrin Bonacker

Marburg* ; Luther, Martin* ; Prager Frühling* ; Maiunruhen <1968>* ;
Verführer* (*maschinell ermittelt)

<http://d-nb.info/1156209951>

Drogen	28
Dutschke, Rudi (1940–1979)	30
Ehe	31
Ensslin, Gudrun (1940–1977)	32
Farbfernsehen	33
Film	34
Flokati	35
Flugblatt	36
Frauenbewegung	37
Gassmann, Georg (1910–1987)	38
„Geheime Verführer“	39
„Hair“	40
Hippies	41

„The hidden persuaders“ ist ein Buchtitel (über Werbung)

Lexikon zu 1968:

*ca. 60 Einträge (max. 2 S.),
darunter allein 14 Personen;
berücksichtigt werden vier
Einträge (zwei davon falsch)*

Meinhof, Ulrike (1934–1976)	69
Minirock	70
Notstandsgesetze	72
NS-Vergangenheit	74
Obermaier, Uschi (*1946)	77
Olympia 1968	79
Ordinarien	80
Osterunruhen	82
Paris, Mai '68	86
Pillenknick	87
Pop Art	88
Prager Frühling	89
„Quartett im Bett“	91

125 Jahre Schwäbischer Albverein Ortsgruppe Crailsheim :
1892-2017

Ortsverein* ; Crailsheim* ; Familiengruppe* ; Jugendgruppe* ;
Turnhalle* ; **Nordic Walking*** (*maschinell ermittelt)

<http://d-nb.info/1135937346>

Soll bei einer Recherche nach „Nordic Walking“ die Festschrift einer Ortsgruppe des Schwäbischen Albvereins erscheinen?

Grußworte des Präsidenten des Schwäbischen Albvereins .

Grußworte des Gauvorsitzenden Andreas Raab

Grußworte des Oberbürgermeisters Rudolf Michl

Grußworte des Vorstandsteams

Die Ziele des Vereins

Die Geschichte der Ortsgruppe in den Anfangsjahren

Das Urgestein W. Andrassy 1928 - 1965

Die Aera von Walter Strähle 1965 - 1995

Die 75 Jahr Feier im Turnhallensaal Altenmünster

Die 100 Jahr Feier in der Festhalle Ingersheim

Albvereinsfest des Schwäb. Albvereins 2001 in Crailsheim .

Die Mobilität nimmt zu, Wechsel auch beim Albverein

Zahlen, Fakten, Daten zur Ortsgruppe Crailsheim.....

Wege und Markierungen: vom Farbeimer zum GPS

Aufgaben des Wanderwarts

Von der Nordsee bis nach Südtirol

Jugendgruppe im Wandel der Zeit

Familiengruppe: Die Junge Familie

Radfahren, umweltschonend durch die Natur

****Nordic Walking,** fit mit zwei Stöcken (oder auch ohne)**

Die Natur erhalten, der Naturschutz in der Ortsgruppe

Gesundheitswandern, wandern für Körper, Seele und Geist ..

Frauenwandern im Trend

Seniorengruppe, das Durchschnittsalter wird höher

Bergfahrten

Der Rechner - eine vertrauensvolle Aufgabe

Die Zukunft des Schwäbischen Albvereins

... und der Ortsgruppe


Firmen die unseren Verein unterstützten und unterstützen ..

Kapitel jeweils 1-2 Seiten

Agenda

1. Nicht alle Wörter stehen für Themen
2. Nicht alle Themen kann man an Wörtern ablesen
3. Falsche oder fehlende Erkennung von Konzepten
4. Gewichtung und Auswahl von Schlagwörtern
5. Überlegungen und Anregungen

Drei grundsätzliche Ratschläge

- **Mit Clustering unnötige maschinelle Erschließung vermeiden**
Übernahme von intellektueller Sacherschließung aus früheren Auflagen bzw. aus Reihe A, wenn eine Veröffentlichung parallel in- und außerhalb des Buchhandels erscheint (z.B. Ausstellungskataloge)
 *wünschenswert wäre außerdem Übernahme intellektueller Erschließung aus Verbänden (dann evtl. maschinelle Indexierung überschreiben?)*
- **Mit mehrdimensionaler Technik an die Aufgabe herangehen**
derzeitiges System nicht ausreichend; sollte nur ein Baustein einer umfassenderen Methodik sein (z.B. lernende Verfahren; Einbezug weiterer Quellen)
- **Nicht so tun, als würde das gewohnte Ergebnis herauskommen**
 - *Beschränkung auf sieben Schlagwörter aufgeben
(spiegelt fälschlich vor, dass nur substantziell Behandeltes abgebildet wird)*
 - *Rechercheoptionen auf die neue Situation abstimmen
(Ranking, evtl. Option zum Ausschluss maschineller Schlagwörter)*

Ideen für das „Named entity“-Problem

- **Abgleichlauf ohne linguistische Verfahren vorschalten**
zur Erkennung von Körperschaften etc., die in der GND vorhanden sind
- **Informationen aus der Formalerschließung nutzen, z.B.**
 - *als VerfasserInnen etc. erfasste Personen nicht als Schlagwort zulassen (Ausnahmen wären bei Formangaben wie „Tagebuch“ möglich)*
 - *in der Formalerschließung erfasste Körperschaften auf Themenhaftigkeit abprüfen*
- **Layout-Analyse von Inhaltsverzeichnissen durchführen**
i.d.R. lassen sich Namen von Beiträgern von Überschriften unterscheiden
- **Erkennung von Named entities außerhalb der GND ermöglichen**
 - *Einbezug weiterer Quellen für den Abgleich (DBpedia, GoogleMaps...)*
 - *zusätzliche Regeln für häufige Fälle, z.B. „...straße“ oder „...platz“*

Nötig sind umfangreiche Plausibilitätsprüfungen

- **Passen die ermittelten Schlagwörter in den fachlichen Kontext?**
Plausibilitäten über vorhandene Information (auch aus Verbänden) abprüfen, z.B.:
 - *welche Notationen wurden in unterschiedlichen Systematiken vergeben?*
 - *in welchen Fachgebieten haben beteiligte Personen bisher publiziert?*

005 Tsv1
 006 <http://d-nb.info/gnd/4328972-1>
 011 s
 012 w;z
 035 gnd/4328972-1
 039 swd/4328972-1\$vzg
 065 **31.5;31.9a**
 083 621.042\$d1\$t2009-06-08
 150 Lastganglinie
 450 Belastungskurve
 450 Lastgangkurve
 450 Lastgangkennlinie
 450 Lastprofil\$gEnergietechnik
 670 B 1986
 680 benutzt in allen Bereichen der Energietechnik, nicht
 nur für Elektroenergietechnik

Leistung <Recht>* ; Soldat* ; Leistungssportler* ; **Lastganglinie*** ;
 Prüfstand* ; Mersch* (*maschinell ermittelt)


<http://d-nb.info/1049547829>

*Wie wahrscheinlich ist ein Schlagwort aus der
 Elektrotechnik, wenn der Herausgeber
 Archäologe ist und einschlägig publiziert hat?*

- **Prüfen von abweichenden Mustern bei maschineller Indexierung**
z.B. *sehr viel häufigere Vergabe eines Schlagworts bei maschineller Indexierung als bei intellektueller Erschließung spricht für systematischen Fehler*

Beispiel: SW „Sprachbezeichnung“ intellektuell von DNB 6mal vergeben, maschinell 12mal (davon nur einmal korrekt)

Thema in	6 Publikationen
	<ol style="list-style-type: none"> 1. <i>Indices to Bantu languages</i> Maho, Jouni F.. - Muenchen : LINCOM Europa, 2008 2. <i>Die Sprachbezeichnungen Latinus und Romanus im Lateinischen und Romanischen</i> Kramer, Johannes. - Berlin : Erich Schmidt, 1998 3. ...
Maschinell verknüpft mit	12 Publikationen
	<ol style="list-style-type: none"> 1. <i>Zur Konstruktion der magischen Welt in den Harry-Potter-Romanen von Joanne K. Rowling</i> Gottschalk, Nina. - München : GRIN Verlag, 2018, 1. Auflage 2. <i>Der Titel ist ein Trailer. Untersuchungen zu Titeln pornographischer Videofilme</i> Kaczmarek, Ludger. - Kiel : Universitätsbibliothek Kiel, 2016 3. ...

 *nur möglich, wenn weiterhin auch intellektuell erschlossen wird*

Persönliches Fazit:

Die derzeitige maschinelle Indexierung der DNB zeigt sehr deutliche Begrenzungen. Sie kann nicht der Endpunkt der Entwicklung sein, sondern nur ein Schritt auf einem wohl noch sehr langen Weg hin zu brauchbaren Systemen für die automatische Schlagwortvergabe.

Literaturhinweise zum Verfahren der DNB:

- *Elisabeth Mödden, Christa Schöningh-Walter, Sandro Uhlmann: Maschinelle Inhaltserschließung in der Deutschen Nationalbibliothek. In: BuB 70 (2018), Heft 1, S. 30-35*
<https://b-u-b.de/wp-content/uploads/2018-01.pdf>
- *Elisabeth Mödden: Die maschinelle Erschließung der Deutschen Nationalbibliothek, Vortrag am 8. Mai 2017 in Stuttgart,*
http://www.ub.uni-stuttgart.de/wirueberuns/downloads/veranstaltungen/da-workshop/Moedden_Vortrag_Stuttgart_DNB_2017-05-08.pdf



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>)