

Kempf, Andreas Oskar

**Book Chapter — Published Version**

**Thesauri**

*Suggested Citation:* Kempf, Andreas Oskar (2023) : Thesauri, In: Kuhlen, R. Semar, W. Womser-Hacker, C. Lewandowski, D. (Ed.): Grundlagen der Informationswissenschaft, 7. Auflage, De Gruyter Saur, Berlin, pp. 183-194,  
<https://doi.org/10.1515/9783110769043-016>

This Version is available at:

<http://hdl.handle.net/11108/545>

**Kontakt/Contact**

ZBW – Leibniz-Informationzentrum Wirtschaft/Leibniz Information Centre for Economics  
Düsternbrooker Weg 120  
24105 Kiel (Germany)  
E-Mail: [info@zbw.eu](mailto:info@zbw.eu)  
<https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw>

**Standard-Nutzungsbedingungen:**

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.*



<https://creativecommons.org/licenses/by/4.0/>

Andreas Oskar Kempf

## B 4 Thesauri

### 1 Einleitung

In der Informationswissenschaft stehen Thesauri für kontrollierte und strukturierte Vokabulare, in denen Begriffe, verstanden als geistige Einheiten, durch Bezeichnungen der natürlichen Sprache repräsentiert werden (International Organization for Standardization – ISO 2011, S. 12). Jedem Begriff werden eine Vorzugsbenennung, ein sog. Deskriptor, der bei der Inhaltserschließung als Schlagwort vergeben wird, und weitere bedeutungsgleiche oder -ähnliche Bezeichnungen, sog. Nicht-Deskriptoren, als Zugangsvokabular bzw. alternative Sucheinstiege zugewiesen. Dieser Art werden der Variabilität und der Mehrdeutigkeit natürlicher Sprache Rechnung getragen. Darüber hinaus werden zwischen Begriffen bzw. ihren Bezeichnungen spezifische, reziproke Relationen kenntlich gemacht (s. Abschnitt 1), die die Bedeutungsbeziehungen bzw. das „semantische Gefüge“ zwischen den Begriffen aufzeigen (Deutsches Institut für Normierung – DIN 1987, S. 10). Diese Kernprinzipien dieser Wissensorganisationsmethode dienen sowohl auf Seiten der Inhaltserschließenden als auch auf Seiten der in einem Informationssystem Recherchierenden in unterschiedlicher Weise der Benutzerführung und Suchunterstützung. Als Grundlage für semantisches Retrieval etwa sorgen sie bei automatischer Erweiterung der Suchanfrage um die hinterlegten Nicht-Deskriptoren für eine erfolgreiche Suche weitgehend unabhängig vom konkreten Suchterm.

Dadurch, dass Begriffe bzw. Konzepte eines bestimmten Gegenstandsbereichs – häufig handelt es sich um eine bestimmte Fachdisziplin oder Wissensdomäne – die wesentlichen Bezugseinheiten eines Thesaurus darstellen, eignen sich Thesauri, anders als Klassifikationen, die oft eine Ressource auf einem relativ hohen Abstraktionsniveau beschreiben und vielfach als Instrument zur Aufstellung von Dokumentbeständen verwendet werden, für die inhaltliche Feinerschießung einer dokumentarischen Bezugseinheit (DBE). Notwendig wurde dies unter anderem dadurch, dass parallel zu dem weltweit zu verzeichnenden enormen Anstieg an Forschung und Entwicklung ab Mitte des 20. Jahrhunderts unselbständige Werke, wie etwa Zeitschriftenaufsätze, die häufig eine Vielzahl unterschiedlicher Aspekte und Perspektiven thematisierten, zunahmen und damit nach einer flexibleren und ausdrucksfähigeren Dokumentationsprache gesucht wurde (Bertram 2005, S. 217, 255–256).

In diesem Beitrag werden zunächst die wesentlichen Strukturprinzipien eines Thesaurus genauer vorgestellt. Da die Methodik im Kern unverändert geblieben ist, sei ergänzend auf das diesbezüglich ausführlichere Kapitel B 2 Thesaurus aus KSS-5 verwiesen (Burkart 2004). Anschließend wird ausgeführt, wie sich aufgrund geänderter Rahmenbedingungen und technischer Umgebungen oftmals der konkrete Anwendungskontext und damit auch der Bezugsrahmen von Thesauri geändert hat. Vor diesem Hintergrund werden abschließend zentrale Herausforderungen und zukünftige Aufgabenfelder für den Einsatz und die Weiterentwicklung von Thesauri angerissen.

## 2 Terminologische und begriffliche Kontrolle

Die Verwendung der natürlichen Sprache bei dieser Methode der Wissensorganisation macht es erforderlich, mit Mehrdeutigkeiten, der großen Variationsbreite sprachlicher Ausdrücke und dem Umstand, dass Bezeichnungen aus mehreren Bedeutungsbestandteilen zusammengesetzt sein können, umzugehen. Zusammengefasst wird der methodische Umgang damit als *terminologische Kontrolle* bezeichnet. Im Einzelnen handelt es sich dabei um die Kontrolle von Homonymen und Polysemen sowie um Synonym- und Zerlegungskontrolle. Vorab gilt es allerdings, im Zuge der sog. Normung einheitliche Festlegungen hinsichtlich der Ansetzung bzw. äußeren Form von Deskriptoren zu treffen. Dies betrifft etwa die Wortfolge, Wortart sowie Kasus, Numerus und Groß- und Kleinschreibung. Auch muss entschieden werden, ob etablierte Formulierungen aus anderen Sprachen, etwa dem Englischen, übernommen werden (Stock & Stock 2008, S. 244).

Bei der *Homonym- bzw. Polysemkontrolle* wird mehrdeutigen sprachlichen Ausdrücken eine eindeutige Bedeutung zugewiesen. Dies kann dadurch erfolgen, dass nur eine der Bedeutungen im Thesaurus aufgenommen und andere ausdrücklich ausgeschlossen werden, oder die mehrdeutige Bezeichnung durch eine alternative eindeutige Bezeichnung ersetzt wird. In anderen Fällen kann die Disambiguierung durch einen sog. Homonymzusatz in Klammern erfolgen, der verdeutlicht, mit welcher Bedeutung die Bezeichnung im Thesaurus verwendet wird, z. B. *Zug (Kanton)*.

Die *Synonymkontrolle* bezeichnet die Auswahl weiterer Bezeichnungen, die neben der Vorzugsbenennung ebenfalls den Begriff repräsentieren. Die ausgewählten Bezeichnungen zusammengefasst bilden die sog. Äquivalenzklasse. Für die geordnete Darstellung der Bezeichnungen und Begriffe gibt es genormte Relationenkürzel (siehe Tabelle 1). Neben bedeutungsgleichen können auch bedeutungsähnliche Bezeichnungen, sog. Quasi-Synonyme, darin enthalten sein. Des Weiteren kann es sich auch um spezifischere Bezeichnungen handeln, bei denen davon ausgegangen wird, dass sich ihre Aufnahme als Repräsentation für ein eigenes Konzept nicht lohnen würde, da in dem entsprechenden Informationssystem, in dem der Thesaurus eingesetzt wird, nicht genug Ressourcen zu einem Thema vorhanden sind (ISO 2011, S. 45–47).

Bei der *Zerlegungskontrolle* wird überprüft, inwiefern Mehrwortausdrücke, die aus unterschiedlichen Bedeutungsbestandteilen bestehen, präkombiniert oder postkoordiniert Eingang in das Thesaurusvokabular finden sollten. Ersteres meint, dass die Bedeutungsbestandteile zusammengesetzt bleiben (z. B. *Leiharbeitskräfte*). Letzteres steht dafür, dass die komplexe Benennung in einzelne Bedeutungsbestandteile zerlegt wird (z. B. *Leiharbeit + Arbeitskräfte*). Anders als bei einer morphologischen Zerlegung, die die Ausgangsbenennung in die einzelnen Wortbestandteile zerlegt (z. B. *Heimtextilien* in *Heim + Textilien*), handelt es sich im Rahmen der terminologischen Kontrolle um eine begriffsorientierte Zerlegung (für das Beispiel *Heimtextilien* z. B. *Haushalt + Textilien*). Dass Deskriptoren aus einer morphologischen Zerlegung hervorgehen, ist nur denkbar, wenn sie auch semantisch korrekt sind (z. B. *Fischereiprodukt – Fischerei + Produkt*). Bei der semantischen Zerlegung lässt sich des Weiteren zwischen einer disjunkten Zerlegung, bei der die Begriffe, in die zerlegt wurde, keine Bedeutungsüberschneidung aufweisen, und einer konjunkten Zerlegung, bei der dies der Fall ist, unterscheiden. Bertram (2005, S. 223) führt hierzu das Beispiel *Informationsbedarfsanalyse* an. Als disjunkte Zerlegung sei die Teilung in *Information* und *Bedarfsanalyse* denkbar. Eine konjunkte Zerlegung würde eine Zerlegung in *Informationsbedarf* und *Bedarfsanalyse* vorsehen. Auch wenn hierdurch Redundanzen erzeugt würden, beuge dies fehlerhaften Verknüpfungen vor.

Unterschiedliche Aspekte, deren Gewichtung je nach Gegenstandsbereich des Thesaurus variieren kann, sollten insbesondere aus Sicht der Nutzenden bei der Zerlegungskontrolle berücksichtigt werden. Einerseits hält eine Zerlegung den Thesaurus übersichtlich und schafft zusätzliche Einstiegspunkte, wodurch sich die Belegungshäufigkeit der einzelnen Deskriptoren allerdings erhöhen kann. Andererseits besteht die Gefahr fehlerhafter Verknüpfungen, die bei der Recherche Ballast erzeugen können. Stets, so Bertram (2005, S. 219), handele es sich um einen „Balanceakt“. Je stärker zerlegt wird, desto größer sind der semantische Informationsverlust und die Gefahr einer sinnverfälschenden Zerlegung. Zu prüfen ist etwa, ob es sich bei einer komplexen Benennung um einen Fachausdruck handelt oder ob eine Benennung durch eine Kombination bereits bestehender Deskriptoren gebildet werden kann. Auch lässt sich für den Gegenstandsbereich des Thesaurus mitunter zwischen einem Kern- und einem Randbereich unterscheiden. In diesem Fall ist zu empfehlen, für den Kernbereich Inhalte tendenziell zu präkombinieren, um einen hohen Grad an Spezifität zu gewährleisten. In einem wirtschaftswissenschaftlichen Fachthesaurus das Teilgebiet Haushaltsökonomik zu zerlegen, ist vermutlich wenig sinnvoll, da die Verwendung der einzelnen Begriffsbestandteile bei der Suche auch Rechercheergebnisse erzeugen kann, die sich nicht mit dieser Teildisziplin beschäftigen. Bei mehrsprachigen Thesauri kann bei der Abwägung zusätzlich der Aspekt eine Rolle spielen, inwiefern eine Bezeichnung auch in den weiteren Sprachen des Thesaurus etabliert ist. Je nach Anspruch und Zielgruppe sollte etwa auf Wortprägungen verzichtet werden, die lediglich für einen Sprachraum gelten (Stock & Stock 2008, S. 244–245).

Die *begriffliche Kontrolle* stellt das zweite wesentliche Strukturmerkmal eines Thesaurus dar. Sie steht für die Auszeichnung zentraler semantischer Beziehungen zwischen den Begriffen mit dem Ziel, die semantische Struktur des Begriffsumfelds kenntlich zu machen. Die sog. *Äquivalenzrelation* besteht zwischen den einzelnen Bezeichnungen der oben bereits erwähnten Äquivalenzklasse – es handelt sich hier folglich eigentlich um einen Relationstyp zwischen einzelnen Bezeichnungen und nicht zwischen Begriffen. Daneben gibt es *hierarchische Relationen* in einem Thesaurus, die zwischen den Begriffen angelegt werden können. Sie können entweder als einfache Ober- und Unterbegriffsrelationen oder weiter spezifiziert angelegt werden. In diesem Fall wird zwischen einer generischen bzw. Gattung-Art-Relation und einer partitiven bzw. Teil-Ganzes-Relation unterschieden. Bei der generischen Relation oder auch Abstraktionsrelation besitzt der untergeordnete alle Merkmale des übergeordneten Begriffs und zusätzlich mindestens noch ein weiteres Merkmal (z. B. *Bargeldloser Zahlungsverkehr – Elektronisches Geld*). Bei der partitiven Relation entspricht der übergeordnete Begriff einem Ganzen und der untergeordnete Begriff einem Bestandteil dieses Ganzen (z. B. *Schiffstechnik – Schiffsantrieb*). Für den Fall, dass ein Begriff mehrere Ober- oder Unterbegriffsrelationen aufweist, wird von einem polyhierarchischen Thesaurus gesprochen. Bei der Suche ermöglichen die hierarchischen Untergliederungen, die Suchanfrage allgemeiner oder präziser gefasst zu formulieren. Schließlich gibt es die *Assoziations-* oder auch *Verwandtschaftsrelation*. Ausschlaggebend für das Anlegen dieses Relationstyps sollte sein, dass sich Begriffe semantisch nah sind oder häufig gemeinsam auftreten, weswegen es für die Inhaltserschließung und Recherche aus dem Thesaurus heraus besonders wertvoll sein kann, sie aufzuführen. Eine Verwandtschaftsrelation ist etwa zwischen einem Wirtschaftszweig und einem entsprechenden Produkt (z. B. *Chemieindustrie – Chemikalie*) oder zwischen einer Fachdisziplin und ihrem Gegenstand (z. B. *Haushaltsökonomik – Privater Haushalt*) denkbar.

Home  
Alphabetische Liste der Deskriptoren  
Mappings  
Versionen  
Web Services  
Downloads  
Mehr zum STW

- ▶ V Volkswirtschaft
- ▶ B Betriebswirtschaft
- ▶ W Wirtschaftssektoren
- ▶ P Produkte
- ▶ N Nachbarwissenschaften
- ▶ G Geographische Begriffe
- ▶ A Allgemeinwörter

## Wirtschaftsstruktur EB

**Economic structure** (engl.)

benutzt für: Sektorale Wirtschaftsstruktur, Branchenstruktur, Sectoral structure

*Für die Struktur einer einzelnen Branche BENUTZE Marktstruktur*

*für die Struktur einer einzelnen Region BENUTZE Regionale Wirtschaftsstruktur*

### Unterbegriffe

- Agrarstruktur EB
- Beschäftigungsstruktur EB
- Betriebsgrößenstruktur EB
- Industriestruktur EB
- Marktstruktur EB
- Produktionsstruktur EB
- Regionale Wirtschaftsstruktur EB

### Verwandte Begriffe

- Branchenklassifikation EB
- Input-Output-Analyse EB
- Mehrsektoren-Modell EB
- Mesoökonomie EB
- Strukturpolitik EB
- Strukturwandel EB

### Thesaurus Systematik

- V.04.03 Wirtschaftsstruktur ▼
- W.00 Wirtschaftssektoren ▼

### Links zu anderen Thesauri und Vokabularen

- = Wirtschaftsstruktur (aus GND)
- > Branchenstruktur (aus GND)
- > Kommunale Wirtschaftsstruktur (aus GND)
- > Sektorale Wirtschaftsstruktur (aus GND)
- ~ Produktionsweise (aus GND)
- ~ Struktur Anpassung (aus GND)
- ~ Strukturkrise (aus GND)
- ~ Strukturberichterstattung (aus GND)
- = Wirtschaftsstruktur (aus Wikidata) **W**
- = Wirtschaftsstruktur (aus EuroVoc)
- = Wirtschaftsstruktur (aus TheSoz)
- > Monostruktur (aus TheSoz)
- = Wirtschaftsstruktur (aus AGROVOC)

### Persistenter Identifier (für Bookmarks und zum Verlinken)

- <http://zbw.eu/stw/descriptor/11782-3>

Standard-Thesaurus Wirtschaft (v 9.12, 2021-10-15) · Anregungen und Kommentare an das Thesaurus Team ·  
Mailinglisten: stw-announce, stw-user  
ZBW - Leibniz-Informationszentrum Wirtschaft - Impressum - Datenschutz



Der Standard-Thesaurus Wirtschaft steht unter CC BY 4.0.

Abb. 1: Deskriptorsatz *Wirtschaftsstruktur* aus dem Standard-Thesaurus Wirtschaft inkl. Links zu anderen Thesauri und Vokabularen

Abbildung 1 zeigt die Web-Darstellung eines sog. Deskriptorsatzes aus dem Standard-Thesaurus Wirtschaft, einem bilingualen wirtschaftswissenschaftlichen Fachthesaurus. Als Vorzugsbenennung der jeweiligen Sprache, hier dick gedruckt, wird die jeweils gebräuchlichste Benennung verwendet. Ein Klick auf das EB-Icon daneben, stößt eine Recherche nach sämtlichen damit erschlossenen Ressourcen im Fachportal EconBiz an. Die Nicht-Deskriptoren unter „benutzt für“ enthalten alternative Benennungen, deren Anzahl je nach Sprachgebrauch der jeweiligen Sprache variieren kann. In dieser Web-Ansicht, anders als für gewöhnlich in Print-Darstellungen, gibt es keine Nicht-Deskriptorsätze. Enthalten sind zwei Erläuterungen, auch Scope Notes genannt, die als kontextabhängige Erläuterung hier jeweils einen Verwendungshinweis auf einen anderen Deskriptor enthalten. Dem Deskriptor sind zwei Stellen in der Thesaurus-Systematik zugeordnet, die als weitere Kontextinformation verdeutlichen sollen, wo in der disziplinspezifischen Unterteilung in Fachteilgebiete der Deskriptor verortet ist. Auf der linken Seite findet sich die oberste Ebene dieser Systematik, entlang derer auch durch die einzelnen Teilgebiete des Thesaurus navigiert werden kann, etwa um sich einen Überblick über die im Thesaurus enthaltenen Themenbereiche zu verschaffen. Sinnvoll ist, sich hierbei an disziplinspezifischen Sachgebietsunterteilungen zu orientieren. Weiter unten sind Links zu anderen Thesauri und Vokabularen aufgeführt (s. Abschnitt 3). In der Web-Präsenz finden sich darüber hinaus weitere Menüpunkte, die die Verwendung erleichtern und ggf. auch die Nachnutzung des Thesaurus ermöglichen sollen. Diese beinhalten eine alphabetische Ansicht über die im Thesaurus enthaltenen Deskriptoren, sog. Mappings zu anderen Vokabularen (s. Abschnitt 3), die Möglichkeit des Downloads, die Anzeige früherer Versionen sowie Hintergrundinformationen u. a. zur Geschichte des Thesaurus. Ferner gibt es die Möglichkeit, mit dem Team bzw. der Thesaurus-Redaktion, etwa für Begriffsvorschläge, in Kontakt zu treten. Am Ende des Deskriptorsatzes findet sich eine Angabe zur Lizenz, unter der der Thesaurus nachgenutzt werden kann.

Neben den Strukturmerkmalen, die anhand des Beispiels veranschaulicht wurden, können weitere Bestandteile Eingang in die Darstellung von Thesauri finden (s. Tabelle 1). Denkbar ist, dass zusätzlich in Form einer Definition die kontextunabhängige lexikalische Bedeutung des Deskriptors angegeben ist, oder dass in einer sog. *History Note* Details zur Anzeige gebracht werden, wann ein Deskriptor aufgenommen oder zuletzt geändert wurde.

### 3 Thesaurusaufbau und -pflege unter Berücksichtigung informationsethischer Aspekte

Vor der eigentlichen Terminologiearbeit zur Erstellung eines Thesaurus sind zahlreiche grundsätzliche Fragen zu klären, die den Bezugsrahmen des Thesaurus hinsichtlich Gegenstandsbereich und Umfang sowie Nutzerschaft, zu erschließender Ressourcen und Erschließungstiefe betreffen. Daran entscheidet sich etwa, ob es sich um einen mehrsprachigen Thesaurus handeln und welchen Grad an Spezifität die Thesaurusbegriffe aufweisen sollten (Burkart 2004, S. 141). Hieran schließen sich die Auswahl geeigneter Quellen, wie Fachwörterbücher, Lehrbücher sowie relevante Teilausschnitte aus ggf. bereits existierenden sog. Makrothesauri, die thematisch meist relativ breit fachlich allerdings häufig eher allgemein gehalten aufgestellt sind, und die Sammlung von Wortmaterial daraus an. Auf Basis geeigneter Textkorpora lassen sich zunehmend auch Daten ge-

triebene Verfahren, etwa aus dem Bereich des Text Mining (s. Kapitel. B 16 Text and Data Mining), für die Generierung von Kandidatenvokabular verwenden.

An dieser Stelle ist es lohnend, sich die Standortgebundenheit des dieser Art zusammengetragenen Wortguts und sich in einem nächsten Schritt die darin enthaltenen sozialen Implikationen dieser Wissensorganisationsmethode bewusst zu machen. Unter Berücksichtigung der zeitlichen Perspektive ist das Thesaurusvokabular einerseits in die Zukunft gerichtet, da es darum geht, zukünftige Suchanfragen zu antizipieren. Andererseits ist das in dieser Form akquirierte Vokabular sehr stark in zurückliegenden Diskursen, in denen es sich sukzessive etabliert hat, verankert (Buckland 2017, S. 96–98). Gerade in wissenschaftlichen Publikationsprozessen ist Sprache, die jeweils in ein spezifisches Diskursfeld eingebettet ist, stets eine wichtige Ressource im Kampf um wissenschaftliche Anerkennung (Barlösius 2012, S. 130). Legt man einen globalen Maßstab an, wird deutlich, wie sehr die Wissenschaftssysteme zahlreicher Länder und Regionen strukturell benachteiligt sind (Schmidt 2020). Für die Orientierung bei der Terminologiearbeit an der – bezogen auf den Gegenstandsbereich des kontrollierten Vokabulars – allgemeinen Gebräuchlichkeit wurde in der Informationswissenschaft das Konzept des sog. *literary warrant* (Barité 2018) geprägt, wonach ein Begriff erst dann eingeführt werden sollte, wenn er in der Literatur häufiger vorkommt bzw. in einschlägigen Referenzwerken verzeichnet ist. Damit wird auf der einen Seite zwar der Nutzerorientierung entsprochen. Unter Berücksichtigung informationsethischer Gesichtspunkte berge dies, laut Rösch (2021, S. 319), allerdings auch die Gefahr, dass dadurch „bestehende Stereotype, Bewertungen und Herrschaftsstrukturen terminologisch reproduziert“ und damit Perspektiven einer Minderheit unterrepräsentiert würden. Tendenziell reproduzieren Thesauri somit eine Mehrheitsmeinung. Da Sprache stets in einen spezifischen zeitlich-kulturellen Kontext eingebettet ist, ist sie immer in gewisser Weise weltanschaulich befangen und verzerrt. Dies gälte es, so Rösch (2021, S. 317), sich bewusst zu machen, nach Kräften zu entschärfen und – wenn möglich – mit dem Ziel einer größeren Ausgewogenheit zu beseitigen. Neben der Termfindung und der sich daran anschließenden terminologischen Kontrolle, bei der das Prinzip der Gleichsetzung von Bezeichnungen bzw. der dahinterliegenden Sachverhalte zum Tragen kommt, birgt auch die begriffliche Kontrolle soziale und ethische Implikationen. So legen Thesauri nicht nur in Form der Äquivalenzrelation, sondern auch durch die hierarchischen Begriffsbeziehungen ein Ordnungssystem an die Welt an, das an den Bedarfen und Relevanzen des Gegenstandsbereichs, der durch den Thesaurus abgedeckt wird, ausgerichtet ist und gewisse Normvorstellungen transportieren kann. Bestimmte Perspektiven, so Olson (1997, S. 183, zitiert nach Rösch 2021, S. 319) unter Verweis auf die *Library of Congress Subject Headings*,<sup>1</sup> seien unterrepräsentiert und was den Normvorstellungen, die darin eingeschlossen seien, nicht entspreche, gelte als Abweichung. Am Beispiel der Gemeinsamen Normdatei (GND),<sup>2</sup> deren Sacherschließungsvokabular sukzessive zu einem Thesaurus ausgebaut werden soll, wurde dies in der Vergangenheit exemplarisch an der Ansetzung und Verortung des Deskriptors *Transsexualismus* deutlich, der lange als Unterbegriff zum Deskriptor *Geschlechtsidentitätsstörung* aufgeführt wurde. Entsprechend dem Nachschlagewerksprinzip der GND bezog man sich als Quelle unter anderem auf die zehnte Ausgabe der *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). Als *Transsexualismus* in der darauffolgenden Ausgabe nicht mehr länger als *Geschlechtsidentitätsstörung* geführt wurde, wurde der Deskriptor in *Transsexualität* umbenannt

1 <https://id.loc.gov/authorities/subjects.html>.

2 [https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd\\_node.html](https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html).

und als Oberbegriff der Deskriptor *Geschlechtsidentität* angegeben. Die ICD wird nicht mehr als Quelle genannt.<sup>3</sup>

Dieses Beispiel veranschaulicht, dass sprachliche Bezeichnungen nicht stabil sind. Der Wandel in der Sprache, der je nach Perspektive auf einen Gegenstandsbereich bzw. eine Fachdisziplin asynchron verlaufen kann, erfordert Aktualisierungen in der Terminologie. Gleichzeitig gilt es, zurückliegende dominante Ausdrucksformen, etwa für wissenschaftsgeschichtliche Fragestellungen, auffindbar zu halten. Der permanente Sprachwandel macht es erforderlich, dass ein Thesaurus kontinuierlich überprüft und weiterentwickelt wird. In der Regel ist dafür eine Redaktion eingesetzt, die bei Änderungen und Weiterentwicklungen die Bedarfe der Nutzenden sowie die Entwicklungen in der entsprechenden Wissensdomäne im Blick hat. Änderungsbedarfe können etwa die Zusammenlegung von bisher getrennt aufgeführten Deskriptoren oder aber die Aufspaltung eines bisher relativ breit angelegten Deskriptors in zwei oder mehr Deskriptoren betreffen. Ebenfalls denkbar ist die Einführung einer neuen Systematikstelle, etwa aufgrund zahlreicher Innovationen oder neuer Forschungsfelder in einem Fachteilgebiet. Im Zuge der zunehmenden (Teil-)Automatisierung der Sacherschließung durch prozessunterstützende Assistenzsysteme, wie den Digitalen Assistenten<sup>4</sup> (siehe auch Beckmann et al. 2019), gilt es, neben Begriffsvorschlägen von Inhaltserschließenden weitere Quellen für Kandidatenvokabular zu nutzen, um die Rückbindung an die Entwicklung des Fachwissens sicherzustellen, die für die Begriffsfindung und terminologische Arbeit unabdingbar ist. Denkbar ist etwa das maschinelle Clustering von Stichwörtern, die von Autorinnen und Autoren vergeben wurden, wodurch die terminologische Weiterentwicklung zusätzlich potentiell stärker an die Perspektive von Forschenden heranrücken könnte, oder die technisch unterstützte Analyse von Recherche-Logfiles der Nutzerinnen und Nutzer eines Informationssystems.

Das Thesaurusmanagement lässt sich durch den Einsatz von Software vereinfachen, die etwa das reziproke Anlegen von Relationen und Konsistenzchecks erlaubt (siehe ISO 2011, Kapitel 14 zu Anforderungen an Thesaurusmanagement-Software). Je nach Gesamtarchitektur des Informationssystems lässt sich zwischen anwendungsabhängiger bzw. integrierter Software, bei der gleichzeitig der Dokumentbestand mit gepflegt wird, und anwendungsunabhängiger bzw. autonomer Thesaurusmanagement-Software unterscheiden (Bertram 2005, S. 230–231). Neben kommerzieller Software stehen zunehmend auch Open-Source-Anwendungen zur Verfügung.

## 4 Öffnung des Bezugsrahmens von Thesauri

Den primären Bezugsrahmen eines Thesaurus stellte lange Zeit ein abgeschlossenes Informationssystem dar. Mittlerweile setzen sich Suchräume häufig aus verteilten Datenbanken zusammen und die inhaltsbeschreibenden Metadaten (s. Kapitel B 9 Metadaten) der in ihnen enthaltenen Informationsressourcen weisen oftmals eine große Heterogenität auf. Um etwa in Fachportalen eine Datenbank übergreifende Suche zu ermöglichen, gilt es, über Querverbindungen bzw. sog. Crosskonkordanzen oder Mapping-Relationen im Sinne von Synonym-, Ähnlichkeits- oder Hierarchie-Beziehungen die Begriffe der unterschiedlichen kontrollierten Vokabulare aufeinander abzubilden. In den Suchindex

<sup>3</sup> <https://d-nb.info/gnd/4185937-6>.

<sup>4</sup> <https://www.eurospider.com/de/relevancy-produkt/digitaler-assistent-da-3>.

aufgenommene Äquivalenzrelationen können so als zusätzliche Sucheinstiege genutzt werden (Kempf & Neubert 2016, S. 166–167). Im Deskriptorsatz als Links zur Anzeige gebracht (siehe Abbildung 1) können sie darüber hinaus als Brücke in weitere Rechercheportale dienen. Insbesondere der zweite Teil des ISO-Standards 25964 liefert Orientierung beim Aufbau derartiger Mapping-Relationen. Es wird zwischen verschiedenen Formen von Äquivalenz und unterschiedlichen Mapping-Modellen unterschieden. So können entweder uni- oder bidirektionale Direktverbindungen zwischen allen beteiligten Wissensorganisationssystemen aufgebaut oder eines der berücksichtigten Systeme als sog. *Hub* bzw. zentraler Knotenpunkt genutzt werden. Die beteiligten Wissensorganisationssysteme sind dann lediglich mit diesem direkt verbunden. Ein Beispiel für eine derartige Datendrehscheibe, mit deren Entitäten mittlerweile zahlreiche Thesauri verlinkt sind, ist die sog. Wissensdatenbank Wikidata.<sup>5</sup> Zu aktuellen Initiativen, die das Ziel verfolgen, die Infrastruktur zu Aufbau, Pflege und Nachnutzung von Mappings wirksam zu verbessern, zählen etwa *coli-conc*<sup>6</sup> und das Projekt *GND-mul*.<sup>7</sup>

Ein weiterer wesentlicher Schritt, wodurch sich der Bezugsrahmen von Thesauri deutlich erweitert hat, bildete die Verabschiedung des sog. SKOS-Standards – das Akronym steht für *Simple Knowledge Organization System* – zur Veröffentlichung von Thesauri und anderen kontrollierten Vokabularen im Web durch das *World Wide Web Consortium* (2009). Hierbei handelt sich um ein RDF-basiertes, maschinenlesbares Datenformat, das für das Semantic Web entwickelt wurde und auf das mittlerweile zahlreiche Linked-Data-Anwendungen zugreifen. Die Deskriptoren bzw. die in den Deskriptorsätzen enthaltenen semantischen Informationen – bei der Anzeige im Web werden sie über RDFa (*Resource Description Framework in attributes*) in die HTML-Webseiten eingebettet – werden durch persistente, im World Wide Web eindeutig referenzierbare Identifier (*Uniform Resource Identifier* – URI) repräsentiert. Ebenso wird es möglich, den Thesaurus in einem Download-Format anzubieten, das eine maschinelle Nachnutzung zulässt. Des Weiteren erlaubt es der Web-Standard für den Fall, dass Änderungen im Thesaurus in aufeinanderfolgenden eigenen Versionen veröffentlicht werden und jede publizierte URI persistent bleibt, diese Änderungen entlang der Versionen transparent zu machen. Indem die HTML/RDFa-Seiten älterer Versionen weiter online verfügbar gehalten werden, können durch sog. Linked-Data-basierte Änderungsberichte sämtliche Änderungen, die von einer auf die andere Version durchgeführt wurden, transparent gemacht und Unterschiede zwischen den Versionen maschinell ausgewertet, analysiert und dargestellt werden (Gastmeyer et al. 2016, S. 236).<sup>8</sup>

Darüber hinaus lassen sich mit Veröffentlichung des SKOS-Standards auch obige Mapping-Relationen zwischen Wissensorganisationssystemen im Web veröffentlichen. Fünf unterschiedliche Typen von Mapping-Relationen stehen zur Verfügung, um im Fall von Thesauri Deskriptoren aus unterschiedlichen Thesauri aufeinander abzubilden und dadurch semantische Interoperabilität herzustellen, die je nach Überlappung des Bedeutungsumfangs weiter ausdifferenziert werden kann. Darüber hinaus lassen sich je nach Eigenschaft der Relationen, die maschinell interpretiert werden können, automatisch Inferenzen ziehen (Hubrich 2018, S. 265). Die beiden sog. intersystemischen Relationstypen *skos:exactMatch* und *skos:closeMatch*, die von Bedeutungsgleichheit bzw. -ähnlichkeit

<sup>5</sup> [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page).

<sup>6</sup> <https://coli-conc.gbv.de/>.

<sup>7</sup> [https://gnd.network/Webs/gnd/DE/Projekte/GNDmul/GNDmul\\_node.html](https://gnd.network/Webs/gnd/DE/Projekte/GNDmul/GNDmul_node.html).

<sup>8</sup> Zum Vorgehen bei der maschinellen Auswertung unterschiedlicher Versionen eines Wissensorganisationssystems im SKOS-Format siehe auch <https://github.com/jneubert/skos-history>.

ausgehen, verfügen über die logische Eigenschaft der Symmetrie. *Skos:exactMatch* verfügt darüber hinaus über die logische Eigenschaft der Transitivität. Wenn Deskriptor  $D^A$  aus dem Thesaurus A über *skos:exactMatch* mit einem in dem Thesaurus B modellierten Deskriptor  $D^B$  verbunden ist und letzterer auch über *skos:exactMatch* mit einem in Thesaurus C modellierten Deskriptor  $D^C$  verlinkt ist, kann aufgrund der Transitivität der Relation der Rückschluss gezogen werden, dass auch Deskriptor  $D^A$  über *skos:exactMatch* mit Deskriptor  $D^C$  verbunden ist. Transitive Relationen, die sich insbesondere aus Mappings ziehen lassen, die dem obigen Mapping-Modell des *Hub* entsprechen, können für automatische Sucherweiterungen genutzt werden (Hubrich 2018, S. 266). Zusätzlich lassen sich Mapping-Relationen zwischen Deskriptoren aus unterschiedlichen Thesauri erstellen, die in einer hierarchischen oder assoziativen Beziehung zueinander stehen. Hierarchische intersystemische Beziehungen werden durch die als invers definierten Relationstypen *skos:narrowMatch* und *skos:broadMatch* zum Ausdruck gebracht, assoziative intersystemische Relationen durch den mit der logischen Eigenschaft der Symmetrie ausgezeichneten Relationstyp *skos:relatedMatch*.

Durch die Web-Veröffentlichung und die Anzeige von Links zu anderen Thesauri und Vokabularen wird der Thesaurus selbst zu einer Art Daten-Drehscheibe. Zum einen schlägt er innerhalb des Vokabulars eine Brücke zum Rechercheportal, in das er eingebunden ist. Zum anderen bündelt er Verknüpfungen zu anderen Datenbanken und Katalogen.

## 5 Ausblick

Verbunden mit den aufgezeigten veränderten Rahmenbedingungen und Anwendungskontexten ist die Diskussion um die Rolle von Thesauri in modernen Information Retrieval-Systemen längst eröffnet (Dextre Clarke 2016). Da Aufbau, Pflege und Weiterentwicklung von Thesauri mit einem hohen Ressourcenaufwand verbunden sind, wird es zunehmend darauf ankommen, sämtliche dieser Prozessschritte zusätzlich stärker datengetrieben auszurichten und auf (teil-)automatisierte Unterstützungsverfahren zu setzen, etwa bei der Generierung von Kandidatenvokabular oder bei dem Aufspüren von Bedeutungsveränderungen von Begriffen bzw. ihren sprachlichen Repräsentationen, etwa durch Auswertung ihrer textlichen Einbettung bzw. ihrer sog. *Word Embeddings*. Die Veröffentlichung von Thesauri in einem maschinenlesbaren Format und der Ausbau Web basierter maschinenlesbarer Vokabular-Mappings – ggf. zusätzlich angereichert mit semantischen Kontextinformationen – sowie der Aufbau nachhaltiger Mapping-Infrastrukturen bilden hierfür wichtige Ausgangs- und Rahmenbedingungen. Gleichzeitig kann der Auf- und Ausbau von Mapping-Relationen zu anderen Vokabularen wichtige Unterstützung dabei bieten, die Interoperabilität von Begriffsansetzungen zu schärfen, die mit zunehmender Vernetzung offen und anschlussfähig gestalteter Informationssysteme an Bedeutung gewinnt. Daneben können von der Ausweitung der Thesaurus-Methodik auf weitere Datentypen, etwa Forschungsdaten, und von den damit verbundenen Erschließungs- und Kuratierungsaufgaben wichtige Impulse für eine zukunftsorientierte Ausrichtung der Thesaurusmanagementprozesse ausgehen.<sup>9</sup>

---

<sup>9</sup> Als ein Projekt zur Thesaurus gestützten Erschließung von Forschungsdaten sei beispielhaft auf die Pilotstudie *Linked Open Research Data* (LORDpilot) verwiesen.

Schließlich wird es zukünftig weiterhin darum gehen, Thesauri optimal in moderne Information Retrieval-Anwendungen zu integrieren, um – zumeist quasi im Hintergrund – auf Basis von terminologischer und begrifflicher Kontrolle eine optimale Retrieval-Unterstützung, etwa bei Mehrsprachigkeit und Fremddatennutzung, zu erzielen. Der zweite Teil der ISO-Norm 25964 (ISO 2013) zur Interoperabilität von Thesauri mit anderen Vokabularen liefert hierzu wichtige Leitlinien. Denkbar ist die Einbindung der Thesaurus-Methodik in „hybride Wissensorganisationssysteme“ (Dextre Clarke 2019, S. 453), etwa indem fachbezogene Ontologien oder Wissensgraphen um thesauruspezifische Struktur- und Relationsbeziehungen ergänzt werden.

**Tab. 1:** Relationenkürzel und Erläuterungen zu Angaben in einem Deskriptorsatz (in Anlehnung an ISO 25964, 2011 und DIN 1463-1, 1987, DIN 1463-2, 1988)

<i>Deutsch</i>	<i>Englisch</i>	<i>Erläuterung</i>
<b>Hinweise und Definitionen</b>		
H Hinweis	SN Scope Note	Erläuterung
D Definition	D Definition	Definition
-	HN History Note	Hinweis zur Einführung von oder Änderung an einem Deskriptorsatz (in DIN 1463, 1987 ist keine Abkürzung enthalten)
<b>Codes</b>		
- Systematikstelle	SC Subject Category	Notation einer Systematik
C Begriffscode oder Notation	CC Concept Code or Notation	kodierte Bezeichnung aus anderen Systemen
<b>Verweisungen</b>		
BS Benutze Synonym / Quasisynonym	USE Use	Äquivalenz: ND – Deskriptor
BF Benutzt für Synonym / Quasisynonym	UF Used for	Äquivalenz: Deskriptor – ND
BK Benutze Kombination	USE...+ Use	partielle Äquivalenz: ND – mehrere Deskriptoren
KB Benutzt in Kombination	UF+ / UFC Used for combination	partielle Äquivalenz: Deskriptor – ND
<b>Relationen</b>		
TT Top Term	TT Top Term	oberster Begriff
OB Oberbegriff	BT Broader Term	übergeordneter Begriff
OA Oberbegriff	BTG Broader Term (generic)	Oberbegriff (Abstraktionsrelation)
–	BTI Broader Term (instantial)	Oberbegriff (Instanzrelation)
SP Verbandsbegriff	BTP Broader Term (partitive)	Verbandsbegriff (Bestandsrelation)
UB Unterbegriff	NT Narrower Term	untergeordneter Begriff
UA Unterbegriff	NTG Narrower Term (generic)	Unterbegriff (Abstraktionsrelation)
-	NTI Narrower Term (instantial)	Unterbegriff Instanzrelation (Individualbegriff)

<i>Deutsch</i>	<i>Englisch</i>	<i>Erläuterung</i>
TP Teilbegriff	NTP Narrower Term (partitive)	Teilbegriff
VB Verwandter Begriff	RT Related Term	Assoziationsrelation

## 6 Literaturverzeichnis

- Barité, M. (2018). Literary warrant. *Knowledge Organization*, 45(6), 517–536. <https://doi.org/10.5771/0943-7444-2018-6-517>.
- Barlösius, E. (2012). Wissenschaft als Feld. In S. Maasen, M. Kaiser, M. Reinhart & B. Sutter (Hrsg.), *Handbuch Wissenschaftssoziologie* (S. 125–135). Springer VS. [https://doi.org/10.1007/978-3-531-18918-5\\_10](https://doi.org/10.1007/978-3-531-18918-5_10).
- Beckmann, R., Hinrichs, I., Janßen, M., Milmeister, G. & Schäuble, P. (2019). Der Digitale Assistent DA-3: Eine Plattform für die Inhaltserschließung. *o-bib. Das offene Bibliotheksjournal*, 6(3), 1–20. <https://doi.org/10.5282/o-bib/2019H3S1-20>.
- Bertram, J. (2005). *Einführung in die inhaltliche Erschließung: Grundlagen – Methoden – Instrumente*. Ergon.
- Buckland, M. (2017). *Information and Society*. The MIT Press. <https://doi.org/10.7551/mitpress/10922.001.0001>.
- Burkart, M. (2004). Thesaurus. In R. Kuhlen, T. Seeger & D. Strauch (Hrsg.), *Grundlagen der praktischen Information und Dokumentation* (5. Aufl., S. 141–154). K. G. Saur. <https://doi.org/10.1515/9783110964110.141>.
- Deutsches Institut für Normierung (1987). *Erstellung und Weiterentwicklung von Thesauri: Einsprachige Thesauri* (DIN 1463-1:1987-11). Beuth.
- Deutsches Institut für Normierung (1988). *Erstellung und Weiterentwicklung von Thesauri: Mehrsprachige Thesauri* (DIN 1463-2:1988-12). Beuth.
- Dextre Clarke, S. G. (2016). Origins and Trajectory of the Long Thesaurus Debate. *Knowledge Organization*, 43(3), 138–144. <https://doi.org/10.5771/0943-7444-2016-3-138>.
- Dextre Clarke, S. G. (2019). The Information Retrieval Thesaurus. *Knowledge Organization*, 46(6), 439–459.
- Gastmeyer, M., Wannags, M.-M. & Neubert, J. (2016). Relaunch des Standard-Thesaurus Wirtschaft: Dynamik in der Wissensrepräsentation. *Information. Wissenschaft & Praxis*, 67(4), 217–240. <https://doi.org/10.1515/iwp-2016-0039>.
- Hubrich, J. (2018). Semantische Interoperabilität zwischen Klassifikationen und anderen Wissenssystemen. In H. Alex, G. Bee & U. Junger (Hrsg.) *Klassifikationen in Bibliotheken: Theorie – Anwendung – Nutzen* (S. 235–276). De Gruyter Saur. <https://doi.org/10.1515/9783110299250-008>.
- International Organization for Standardization (2011). *Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval* (ISO Standard No. 25964-1:2011). <https://www.iso.org/standard/53657.html>.
- International Organization for Standardization (2013). *Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies* (ISO Standard No. 25964-2:2013). <https://www.iso.org/standard/53658.html>.
- Kempf, A. O. & Neubert, J. (2016). The role of thesauri in an open web: A case study of the STW Thesaurus for Economics. *Knowledge Organization*, 43(3), 160–173. <https://doi.org/10.5771/0943-7444-2016-3-160>.
- Olson, H. A. (1997). The feminist and the emperor's new clothes: Feminist deconstruction as a critical methodology for library and information studies. *Library & Information Science Research*, 19(2), 181–198. [https://doi.org/10.1016/S0740-8188\(97\)90042-6](https://doi.org/10.1016/S0740-8188(97)90042-6).
- Rösch, H. (2021). *Informationsethik und Bibliotheksethik: Grundlagen und Praxis*. De Gruyter. <https://doi.org/10.1515/9783110522396>.

- Schmidt, N. (2020). *The Privilege to Select: Global Research System, European Academic Library Collections, and Decolonisation*. Lund University. <https://doi.org/10.5281/zenodo.4011295>.
- Stock, W. G. & Stock, M. (2008). *Wissensrepräsentation. Informationen auswerten und bereitstellen*. Oldenbourg.
- World Wide Web Consortium. (2009). *SKOS Simple Knowledge Organization System Reference. W3C Recommendation*. Retrieved November 29, 2021, from <http://w3.org/TR/skos-reference/>.