

Erstellung und Weiterentwicklung von Thesauri

Einsprachige Thesauri

DIN
1463
 Teil 1

Guidelines for the establishment and development of thesauri; Monolingual thesauri

Ersatz für DIN 1463/03.76

Zusammenhang mit der von der International Organization for Standardization (ISO) herausgegebenen Internationalen Norm ISO 2788 – 1986 siehe Erläuterungen.

Das Einrahmen von Textteilen dient der Hervorhebung.

Inhalt

	Seite		Seite
1 Anwendungsbereich und Zweck	2	5.5 Verweisungen	6
2 Begriffe	2	5.5.1 Verweisungen in der Äquivalenzrelation	7
3 Arten von Thesauri	2	5.5.2 Verweisungen in der Hierarchierelation	7
4 Auswahl und Bildung von Deskriptoren	2	5.5.3 Verweisungen in Begriffskombinationen	7
4.1 Arten von Deskriptoren und Nicht-Deskriptoren	2	6 Darstellungsform des Thesaurus	7
4.2 Formale Festlegungen	2	6.1 Einleitung zum Thesaurus	7
4.2.1 Zusammengesetzte Ausdrücke	2	6.2 Thesaurus-Hauptteil	7
4.2.2 Begriffsdarstellung durch mehrere Deskriptoren	3	6.3 Ergänzende Thesaurusteile	8
4.2.2.1 Prinzipien der Zerlegung von Bezeichnungen	3	6.4 Alphabetische Darstellung	8
4.2.2.2 Postkoordination und Präkoordination	3	6.5 Systematische Darstellung	8
4.2.3 Wortform	3	6.6 Beziehungsgraphen	8
4.2.3.1 Schreibweise	3	7 Erstellung eines Thesaurus	8
4.2.3.2 Fremdsprachige Bezeichnungen	3	7.1 Vermeidung von Doppelarbeit	8
4.2.3.3 Übersetzung und Transliteration	3	7.2 Bekanntgabe	9
4.2.3.4 Zeichenvorrat	4	7.3 Einbeziehung von Fachleuten	9
4.2.3.5 Wortarten	4	7.4 Verfahren	9
4.2.3.6 Numerus	4	7.5 Gewinnung von Deskriptoren	9
4.2.3.7 Abkürzungen und Akronyme	4	7.5.1 Sammlung	9
4.3 Verfahren zur Vermeidung von Mehrdeutigkeit	4	7.5.2 Test	9
4.3.1 Kennzeichnung von Homonymen bzw. Polysemen	4	7.5.3 Auswahl und Bewertung	10
4.3.2 Erläuterungen (Scope notes)	4	7.6 Erfassungsverfahren	10
4.3.3 Definitionen	4	7.7 Anwendung	10
4.3.4 Übersetzungen	4	8 Weiterentwicklung des Thesaurus	10
4.3.5 Quellenhinweise	5	8.1 Regelmäßige Überprüfung	10
5 Relationen zwischen Begriffen und Verweisungen zwischen Bezeichnungen	5	8.2 Eliminierung von Deskriptoren	10
5.1 Allgemeines	5	8.2.1 Eliminierung	10
5.2 Die Äquivalenzrelation	5	8.2.2 Differenzierung	10
5.3 Die Hierarchierelationen	5	8.3 Aufnahme neuer Deskriptoren	10
5.3.1 Die Abstraktionsrelation (generische Relation)	5	9 Kurzzeichen	10
5.3.2 Die Bestandsrelation (partitive Relation)	5	9.1 Kurzzeichen für Relationen und Verweisungen	10
5.3.3 Darstellung der Hierarchierelationen	6	9.2 Kurzzeichen für weitere Angaben	10
5.4 Die Assoziationsrelation	6	Zitierte Normen	12
		Weitere Normen	12

Fortsetzung Seite 2 bis 12

Normenausschuß Bibliotheks- und Dokumentationswesen (NABD) im DIN Deutsches Institut für Normung e.V.

1 Anwendungsbereich und Zweck

Diese Norm gilt im Bereich Information und Dokumentation (z. B. für Dokumentationsstellen, Bibliotheken, Archive, Hersteller von Datenbasen) für die Erstellung und Weiterentwicklung von einsprachigen Thesauri¹⁾, unabhängig von den bearbeiteten Fachgebieten. Sie soll dazu dienen, die möglichst einheitliche Erstellung und Weiterentwicklung von Thesauri zu erleichtern, unabhängig davon, ob sie manuell oder maschinell geführt werden; sie gilt nicht für Thesauri im Sinne der Sprachwissenschaft (z. B. Synonymwörterbücher u. ä.).

2 Begriffe

2.1 Dokumentationssprache

Eine Dokumentationssprache ist eine Menge sprachlicher Ausdrücke (Bezeichnungen²⁾), die, nach bestimmten Regeln angewendet, der Beschreibung von Dokumenten zum Zweck des Speicherns und einer gezielten Wiederauffindung (Retrieval) dient. Dokumentationssprachen können dargestellt werden durch Schlagwortsysteme, Thesauri und Klassifikationen.

2.2 Thesaurus

Ein Thesaurus im Bereich der Information und Dokumentation ist eine geordnete Zusammenstellung von Begriffen und ihren (vorwiegend natürlichsprachigen) Bezeichnungen, die in einem Dokumentationsgebiet zum Indexieren, Speichern und Wiederauffinden dient.

Er ist durch folgende Merkmale gekennzeichnet:

- a) Begriffe und Bezeichnungen werden eindeutig aufeinander bezogen („terminologische Kontrolle“), indem
 - Synonyme möglichst vollständig erfaßt werden,
 - Homonyme und Polyseme besonders gekennzeichnet werden,
 - für jeden Begriff eine Bezeichnung (Vorzugsbenennung, Begriffsnummer oder Notation) festgelegt wird, die den Begriff eindeutig vertritt,
- b) Beziehungen zwischen Begriffen (repräsentiert durch ihre Bezeichnungen) werden dargestellt.

2.3 Vokabular

Das Vokabular einer Dokumentationssprache ist die Menge der Deskriptoren einer Dokumentationssprache.

2.4 Deskriptor

Deskriptor ist eine Bezeichnung (Vorzugsbezeichnung), die zur Inhaltskennzeichnung (Indexierung) zugelassen ist.

2.5 Nicht-Deskriptor

Nicht-Deskriptor ist eine Bezeichnung, die nicht zur Indexierung zugelassen, aber im Thesaurus aufgeführt und entsprechend gekennzeichnet ist, um dem Benutzer den Zugang zu den Deskriptoren des Thesaurus zu erleichtern.

2.6 Arten von Relationen

2.6.1 Abstraktionsrelation

Definition siehe Abschnitt 5.3.1

2.6.2 Äquivalenzrelation

Definition siehe Abschnitt 5.2

2.6.3 Assoziationsrelation

Definition siehe Abschnitt 5.4

2.6.4 Bestandsrelation

Definition siehe Abschnitt 5.3.2

2.6.5 Hierarchierelation

Definition siehe Abschnitt 5.3

3 Arten von Thesauri

Hinsichtlich der terminologischen Kontrolle werden zwei Arten von Thesauri unterschieden:

- Thesauri, welche für die Indexierung und Suche für jeden Begriff nur **eine** Bezeichnung (Vorzugsbezeichnung) zulassen;
- Thesauri, welche alle im Thesaurus befindlichen Bezeichnungen für einen Begriff zur Indexierung zulassen unter der Voraussetzung, daß in einem der Indexierung nachfolgenden Schritt die Verknüpfung der gewählten Bezeichnung z. B. mit einer Begriffsnummer, einer Notation oder einer Vorzugsbenennung hergestellt wird. (Dieses Verfahren gestattet die Suche sowohl nach der Gesamtheit aller Bezeichnungen eines Begriffes (Äquivalenzklasse) als auch nach einer einzelnen Bezeichnung und bietet so u. a. die Möglichkeit, früher festgelegte Synonymie- und Begriffsrelationen im Thesaurus und im Speicher gezielt zu korrigieren.)

Während beim Thesaurus der ersten Art Pflege, Indexierung und Retrieval auch manuell durchgeführt werden können, ist beim Thesaurus der zweiten Art maschinelle Unterstützung erforderlich.

4 Auswahl und Bildung von Deskriptoren

4.1 Arten von Deskriptoren und Nicht-Deskriptoren

Deskriptoren können sein:

- a) Bezeichnungen von Begriffen
- b) Eigennamen, wie Projektnamen, geographische oder geopolitische Bezeichnungen (siehe DIN 3166), Produktnamen, Namen von Personen und Organisationen.

Nicht-Deskriptoren können u. a. sein:

- a) unterschiedliche Schreibweisen von Deskriptoren (siehe Abschnitt 4.2.3.1)
- b) Synonyme oder Quasi-Synonyme zu Deskriptoren (siehe Abschnitt 5.2)
- c) Begriffsbenennungen, für die eine Kombination von mehreren Deskriptoren verwendet werden soll (siehe Abschnitt 4.2.2).

4.2 Formale Festlegungen

4.2.1 Zusammengesetzte Ausdrücke

Ein Deskriptor kann aus einem Wort (Einwort-Deskriptor) oder mehreren Wörtern (Mehrwort-Deskriptor), z. B.: Wortgruppen oder adjektivischen Phrasen, bestehen. Im allgemeinen sollte ein Deskriptor die in der Fachliteratur übliche Terminologie widerspiegeln, und zwar unabhängig davon, wie viele einzelne Wörter im gegebenen Fall zum Darstellen des Begriffes notwendig sind. Es ist jedoch anzustreben, daß ein Deskriptor den gemeinten Begriff mit möglichst wenigen Wörtern, vorzugsweise mit einem Wort, bezeichnet. Hierbei ist allerdings zu beachten, daß bei Verkürzung eines Ausdrucks die entstehende Bezeichnung an Klarheit verlieren kann. Zusammengesetzte Deskriptoren werden in ihrer natürlichen Wortfolge aufgenommen (z. B. OPTISCHE AKTIVITÄT)³⁾. Es kann nützlich sein, die invertierten Formen (z. B. Aktivität, optische) als Nicht-Deskriptoren aufzunehmen und als Synonym zur nicht invertierten Form zu behandeln. Wenn

- 1) Eine weitere Norm der Reihe DIN 1463 „Erstellung und Weiterentwicklung von Thesauri; Mehrsprachige Thesauri“ ist in Vorbereitung.
- 2) Nach DIN 2330 wird im folgenden meist der Oberbegriff „Bezeichnung“ verwendet, der alle Formen von Begriffsrepräsentationen umfaßt.
- 3) Deskriptoren sind in dieser Norm in VERSALIEN, Nicht-Deskriptoren in Normalschrift wiedergegeben.

in Sonderfällen die invertierte Form als die verbindliche Deskriptorform gewählt wird, ist die Aufnahme der nicht invertierten Form als Synonym zur invertierten Form notwendig.

4.2.2 Begriffsdarstellung durch mehrere Deskriptoren

4.2.2.1 Prinzipien der Zerlegung von Bezeichnungen

Um die Anzahl der Deskriptoren überschaubar zu halten, kann es nützlich sein, Begriffe nicht durch die sie üblicherweise repräsentierenden Bezeichnungen, sondern durch Kombination bereits vorhandener Deskriptoren darzustellen. Hierbei gibt es grundsätzlich zwei Möglichkeiten:

- a) Die morphologische Zerlegung: Wenn der Begriff durch ein Kompositum oder eine Wortkombination repräsentiert wird, ist es oft möglich, das Kompositum oder die Wortkombination (Mehrwortdeskriptor) in ihre Bestandteile zu zerlegen. Die Kombination der dabei entstehenden Wörter repräsentiert den Ursprungsbegriff jedoch nur dann, wenn die morphologische Zerlegung mit der semantischen übereinstimmt. Nur in diesem Fall führt die morphologische Zerlegung zu einem sachlich richtigen Resultat.

Beispiel 1:

BLUMENGESTECK = BLUME + GESTECK

Beispiel 2, negatives Beispiel für a (≠ b):

EISENBAHN ≠ EISEN + BAHN

- b) Die semantische Zerlegung (semantic factoring): Hierbei wird ein Begriff in seine Bedeutungskomponenten zerlegt, die in ihrer Kombination den ursprünglichen Begriff ergeben müssen. Bei der semantischen Zerlegung sind die Deskriptoren, die den Begriff repräsentieren sollen, nicht notwendigerweise in der ursprünglichen Bezeichnung enthalten.

Eine mögliche semantisch korrekte Zerlegung wäre:

Beispiel 3:

EISENBAHN =
SCHIENENBAHN + ÜBERLANDVERKEHR

4.2.2.2 Postkoordination und Präkoordination

Werden Bezeichnungen, die Begriffskombinationen darstellen (z. B. Komposita, Mehrwortbenennungen), in dieser Form als Deskriptoren zugelassen (z. B. TIERPSYCHOLOGIE, z. B. HAKEN UND ÖSEN), so spricht man von „Präkoordination“. Stellt man dagegen eine Begriffskombination durch mehrere Deskriptoren dar, die erst während des Suchvorgangs miteinander kombiniert werden, um einen gesuchten Begriff darzustellen, handelt es sich um „Postkoordination“. Für die Entscheidung, ob präkombinierte Deskriptoren oder zu kombinierende Deskriptoren verwendet werden, können folgende Hinweise gegeben werden:

- a) Postkoordination

Postkoordination darf nur dann vorgenommen werden, wenn durch die Kombination der Deskriptoren tatsächlich der gleiche Begriff repräsentiert wird, wie durch einen präkombinierten Deskriptor.

Morphologische Zerlegung zum Zwecke der Postkoordination ist nur dann erlaubt, wenn diese mit der semantischen Zerlegung übereinstimmt.

Postkoordination sollte nicht vorgenommen werden, wenn

- eine hohe begriffliche Spezifizierung angestrebt wird;
- eine der Zerlegungskomponenten im hierarchischen Niveau zu weit vom ursprünglichen Begriff entfernt ist;
- Ballast aus ungewollten Kombinationen weitgehend vermieden werden soll.

- b) Präkoordination

Deskriptoren sollten immer in ihrer präkombinierten (präkoordinierten) Form verwendet werden, wenn

- die einzelnen Deskriptoren für sich eine andere Bedeutung haben als in der Kombination des präkombinierten Deskriptors;

Beispiel 4:

SCHLÜSSELBEIN

Beispiel 5:

ÖFFENTLICHKEITSARBEIT

- die einzelnen Deskriptoren in anderen hierarchischen Zusammenhängen stehen als der präkombinierte Deskriptor;

Beispiel 6:

HAUSFRAU

- der präkombinierte Deskriptor ein Eigenname ist;
- Zweifel bestehen, ob die Kombination der einzelnen Deskriptoren genau und ausschließlich den Begriffsinhalt des präkombinierten Deskriptors wiedergibt;
- wenn eine spezifische ballastarme Suche nach dem repräsentierten Begriff ermöglicht werden soll.

Wird ein Begriff durch die Kombination von einzelnen Deskriptoren dargestellt, ist dies durch die Verweisung „Benutze Kombination“ auszudrücken (siehe Abschnitt 5.5.3).

Beispiel 7:

Bibliotheksstatistik

Benutze Kombination STATISTIK + BIBLIOTHEK

Die nicht benutzte präkombinierte Bezeichnung muß als Nicht-Deskriptor in den alphabetischen Teil des Thesaurus aufgenommen werden, es kann darüber hinaus zweckmäßig sein, die Kombination von einzelnen Deskriptoren im systematischen Teil nachzuweisen.

4.2.3 Wortform

4.2.3.1 Schreibweise

Die am meisten verbreitete Schreibweise der Benennung ist zu benutzen. In den Fällen, in denen mehr als eine Schreibweise einer Benennung üblich ist, sollten diese wie Synonyme behandelt werden (siehe auch Abschnitt 5.2).

Beispiel 8:

PHOTOGRAPHIE
FOTOGRAFIE

Schreibvarianten können von Synonymen durch gesonderte Kennzeichnung abgehoben werden.

4.2.3.2 Fremdsprachige Bezeichnungen

Wenn fremdsprachige Ausdrücke in einer Fachterminologie benutzt werden, können diese in den Thesaurus aufgenommen werden, gegebenenfalls als Synonyme neben ihren deutschsprachigen Entsprechungen.

Beispiel 9:

RECHENANLAGE
COMPUTER

4.2.3.3 Übersetzung und Transliteration

Bei der Übernahme fremdsprachiger Bezeichnungen in einen Thesaurus kann in der Übersetzung oder Transliteration⁴⁾ eine Fehlerquelle liegen. Gegebenenfalls ist die Aufnahme von fremdsprachigen (z. B. englischsprachigen) Bezeichnungen in einen Thesaurus zu empfehlen.

⁴⁾ Bei Transliterationen sollten die Empfehlungen des DIN Deutsches Institut für Normung e.V. und der International Organization for Standardization (ISO) angewendet werden.

4.2.3.4 Zeichenvorrat

Bei Anwendung von Datenverarbeitungsanlagen kann der dort jeweils verarbeitbare Zeichenvorrat Beschränkungen mit sich bringen. In diesem Fall ist dafür zu sorgen, daß keine Bedeutungsveränderungen der betroffenen Thesaurusbezeichnungen (Deskriptoren, Nicht-Deskriptoren) entstehen. Dieses kann z. B. durch Verwendung verschlüsselter Zeichen (Prototypen) geschehen. Es ist empfehlenswert, den gleichen Zeichenvorrat zu benutzen wie bei der bibliographischen Beschreibung von Dokumenten.

4.2.3.5 Wortarten

Deskriptoren sind vorzugsweise in substantivischer Form zu bilden.

Beispiel 10:

THERAPIE statt THERAPIEREN

Falls notwendig, können Deskriptoren in nichtsubstantivischer Form gebildet werden. Wenn eine Tätigkeit gemeint ist, soll möglichst der Infinitiv des betreffenden Verbs benutzt werden, insbesondere wenn das Substantiv mehrdeutig ist.

Beispiel 11:

IMPRÄGNIEREN statt IMPRÄGNIERUNG

Adjektive sollen möglichst nicht isoliert, sondern in Verbindung mit einem Substantiv angegeben werden.

Beispiel 12:

INTERNATIONALE BEZIEHUNGEN
statt INTERNATIONAL

4.2.3.6 Numerus

Im allgemeinen sollen Deskriptoren im Nominativ Singular angegeben werden. Die Pluralform ist anzuwenden, wenn die Singularform des Wortes nicht gebräuchlich ist oder wenn der Plural eine andere Bedeutung hat als der Singular.

Beispiele 13 und 14 für Beibehaltung des Plurals:

KOSTEN
ELTERN

Beispiel für unterschiedliche Bedeutung der Plural-Singularform eines Wortes:

Beispiel 15:

GESCHICHTE – GESCHICHTEN

(Geschichte im Sinne von Historie, Geschichten im Sinne von Erzählungen.)

4.2.3.7 Abkürzungen und Akronyme

Abkürzungen und Akronyme sollen in einem Thesaurus als Synonyme zur nicht abgekürzten Bezeichnung aufgenommen werden, wenn ihre Bedeutung im betreffenden Benutzerkreis entweder als bekannt vorausgesetzt werden kann oder wenn sie in den allgemein-fachlichen Sprachgebrauch eingegangen sind.

Beispiel 16:

UKW

Beispiel 17:

EDV

Abkürzungen, die mehrere Bedeutungen haben, sind wie Homonyme zu behandeln (siehe Abschnitt 4.3.1). Allgemein bekannte Akronyme können als Deskriptoren verwendet werden.

Beispiel 18:

LASER

Beispiel 19:

RADAR

4.3 Verfahren zur Vermeidung von Mehrdeutigkeit**4.3.1 Kennzeichnung von Homonymen bzw. Polysemen**

Die unterschiedlichen Bedeutungen von Homonymen (Homographen) und Polysemen müssen durch näher bestimmende Zusätze kenntlich gemacht werden, die in Klammern unmittelbar hinter das Homonym oder Polysem gesetzt werden. Sie können aus einem spezifizierenden Wort, das selbst nicht homonym oder polysem ist, oder anderen geeigneten Kennzeichen bestehen. Homonym oder Polysem und der jeweils folgende Klammersausdruck bilden zusammen einen Deskriptor. Soweit möglich, sollten statt der Homonyme oder Polyseme eindeutige Bezeichnungen als Deskriptoren (Vorzugsbezeichnungen) verwendet werden.

Beispiel 20 für die Kennzeichnung von Homonymen:

KIEFER (KNOCHEN)
KIEFER (NADELHOLZ)

Beispiel 21 für die Kennzeichnung von Polysemen:

KNIE (KÖRPERTEIL)
KNIE (FORMSTÜCK)

Beispiel 22 für die Verwendung eindeutiger Bezeichnungen anstelle mehrdeutiger:

Eiweiß	Benutze Synonym	EIKLAR
Eiweiß	Benutze Synonym	PROTEIN

4.3.2 Erläuterungen (Scope notes)

Eine Erläuterung in diesem Sinne ist eine kurze Erklärung des beabsichtigten Gebrauchs eines Deskriptors. Sie kann verwendet werden um

- den Anwendungsbereich eines Deskriptors einzuschränken,
- Abkürzungen und Akronyme zu erläutern,
- eine andere Bedeutung einer Bezeichnung auszuschließen, insbesondere in den Fällen, in denen Bezeichnungen unterschiedlich in mehreren Fachgebieten verwendet werden,
- die Hinzufügung oder Eliminierung von Bezeichnungen zu datieren oder Änderungen im Gebrauch der Bezeichnungen festzuhalten.

Erläuterungen können dem Deskriptor im Thesaurus-Hauptteil beigegeben werden, bilden aber keinen Bestandteil des Deskriptors. Sie werden durch das Kurzzeichen H (Hinweis) gekennzeichnet.

Beispiel 23:

KUPFERLEGIERUNG
H Legierung, mit Kupfer als Hauptbestandteil

Beispiel 24:

SÄUGETIER
H nur auf dem Land lebendes

4.3.3 Definitionen

Der Begriffsinhalt eines Deskriptors in einem Thesaurus wird hauptsächlich durch die dargestellten Relationen zu anderen Thesauruswörtern angegeben. Wenn Zweifel an den einheitlichen Interpretationen eines Deskriptors bestehen, soll eine Definition, d.h. genaue Bestimmung des Begriffsinhaltes, beigelegt werden.

Beispiel 25:

DESKRIPTOR
D Ein Deskriptor ist eine Bezeichnung (Vorzugsbezeichnung), die zur Inhaltskennzeichnung (Indexierung) zugelassen ist.

4.3.4 Übersetzungen

In vielen Fällen wird es sinnvoll sein, die äquivalenten Termini anderer Sprachen aufzuführen, um die korrekte Verwendung des Deskriptors bei der Analyse fremdsprachiger Texte sicherzustellen. Sollte keine vollständige Bedeutungs-gleichheit vorliegen, kann darauf in Form einer Erläuterung