

Indexierung zur inhaltlichen Erschließung von Dokumenten

Gleichordnende Indexierung mit Deskriptoren

DIN
31 623
Teil 2

Indexing for describing the contents of documents;
coordinate indexing with descriptors

Ersatz für Ausgabe 06.83

Inhalt

	Seite		Seite
1 Anwendungsbereich	1	3.3.6 Eponyme	8
2 Begriffe	1	3.3.7 Abkürzungen und Akronyme	8
3 Regeln zur gleichordnenden Indexierung mit Deskriptoren	1	4 Kriterien zur Begriffsabgrenzung durch Deskriptoren	8
3.1 Allgemeines zum Indexieren	2	4.1 Begriffseinheit, Begriffszerlegung, Begriffs-	
3.1.1 Verstehen des Dokumenteninhalts	2	zusammenführung	8
3.1.2 Ermitteln der Begriffe	2	4.2 Morphologische und semantische Zerlegung	10
3.1.3 Aufsuchen oder Formulieren der Deskriptoren ..	3	4.3 Lexembegriff zur sprachlichen Benennungsanalyse	10
3.2 Kriterien zur Deskriptoren-Auswahl	3	4.4 Regelwerk zur Begriffs- und Benennungsanalyse	11
3.2.1 Indexierungsbreite	3	4.4.1 Überprüfung der Begriffsbenennung	11
3.2.2 Indexierungsspezifität	3	4.4.2 Zerlegungskriterien	11
3.2.3 Auswahl freier Deskriptoren	4	4.4.2.1 Lexem-Anzahl	11
3.2.3.1 Wiedergabetreue der Deskriptoren	4	4.4.2.1.1 Grundschemata des Vorgehens bei stärker	
3.2.3.2 Eindeutigkeit der Deskriptoren	4	postkoordinierenden Systemen	11
3.2.3.3 Verlässlichkeit der Deskriptoren	4	4.4.2.1.2 Grundschemata des Vorgehens bei stärker	
3.2.3.4 Prägnanz der Deskriptoren	4	präkoordinierenden Systemen	11
3.2.3.5 Gebräuchlichkeit der Deskriptoren	4	4.4.2.2 Sprachform der Deskriptoren	11
3.2.4 Kategorieller Gesichtspunkt der Deskriptoren-		4.4.2.3 Kategoriale Sonderregelungen: siehe	
Auswahl	5	Abschnitt 4.4.4	12
3.2.4.1 Fundamentalkategorie	5	4.4.3 Begriffsanalyse und Regeln zur Begriffszerlegung	12
3.2.4.2 Indexierung in den zutreffenden Kategorien ..	5	4.4.3.1 Methode der Begriffsanalyse	12
3.2.5 Behandlung quantitativer und gegensätzlicher		4.4.3.2 Regeln zur Begriffszerlegung	12
Eigenschaftsbegriffe	5	4.4.3.2.1 Grundbedingungen der Zerlegung	13
3.2.5.1 Quantitative Angaben von Eigenschaften	5	4.4.3.2.2 Hierarchisches Niveau der Zerlegungs-	
3.2.5.2 Gegensätze in Eigenschaftsbegriffen	6	komponenten	13
3.2.6 Angabe von implizit enthaltenen Inhaltsbegriffen	6	4.4.3.2.3 Möglichkeit der ausschließlichen Zerlegung	
3.3 Kriterien zur sprachlichen Formulierung der		in Oberbegriffe	13
Deskriptoren	7	4.4.3.2.4 Adjektive	13
3.3.1 Sprachlicher Aufbau von Deskriptoren	7	4.4.3.3 Beispiele zur Begriffsanalyse und -zerlegung,	
3.3.2 Wortfolge	7	vorzugsweise für stärker präkoordinierendes	
3.3.3 Fremdsprachige Deskriptoren	7	Vorgehen	13
3.3.4 Numerus	7	4.4.4 Sonderregeln für bestimmte Fundamental-	
3.3.5 Wortarten	8	kategorien	14
		4.5 Begriffszusammenführung	15

1 Anwendungsbereich

Diese Norm gilt im Arbeitsbereich Information und Dokumentation. Sie behandelt in erster Linie die Probleme der Literaturdokumentation. Im Zusammenhang mit anderen Dokumentationsarten (z.B. Datendokumentation, Objektdokumentation, Projektdokumentation) kann sie immer dann sinngemäß angewendet werden, wenn hier Informationen verbal erschlossen werden sollen.

Der vorliegende Teil 2 der Norm betrifft die intellektuelle gleichordnende Indexierung mit Deskriptoren (im weiteren Text kurz „Indexierung“).

2 Begriffe

Zur Definition des Begriffs „gleichordnende Indexierung“ und weiterer grundlegender Begriffe der Indexierung siehe DIN 31 623 Teil 1.

3 Regeln zur gleichordnenden Indexierung mit Deskriptoren

In den nachfolgenden Hinweisen und Regeln sind die Extraktions- und die Additionsmethode (siehe DIN 31 623 Teil 1/09.88, Abschnitt 4.5) sowie die Zuteilung freier Deskriptoren und die Zuteilung verbindlicher Deskriptoren (siehe DIN 31 623 Teil 1/09.88, Abschnitt 4.7) und deren Mischformen berücksichtigt.

Die Hinweise und Regeln gelten als Rahmennormen und sind gegebenenfalls durch systemspezifische Regelungen zu ergänzen (siehe DIN 31 623 Teil 1/09.88, Abschnitt 6).

Zur Inhaltserschließung eines Dokuments können neben einem konkreten Verfahren der intellektuellen, gleichordnenden Indexierung mit Deskriptoren, wenn es zweckmäßig erscheint, zusätzlich weitere Indexierungsmethoden angewendet werden.

Fortsetzung Seite 2 bis 16

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

3.1 Allgemeines zum Indexieren

Das Indexieren umfaßt mehrere Schritte:

- Das **Verstehen** des Dokumenteninhalts
- Das **Ermitteln** der Begriffe, die den wesentlichen Inhalt des vorliegenden Dokuments wiedergeben
- Je nach Indexierungsmethode
 - das **Aufsuchen** der entsprechenden bestpassenden Deskriptoren in einem vorgegebenen Vokabular bzw.
 - das **Formulieren** der entsprechenden bestpassenden Deskriptoren, wenn ein Vokabular noch nicht vorhanden ist (Ausnahmefall) oder nicht verbindlich vorgeschrieben ist und die besttreffenden Deskriptoren (noch) nicht enthält.

3.1.1 Verstehen des Dokumenteninhalts

Das Indexieren beginnt mit dem Lesen und Verstehen des Dokumenteninhalts.

Bei umfangreichen Dokumenten können Titel, Zusammenfassung, Kapitelüberschriften, Bilder und durch Auszeichnungsschrift hervorgehobene Textstellen für das Verständnis des Dokumenteninhalts von besonderem Nutzen sein. In jedem Fall muß sichergestellt werden, daß keine wesentlichen Informationen übersehen werden. Wenn statt des Originaldokuments nur ein Referat zur Verfügung steht, kann ein entsprechender Informationsverlust unvermeidbar sein.

3.1.2 Ermitteln der Begriffe

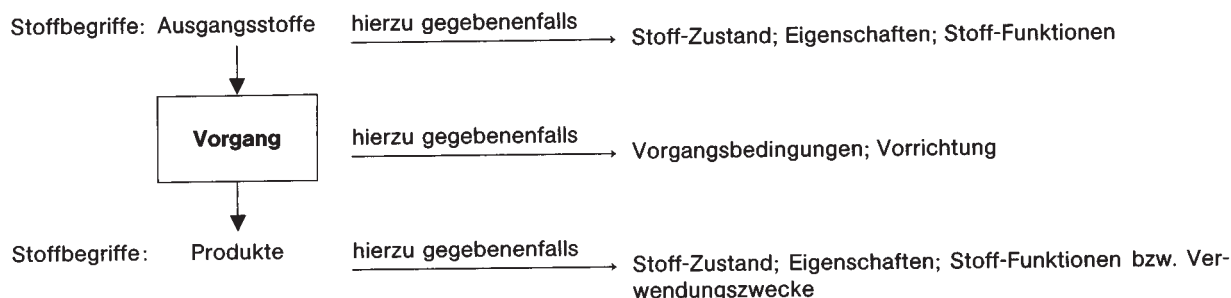
Das Ermitteln von wesentlichen Inhaltsbegriffen kann durch Indexierungshilfen unterstützt werden, z.B. durch ein Schema der Fundamentalkategorien (siehe Abschnitt 3.2.4.1), die für die inhaltliche Beschreibung von Dokumenten eines bestimmten Fachgebietes charakteristisch sind.

Beispiel 1:

Im Fachgebiet **Pharmazie** kann der Indexierer prüfen, ob folgende Angaben für eine Inhaltswiedergabe in Frage kommen: Arzneimittelname, Arzneimittelart, Applikationsform, behandelte Krankheit, Nebenwirkungen usw.

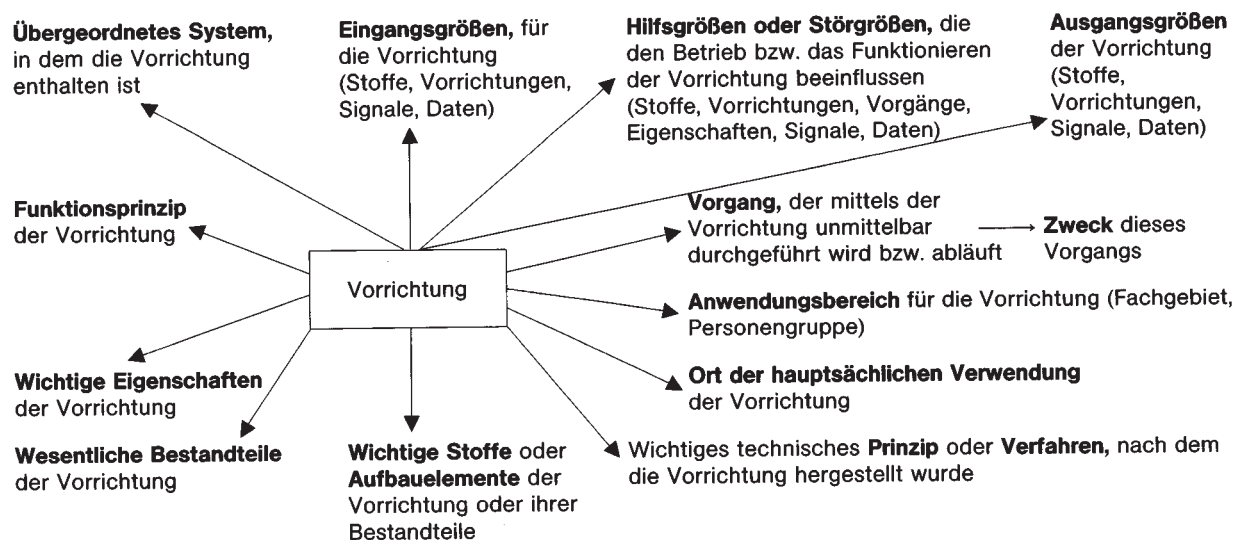
Beispiel 2:

Bei der Indexierung eines **chemischen Vorgangs** ist auf Angaben in folgenden Fundamentalkategorien zu achten:



Beispiel 3:

Stehen in einem technischen Dokumentationsgebiet **Vorrichtungen** im Mittelpunkt des Interesses, so können bei der Inhaltsschließung von Dokumenten Angaben über folgende Fundamentalkategorien und Systemgrößen wichtig sein:



Im übrigen gelten die Überlegungen zur Auswahl von Deskriptoren in den folgenden Abschnitten 3.2, 3.2.1 und 3.2.2 auch schon für die Ermittlung der zugrundeliegenden Begriffe.

3.1.3 Aufsuchen oder Formulieren der Deskriptoren

Das Vokabular der angewendeten Dokumentationssprache sollte in jedem Fall als Indexierungshilfsmittel (Nachschlagewerk) angewendet werden, auch soweit enthaltene Deskriptoren nicht verbindlich vorgeschrieben sind. Dürfen **ausschließlich** Deskriptoren eines verbindlich vorgegebenen Vokabulars verwendet werden, so ist immer dann mit Verlust an semantischer Information zu rechnen, wenn die zur Kennzeichnung des Dokumenteninhalts besttreffenden Deskriptoren im Vokabular nicht enthalten sind (siehe DIN 31 623 Teil 1/09.88, Abschnitt 4.7.2).

Soweit das Aufsuchen von Deskriptoren in einem vorgegebenen Vokabular ein „mechanischer“ Vorgang ist, bedarf dieser keines besonderen Regelwerks. Weitergehende Überlegungen beim Aufsuchen sind denen analog, die auch mit dem Formulieren von Deskriptoren verbunden sind. Die Wahl und Formulierung von Deskriptoren wird in den folgenden Abschnitten eingehend behandelt. In den Beispielen sind die Deskriptoren frei gewählt und nicht einem vorgegebenen Vokabular entnommen worden.

Die bei der Indexierung einem Dokument zugeordneten Deskriptoren werden zusammen mit den bibliographischen Daten (siehe DIN 1505 Teil 1) und gegebenenfalls weiteren Daten in einen Dokumentationspeicher eingegeben und dienen dort zur Recherche. Die Deskriptoren müssen außerdem in einem Vokabular (z. B. Thesaurus) registriert sein bzw. werden, damit eine terminologische Kontrolle (Klären der Begriffsbedeutung, Zusammenführen von Synonymen, Kennzeichnen von Homonymen, Spezifizieren von Begriffsrelationen) möglich ist (siehe auch DIN 1463 Teil 1).

3.2 Kriterien zur Deskriptoren-Auswahl

Zur Erreichung einer möglichst präzisen Wiedergabe des Dokumenteninhalts sollte die Indexierungstiefe (Indexierungsbreite und -spezifität) so groß sein, wie es unter den zu erwartenden Fragestellungen der Benutzer des betreffenden Dokumentationssystems noch sinnvoll erscheint. Beachtet werden sollte auch, daß sich die Interessen der Benutzer wandeln können und bei der Indexierung ausgelassene Themenbereiche (perspektivisches Indexieren siehe DIN 31 623 Teil 1/09.88, Abschnitt 6) später nicht mehr oder nur unter sehr großem Aufwand erschlossen werden können.

3.2.1 Indexierungsbreite

Bei der Analyse und Interpretation eines Dokumententextes ist das Wesentliche vom Unwesentlichen zu scheiden, und implizit enthaltene Begriffe sind aufzuspüren. Dabei steht der Gesichtspunkt des Retrievals im Vordergrund.

Ein inhaltskennzeichnender Deskriptor ist als wesentlich anzusehen, wenn das Dokument als Antwort auf eine Fragestellung mit diesem Deskriptor oder in Verknüpfung mit weiteren Deskriptoren gefunden werden soll.

Ein Inhaltsmerkmal darf dann als unwesentlich betrachtet werden, wenn in dem betreffenden Dokumentationssystem die Relevanz dieses Dokuments in bezug auf jede beliebige Suchfrage mit an Sicherheit grenzender Wahrscheinlichkeit von dem fraglichen Inhaltsmerkmal unabhängig ist.

In Zweifelsfällen hat sich der Indexierer **für** die Vergabe eines Deskriptors zu entscheiden.

Die Anzahl der Deskriptoren je Dokument soll nicht willkürlich begrenzt werden. Sie richtet sich nach dem Informationsgehalt des betreffenden Dokuments.

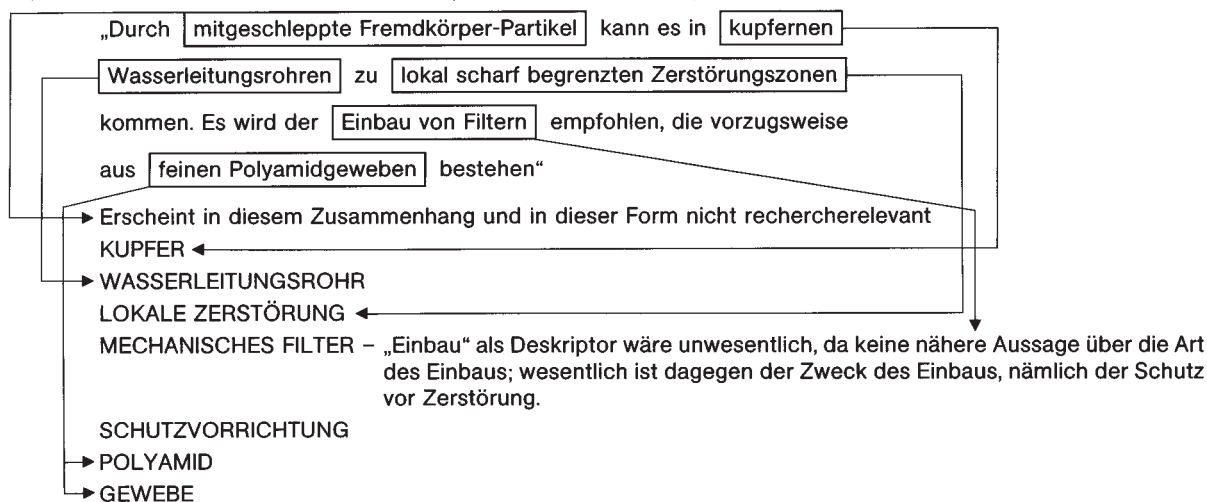
Sind in einem Dokument mehrere voneinander trennbare Inhaltsbereiche angesprochen (z. B. bei interdisziplinären Arbeiten), kann das Dokument entweder als Ganzes von verschiedenen Fachleuten indexiert oder in mehrere einzeln zu indexierende Teile aufgeteilt werden. Besonders bei nichtsyntaktischer Indexierung kann die Aufteilung in einzeln zu indexierende Teile oft empfehlenswert sein, um eine allzu große Indexierungsbreite zu vermeiden und dadurch die Wahrscheinlichkeit von Fehlkombinationen bei der Suche zu verringern.

3.2.2 Indexierungsspezifität

Als Deskriptoren sind jeweils die Bezeichnungen zu wählen, die dem Spezifitätsgrad der Aussagen des Dokuments entsprechen und zugleich für die Inhaltswiedergabe wesentlich (siehe Abschnitt 3.2.1) sind. Entbehrlich sind Inhaltsmerkmale, die im Zusammenhang des Dokumenteninhalts unwesentlich sind.

Beispiel 4: 1)

Text, der den Inhalt eines Dokuments in komprimierter Form wiedergibt:



Zutreffend wären auch Deskriptoren, die den vorstehenden Bezeichnungen äquivalent sind.

1) Dieses Beispiel dient lediglich zur Erläuterung der Frage der Indexierungsspezifität. Hierbei sind weitere, in der Folge behandelte Regeln angewendet worden. Andererseits deckt das Beispiel nicht alle notwendigen Überlegungen und Regeln ab.

Entbehrlich sind Deskriptoren, die mit den inhaltskennzeichnenden Deskriptoren in hierarchischer oder sonstiger semantischer Relation stehen (Begriffsumfeld), aber im Dokument selbst nicht angesprochen sind. Dabei wird vorausgesetzt, daß ein Vokabular vorhanden ist, welches über solche Relationen Auskunft gibt, und daß das Dokumentationssystem hierarchisches Suchen ermöglicht.

Wenn spezielle Angaben im Dokument aus Randgebieten des dokumentierten Fachbereichs stammen und/oder vom Standpunkt des Retrievals abseits des Interesses liegen, kann es zweckmäßig sein, als Deskriptoren Oberbegriffe repräsentierende Bezeichnungen zu wählen, die im hierarchischen Niveau über den explizit enthaltenen Begriffen liegen.

3.2.3 Auswahl freier Deskriptoren

Die folgenden Regeln gelten im wesentlichen für solche Indexierungsmethoden, die die Zuteilung freier Deskriptoren wenigstens teilweise erlauben. So bedarf die in DIN 31 623 Teil 1/09.88, Abschnitt 4.7.3, empfohlene Kombination von verbindlich vorgegebenem und freiem Vokabular einer ergänzenden Regelung für die Zuteilung freier Deskriptoren, selbst wenn diese nur einen relativ kleinen Anteil an der gesamten Indexierung ausmachen sollte.

3.2.3.1 Wiedergabetreue der Deskriptoren

Die Deskriptoren sind – mit Ausnahme offenkundiger Fehler und soweit sie nicht am Rande des Dokumentationsgebiets liegen – in Anlehnung an Titel und Text des Dokuments zu wählen. Bei der Übernahme von Begriffsbezeichnungen aus dem vorliegenden Text ist jedoch kritisch zu prüfen, ob die gewählten Deskriptoren eindeutig (siehe Abschnitt 3.2.3.2), verlässlich (siehe Abschnitt 3.2.3.3) und ausreichend prägnant (siehe Abschnitt 3.2.3.4) sind, sowie in ihrer Fundamentalkategorie dem Dokumenteninhalt entsprechen (siehe Abschnitt 3.2.4). Insoweit müssen die ursprünglich aus Titel oder Text gewählten Benennungen gegebenenfalls modifiziert oder ersetzt werden.

3.2.3.2 Eindeutigkeit der Deskriptoren

Im Hinblick auf die gegebenenfalls nachfolgende Einordnung in das Begriffssystem und auf die Recherche sollen die den Dokumenteninhalt wiedergebenden Deskriptoren auch in isolierter Betrachtung – losgelöst vom Textzusammenhang – Aussagekraft und Informationswert besitzen. Bei der Suche mit einzelnen Deskriptoren kann deren gemeinsames Vorkommen mit anderen Deskriptoren, mit denen das betreffende Dokument ebenfalls indexiert worden ist, im Dokumentationsspeicher nicht vorausgesehen werden. Deshalb soll jeder einzelne Deskriptor das Gemeinte – unabhängig von den weiteren indexierten Begriffen – möglichst klar und eindeutig wiedergeben.

Daraus ergibt sich, daß aus dem Text des Dokuments entnommene Bezeichnungen in dieser Hinsicht zu prüfen und gegebenenfalls zu präzisieren sind, z. B. durch adjektivische Zusätze. Bezeichnungen, die verschiedene Bedeutungen besitzen können (Homonyme), müssen in eine eindeutige Form überführt werden.

3.2.3.3 Verlässlichkeit der Deskriptoren

Im Regelfall ist eine Indexierung in ihrer **Reproduzierbarkeit** („Indexierungskonsistenz“) um so verlässlicher,

- a) je spezifischer die einzelnen Deskriptoren sind und
- b) je genauer die Deskriptoren insgesamt dem Inhalt des Dokuments entsprechen und um so weniger verlässlich,
- a) je allgemeiner²⁾ oder trivialer die Deskriptoren sind und
- b) je weiter sich die Deskriptoren vom Inhalt eines Dokuments entfernen.

Eine spezifische, dem Dokumenteninhalt entsprechende Indexierung hat größere Aussichten auf reproduzierbare Handhabung als eine unspezifische Indexierung.

3.2.3.4 Prägnanz der Deskriptoren

Bei der Indexierung ist eine knappe und treffende (prägnante) Formulierung des gemeinten Begriffs anzustreben. Bei Substantiven mit Attribut kann man z. B. versuchen, ein sinnentsprechendes, gebräuchliches Kompositum zu bilden.

Beispiel 5:

„Bau von Wohnungen“ → WOHNUNGSBAU

Hierbei ist zu beachten, daß Benennungen mit fortschreitender Kürzung (siehe DIN 2330) an Klarheit verlieren können.

Beispiel 6:

FIEBERMITTEL
kann entstanden sein aus „fiebersenkendes Mittel“
oder aus „fiebererzeugendes Mittel“

3.2.3.5 Gebräuchlichkeit der Deskriptoren

Bei der Auswahl von Deskriptoren ist im allgemeinen darauf zu achten, daß die Deskriptoren dem überwiegenden Sprachgebrauch des jeweiligen Fachgebiets entsprechen. Besondere Probleme dabei sind:

- a) Individualprägungen einzelner Autoren oder Schulen (soweit noch nicht in den Sprachgebrauch eingegangen).
- b) Neuschöpfungen des Indexierers. Dies ist zwar manchmal bei Kürzung (siehe Abschnitt 3.2.3.4) nicht zu vermeiden, dabei sollte jedoch darauf geachtet werden, daß der neugeprägte Deskriptor sich an den herrschenden Sprachgebrauch anschließt (z. B. durch Analogiebildung zu eingeführten Termini).
- c) Wörter der Werkstattsprache (Slang, Jargon). Sie sollten nur dann angewendet werden, wenn es keine den vorhergenannten Forderungen entsprechenden Termini gibt.

²⁾ Derartige Allgemeinbegriffe können z. B. sein: Untersuchung, Behandlung, Ergebnis, Problem, Frage, Aufgabe, Bedeutung, Vorgang, Zustand, Abhängigkeit, Einfluß, Folge, Vergleich, Zusammenhang, Beziehung, Verhältnis, Übergang, Änderung, Ursache, Wirkung, Bedingung, Effekt, Faktor, Anlage usw.

3.2.4 Kategorieller Gesichtspunkt der Deskriptoren-Auswahl

3.2.4.1 Fundamentalkategorie

Eine Fundamentalkategorie ist das Ergebnis der Teilung universeller Begriffe in allgemeine Teile, die im Prinzip auf alle Begriffsfelder anwendbar ist.

Solche Fundamentalkategorien sind beispielsweise

- Material (Stoff)
- Eigenschaft
- Vorgang
- Vorrichtung
- Lebewesen

Es ist zu empfehlen, jeweils die Fundamentalkategorien, die für ein bestimmtes Fachgebiet relevant sind (siehe Abschnitt 3.1.2), zusammenzustellen und bei der Indexierung zu berücksichtigen.

3.2.4.2 Indexierung in den zutreffenden Kategorien

In Dokumenten kann ein und derselbe Sachverhalt in verschiedenen Fundamentalkategorien ausgedrückt sein. Ein Vorgang kann sprachlich durch eine Vorrichtung, eine Vorrichtung durch den Vorgang umschrieben sein. Eine stoffliche Erscheinungsform oder eine Stoff-Funktion kann in Form einer Eigenschaft oder eines Vorgangs benannt sein und umgekehrt.

Beispiele 7 bis 10:

FILMBILDENDES MITTEL (Stoff-Funktion) – FILMBILDUNG (Vorgang)
 TENSID (Stoff-Funktion) – OBERFLÄCHENAKTIVITÄT (Eigenschaft)
 KATALYSATOR (Stoff-Funktion) – KATALYSE (Vorgang)
 POLARISATOR (Vorrichtung) – POLARISATION (Vorgang)

Für eine sachgerechte Auswertung und Indexierung ist wesentlich, daß die ausgewählten Begriffe in **den** Fundamentalkategorien formuliert sind, die den Inhaltsschwerpunkt und die gemeinten Aussagen am besten treffen. Durch eine Angabe in der falschen Fundamentalkategorie können Themen vorgetäuscht werden, die im Dokument gar nicht behandelt sind. Andererseits müssen bei der Indexierung alle Fundamentalkategorien vertreten sein, zu denen im Dokument wesentliche Aussagen gemacht sind.

Beispiel 11:

Text aus einem Dokument: „Die Reaktion wurde im Autoklaven durchgeführt.“

Verfälschende Indexierung: AUTOKLAV

Richtige Indexierung: DRUCKREAKTION gegebenenfalls + AUTOKLAV

Begründung:

Über die Vorrichtung „Autoklav“ wird im vorstehenden Dokumententext nichts ausgesagt; die Benennung „Autoklav“ (= Druckgefäß) dient vielmehr zur näheren Charakterisierung des Vorgangs: Reaktion unter Druck. „Autoklav“ kommt jedoch als zusätzliche Angabe in Frage, wenn außer diesem noch andere Typen von Druckgefäßen in Betracht zu ziehen sind.

3.2.5 Behandlung quantitativer und gegensätzlicher Eigenschaftsbegriffe

3.2.5.1 Quantitative Angaben von Eigenschaften

In einem Dokument können zu bestimmten Eigenschaften oder Größen quantitative Angaben (z.B. technische Daten) genannt sein.

Mit Ausnahme der Datendokumentation³⁾ ist es nicht Ziel der Indexierung, die konkreten Werte von Eigenschaften oder Größen wiederzugeben. Vielmehr ist durch die Indexierung zum Ausdruck zu bringen, **welche** Eigenschaften oder Größen nach den Angaben des Autors eine Rolle spielen.

Beispiel 12:

Text des Dokuments: „Der Wert des Kapazitätsnormals ist praktisch unabhängig von der Betriebsspannung. Bei einer Änderung der Betriebsspannung innerhalb weiter Grenzen wurde eine relative Kapazitätsänderung von nur $2 \cdot 10^{-10}$ je Volt gemessen.“

Indexierung: KAPAZITÄTSNORMAL / KAPAZITÄT (KONDENSATOR) /
 SPANNUNGSABHÄNGIGKEIT / BETRIEBSSPANNUNG

In manchen Fällen ist es jedoch nützlich – erforderlichenfalls neben den Eigenschaften oder Größen selbst – deren **pauschalen** Wertebereich zu indexieren. Dabei können gebräuchliche Skalen (z. B. Erdbebenskale, Güteklassen von Waren) durchaus angewendet werden.

Beispiel 13:

Text des Dokuments: „Der Laser arbeitet auf einer Wellenlänge von 320 nm.“

Indexierung: LASER / ULTRAVIOLETTSTRAHLUNG

³⁾ Die Datendokumentation ist nicht Gegenstand dieser Norm (siehe DIN 31 623 Teil 1/09.88, Abschnitt 1).