

ISO 25964 –

Standard der Information und Dokumentation:

*„Thesauri and interoperability
with other vocabularies“*

Jutta Lindenthal, Lübeck

Esther Scheven, Deutsche Nationalbibliothek Frankfurt

Inhalt von ISO 25964

ISO 25964: Thesauri and interoperability with other vocabularies

- Part 1: Thesauri for information retrieval
- Part 2: Interoperability with other vocabularies

Der Teil 1 umfasst monolinguale und multilinguale Thesauri.

Der Teil 2 beschäftigt sich mit "mappings" (Thesaurusförderung) zwischen Thesauri und anderen Typen von Vokabularien, die der Informationsgewinnung (information retrieval) dienen.

Beide Teile gibt es nur als Entwürfe:

Part 1: DIS 25964-1 (September 2009 im ISO Central Secretariat eingereicht)

Part 2: Überarbeitung Ende Juni 2009 begonnen;
erster Committee Draft (CD 25964-2) für Juni 2010 geplant

Vorgänger von ISO 25964

- ISO 2788-1986 Guidelines for the establishment and development of monolingual thesauri

= BS 5723:1987

- ISO 5964-1985 Guidelines for the establishment and development of multilingual thesauri

= BS 6723:1985

BS 8723: Structured vocabularies for information retrieval – Guide

Part 1: Definitions, symbols and abbreviations

Part 2: Thesauri

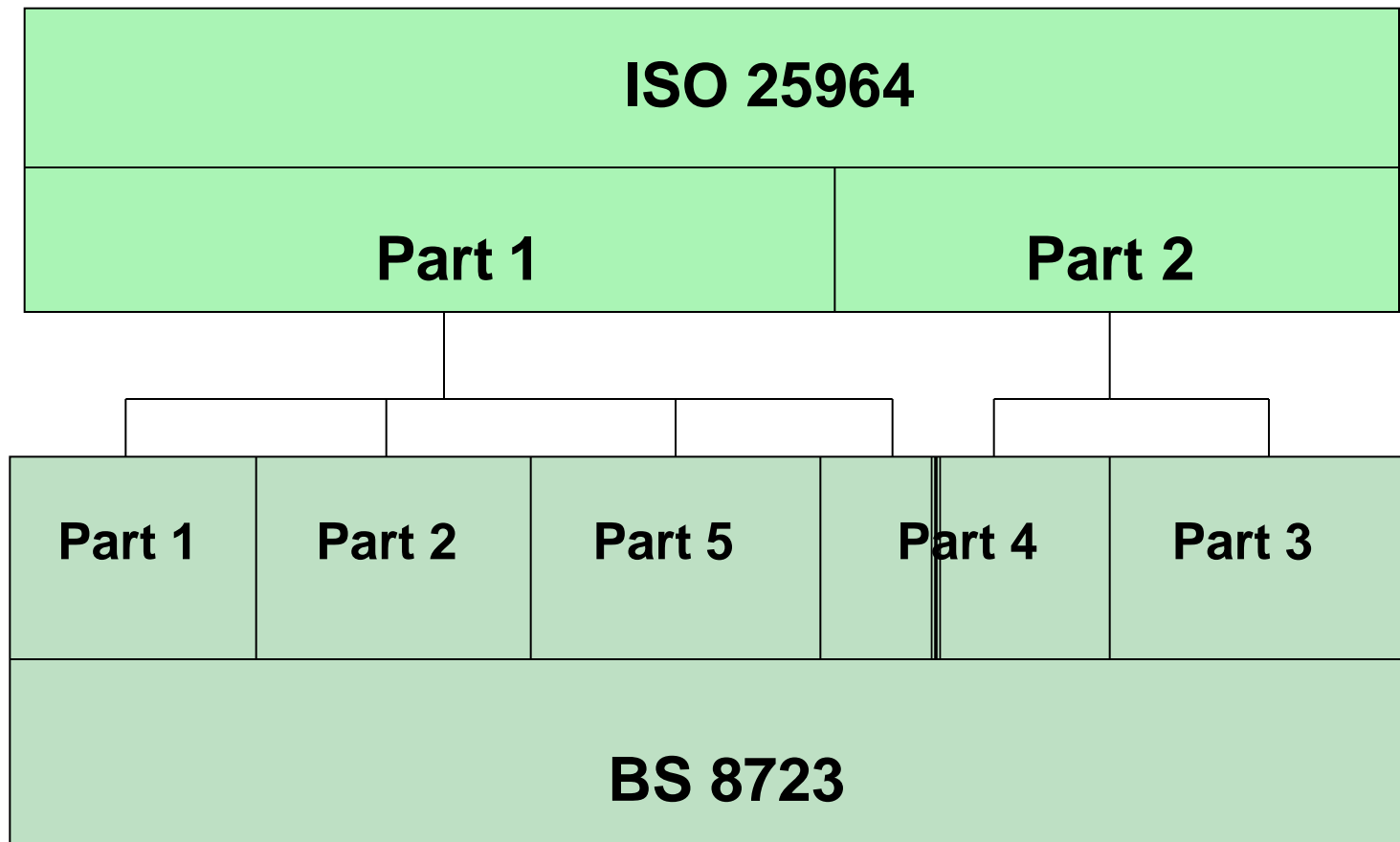
Part 3: Vocabularies other than thesauri

Part 4: Interoperability between vocabularies

Part 5: Exchange formats and protocols for interoperability

veröffentlicht 2005-2008

ISO 29564 basiert auf BS 8723



Part 1: Thesauri for information retrieval

Für die internationalen Standards ISO 2788 (monolinguale Thesauri, 1986) und ISO 5964 (multilinguale Thesauri, 1985) wurde ein Aktualisierungsbedarf festgestellt, besonders im Hinblick auf

- Interoperabilität
 - zwischen verschiedenen Vokabulartypen
 - zwischen verschiedenen Repräsentationsformaten (SKOS, Zthes ...)
- Anpassung an elektronische Umgebungen.

Die Revision erfolgt auf Basis des fünfteiligen britischen Standards BS 8723 (s. Literaturhinweise auf <http://www.jlindenthal.de/mb/index.html#normen>).

Beteiligt sind 15 Länder (in alphabetischer Reihenfolge): Bulgarien, China, Dänemark, Deutschland, Finnland, Frankreich, Großbritannien, Kanada, Korea, Neuseeland, Schweden, Spanien, Südafrika, Ukraine, USA

Projektleiterin: Stella Dextre Clarke, Großbritannien

Einige Erweiterungen in ISO 25964-1

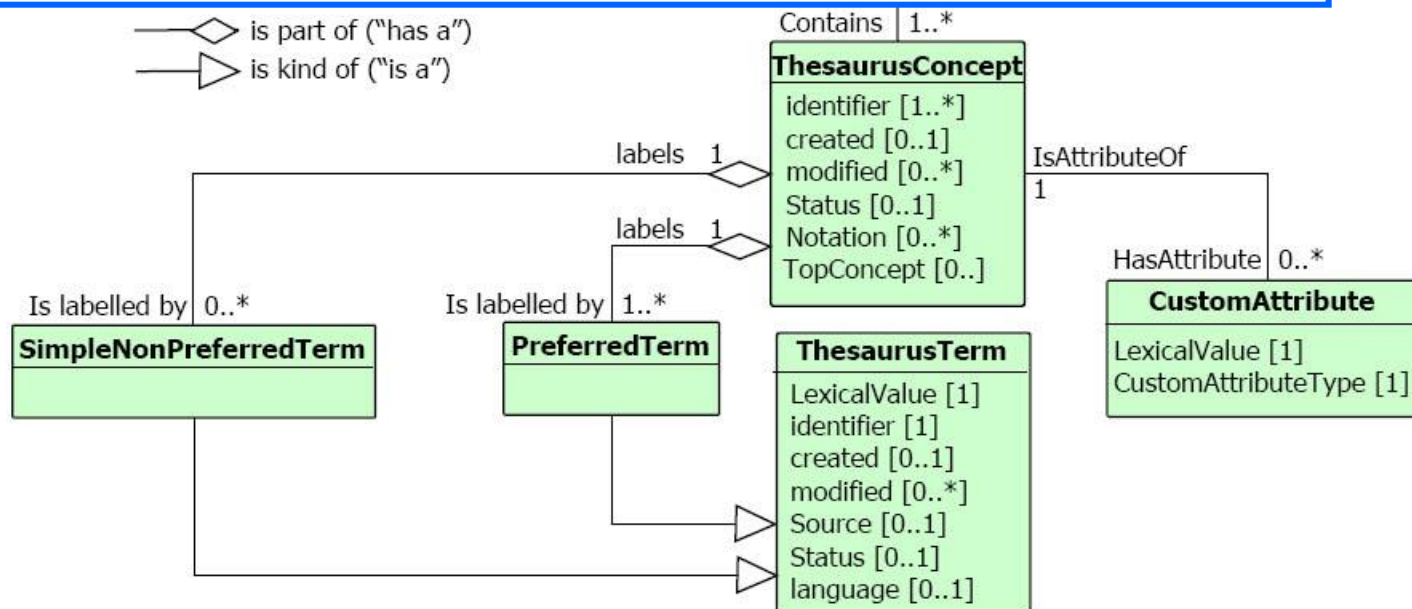
- ausführlichere Darstellung der Facettenanalyse
 - einige Änderungen bei der Begriffszerlegung
 - erweiterte Empfehlungen zur Thesauruserstellung und Thesauruspflege
 - Ergänzungen bei Anforderungen an Thesaurusmanagement-Software
 - Entwicklung eines Datenmodells
 - Das Datenmodell ist dargestellt als UML-Diagramm
 - erweitert BS8723-5-Full
- URL: <http://schemas.bs8723.org/2007-06-01/Documentation/BS8723-5-Full.html>

Das Modell repräsentiert

- Begriffe eines Thesaurus (ThesaurusConcept)
- traditionelle hierarchische und assoziative Relationen sowie nutzerdefinierbare Relationstypen zwischen Begriffen
- textuelle Anmerkungen wie Erläuterungen (Scope notes), Änderungsvermerke, Definitionen, redaktionelle Anmerkungen ...
- bevorzugte Benennungen (Deskriptoren) und nichtbevorzugte Benennungen (Nichtdeskriptoren) sowie nichtbevorzugte Komposita (Verweisungen in Begriffskombinationen)
- Begriffsreihen (Arrays) und Node labels
- Begriffsgruppen, anwendungsspezifisch typisierbar

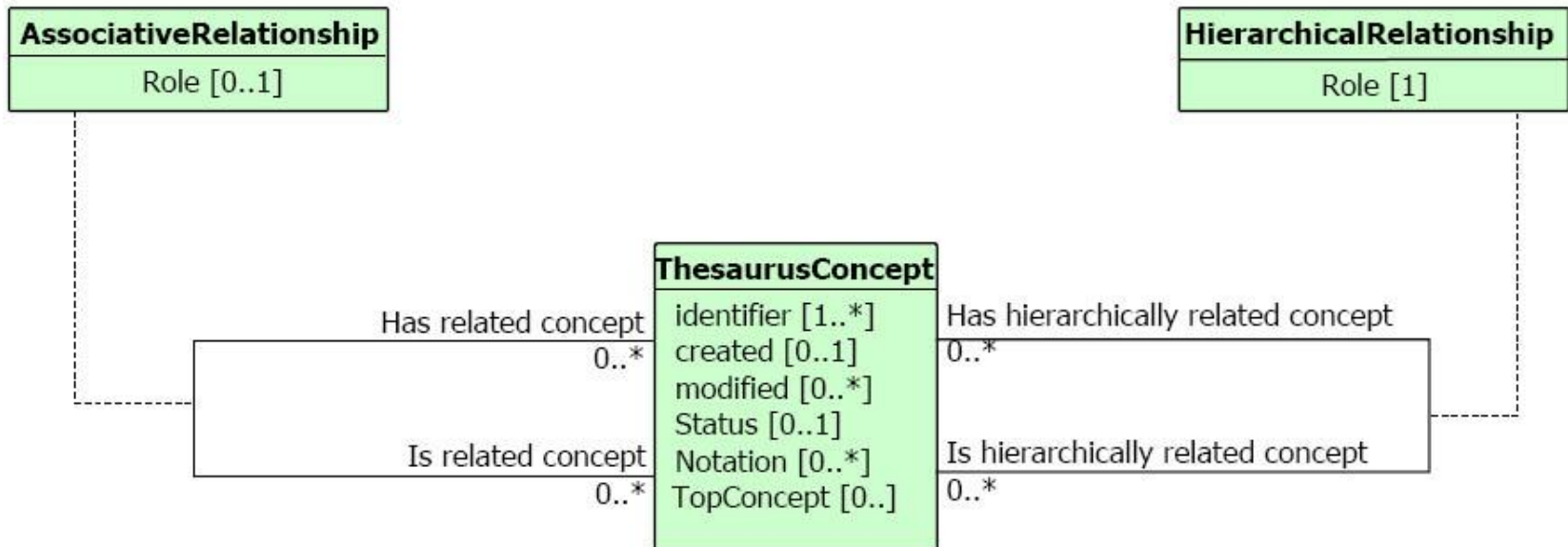
Begriffe (ThesaurusConcept) und Bezeichnungen (ThesaurusTerm)

- Ein Thesaurus besteht aus Begriffen (ThesaurusConcept)
- Begriffe werden durch ihre natürlichsprachigen Bezeichnungen (ThesaurusTerm) repräsentiert
- Jeder Begriff hat genau eine bevorzugte Benennung (PreferredTerm; Deskriptor) pro Sprache
- Jeder Begriff kann beliebig viele alternative Benennungen (SimpleNonPreferredTerm; Nichtdeskriptor) haben



Relationen zwischen Begriffen (ThesaurusConcept)

- Hierarchierelationen können allgemein (BT/NT) oder spezifiziert nach Rolle (BTG/NTG, BTP/NTP, BTI/NTI) ausgewiesen werden
- Assoziationsrelationen können allgemein (RT) oder spezifiziert nach Rolle (Teil/Ganzes, Ursache/Wirkung, ...) ausgewiesen werden

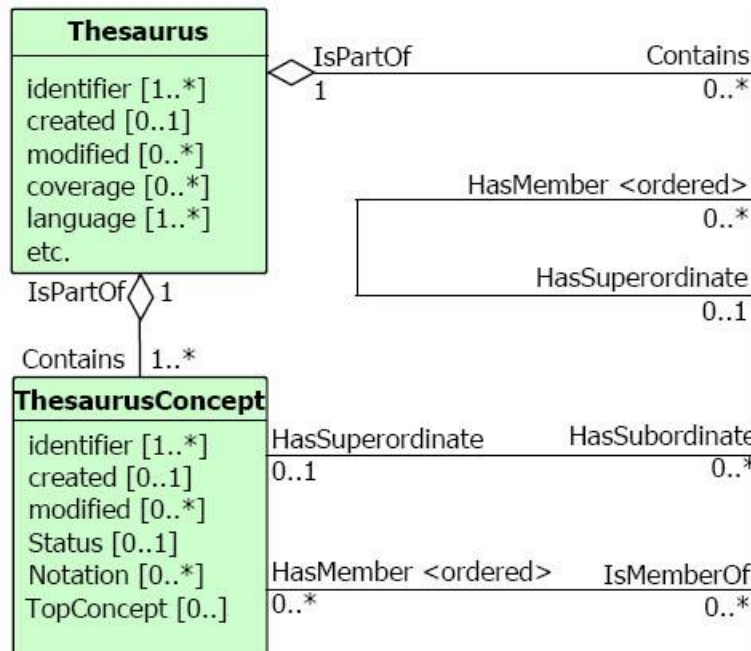
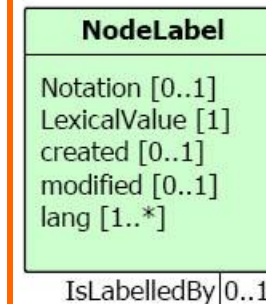


Beispiel Begriffssatz (Deskriptorensatz): Parlamentsthesaurus (Parthes)

Schlagwort: <input checked="" type="checkbox"/>	Konferenz	Bevorzugte Benennung
Thesaurus:	Sachbegriffe	
Definition:	Größere fachliche oder politische Versammlung, Tagung	
Benutzt für:		
	Internationale Konferenz	
	Internationaler Kongress	Nichtbevorzugte Benennung
	Internationale Tagung	
	Kolloquium	
	Kongress	
	Symposium	
	Tagung	
Oberbegriffe:		
<input type="checkbox"/>	→ Veranstaltung	Hierarchierelation
Unterbegriffe:		
<input type="checkbox"/>	→ Gipfeltreffen	
<input type="checkbox"/>	→ Regierungskonferenz der EU	
Verwandte Begriffe:		Assoziationsrelation
<input type="checkbox"/>	→ Diplomatie	
<input type="checkbox"/>	→ Parteitag	
Zuordnung:		Begriffsgruppe
	→ SO.05 - Soziales Leben	

Begriffsreihe und Node label

- Eine Begriffsreihe (ThesaurusArray) ist eine Gruppierung von Geschwisterbegriffen (Begriffe, die einen gemeinsamen Oberbegriff haben).
- Ein Node label bezeichnet das Unterteilungsmerkmal, das zur Bildung der Begriffsreihe herangezogen wurde; z.B. (*Personen nach Alter*)
- Einem Node label können beliebig viele (meist spezifischere) Node labels untergeordnet sein. N.B.: Diese Relation ist keine BT/NT-Relation.



- Einem Node label können beliebig viele Begriffe (ThesaurusConcept) untergeordnet sein. N.B.: Diese Relation ist keine BT/NT-Relation.

Beispiel Begriffsreihe und Node label: Art & Architecture Thesaurus (AAT)

Built Environment (Hierarchy Name)

.... Settlements and Landscapes

..... inhabited places

..... <settlements by function>

Node labels werden im AAT *Guide Terms* genannt. Dem Node label <settlements by function> sind die mit Pfeilen markierten Node labels untergeordnet.

..... <settlements by function: administrative>

..... capital cities

..... cathedral cities

..... colonial cities

Begriffs-
reihe

..... <local administrative centers>

..... missions (settlements)

..... municipalities

..... pueblos (settlements)

Alle hier gelisteten Begriffe, wie z.B. *capital cities*, *municipalities*, *bastides* oder *military towns* sind Unterbegriffe (NT) des Oberbegriffs (BT) *inhabited places*. Node labels gruppieren diese Begriffe unter bestimmten Aspekten (z.B. nach Art der Funktion: *administrativ* oder *militärisch*). Die Begriffe unter einem Node label bilden eine Begriffsreihe (ThesaurusArray).

..... <settlements by function: military>

..... bastides

..... castelli (fortified towns)

..... military towns

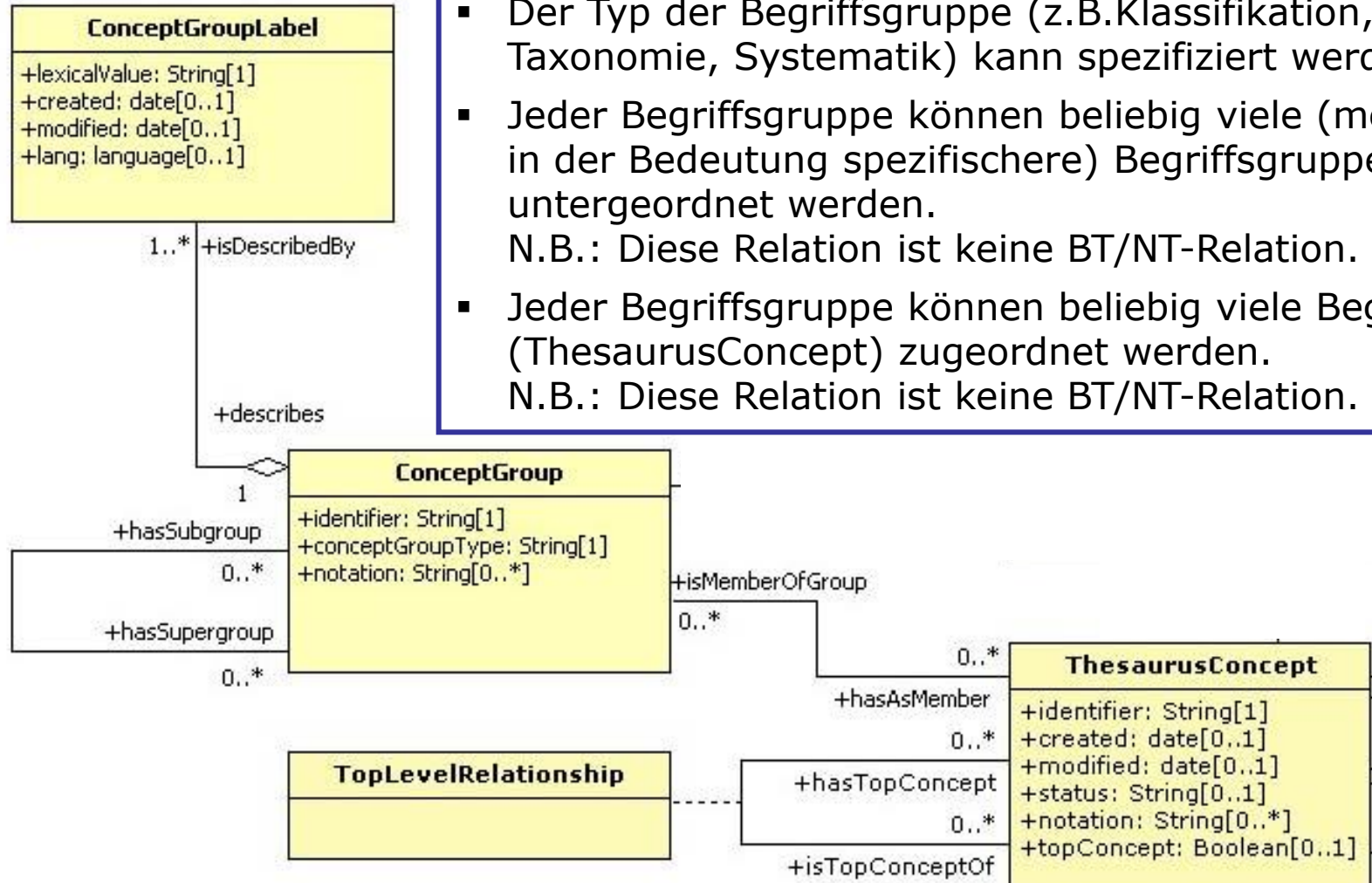
..... presidios

Quelle: <http://www.getty.edu/vow/AATHierarchy?find=&logic=AND¬e=&subjectid=300008407>

Begriffsgruppe (ConceptGroup)



- Begriffsgruppen repräsentieren systematisierende Strukturen wie z.B. Klassen einer Klassifikation, Teilthesauri, Sachgruppen, Kategorien, etc.
- Der Typ der Begriffsgruppe (z.B. Klassifikation, Taxonomie, Systematik) kann spezifiziert werden.
- Jeder Begriffsgruppe können beliebig viele (meist in der Bedeutung spezifischere) Begriffsgruppen untergeordnet werden.
N.B.: Diese Relation ist keine BT/NT-Relation.
- Jeder Begriffsgruppe können beliebig viele Begriffe (ThesaurusConcept) zugeordnet werden.
N.B.: Diese Relation ist keine BT/NT-Relation.



Beispiel Begriffsgruppe: Standard-Thesaurus Wirtschaft (STW)

- Home
- Alphabetische Liste der Deskriptoren

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

Begriffsgruppen; im STW
Hauptteile genannt

B Betriebswirtschaft

B Business economics (engl.)

untergeordnet

- ▶ B.00 Betriebswirtschaft
- ▶ B.01 Unternehmensführung
- ▶ B.02 Investition und Finanzierung
- ▶ B.03 Betriebswirtschaftliches Prüfungswesen
- ▶ B.04 Personalwirtschaft
- ▶ B.05 Materialwirtschaft und Logistik
- ▶ B.06 Produktionswirtschaft
- ▶ B.07 Marketing
- ▶ B.08 Betriebliches Steuerwesen
- ▶ B.09 Wirtschaftsinformatik
- ▶ B.10 Unternehmensforschung

Der Begriffsgruppe
B untergeordnete
spezifischere
Begriffsgruppen

Der Begriffsgruppe
B.07 zugeordnete
Begriffe

Deskriptoren

- Handelsspanne
- Internationales Marketing
- Investitionsgütermarketing
- Investor Relations
- Kauf
- Key-Account Management
- Konsumgütermarketing

Part 2: Interoperability with other vocabularies

Ziel (Scope):



“This part of ISO 25964 deals with thesauri and other types of vocabulary that are commonly used for information retrieval. It describes, compares and contrasts the elements and features of these vocabularies that are implicated when interoperability is needed. It gives recommendations for the establishment and maintenance of mappings between multiple thesauri, or between thesauri and other types of vocabularies.”

- > Handreichung zum “mapping” zwischen Thesauri und anderen Vokabularien; nicht zwischen den Vokabularien unter sich
- > Terminologiarbeit (nicht: Erschließungsarbeit)

Unterschied zu BS 8723: Ausgewählte Vokabularien nur soweit beschreiben, wie notwendig ist, um die speziellen Anforderungen an ein “mapping” mit einem Thesaurus erklären zu können.

Vokabulartypen: solche, die dem “information retrieval” dienen (keine Lexika oder Wörterbücher)

Part 2: Interoperability with other vocabularies



Inhalt:- allgemeine Aussagen zu Thesaurusföderation

- Definitionen, Symbole
- Strukturell verschiedene Modelle zur Erlangung von Interoperabilität zw. Vokabularien
- Anweisungen zur Erstellung von „mappings“
- Einfluss der Anwendung auf das „mapping“
- Datenverwaltung, Darstellung (display), Funktionalität
- Austauschformat

- Darstellung von bestimmten Vokabularien unter dem Gesichtspunkt der Thesaurusföderation
 - Thesauri
 - Klassifikationen
 - Aktenpläne und Ordnungssysteme für die Schriftgutverwaltung
 - Taxonomien
 - Schlagwortlisten (subject heading schemes)
 - Ontologien, Topic Maps, Semantische Netze
 - Terminologiedatenbanken (Termbanken)
 - Normierte Namensverzeichnisse, Normdateien für Individualnamen (name authority lists)
 - Synonymringe

Part 2: Interoperability with other vocabularies

Achtung Baustelle!

Textteile aus dem BS 8723 wurden neu strukturiert zusammengestellt.

-> Umfangreiche inhaltliche und sprachliche Überarbeitungen sind notwendig, um das Ziel der Norm, Anweisungen zu geben für Terminologearbeit, um „mappings“ zwischen den Vokabularien herzustellen, konsequent umzusetzen.

*Aufruf zur ersten Sammlung von „Baustellen“ über ein Formular
Beispiele für „Baustellen“:*

- ausreichende Definitionen ??
- genauere Erläuterungen von Sachverhalten erforderlich
- allgemeiner Teil mit Anweisungen für das „mapping“ mischt sich mit Überlegungen zum Erschließen von Dokumenten mit heterogenen Vokabularien
- ...

Strukturell verschiedene Modelle zur Erlangung von Interoperabilität zw. Vokabularien

Zwei Typen von Vokabularien:

Strukturgleichheit (structural unity):

- Die Äquivalenz der Begriffe zweier oder mehrerer solcher Vokabularien wird durch die Position in der Vokabularstruktur ausgedrückt. Alle Begriffsrelationen entsprechen einander; die Bezeichnungen für die Begriffe können sogar automatisch miteinander zu Pärchen verbunden werden.

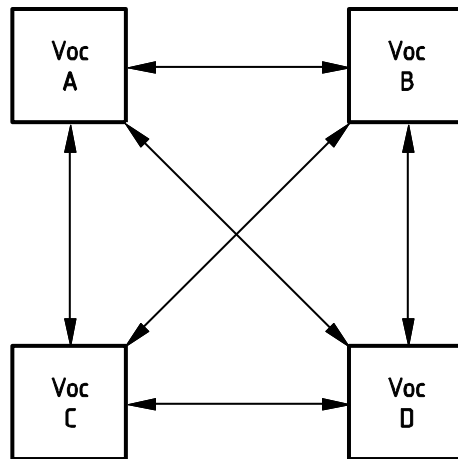
Strukturungleichheit (structural disunity):

- Die Äquivalenz der Begriffe zweier oder mehrerer solcher Vokabularien hat mit der Position in der Struktur nichts zu tun. Die Begriffe müssen unabhängig von der Struktur auf einander bezogen werden. Das „mapping“ erzeugt sowohl richtige Äquivalenzpaare als auch viele Paare, bei denen der Begriffsumfang und der Begriffsinhalt nicht wirklich übereinstimmen.

Strukturell verschiedene Modelle zur Erlangung von Interoperabilität zw. Vokabularien *Fortsetzung*

Zwei Möglichkeiten des „mappings“

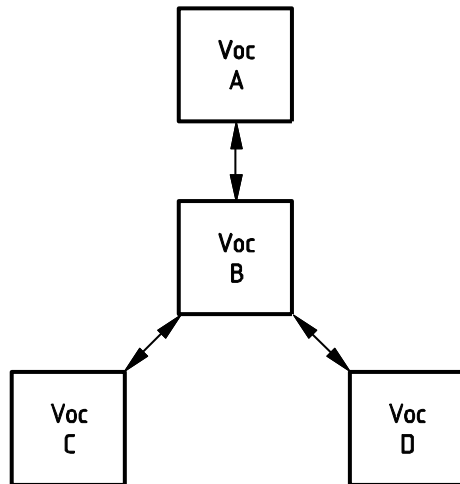
1. Alle Begriffe der verschiedenen Vokabularien werden jeweils miteinander zu Pärchen gebündelt.



- > bei strukturgleichen Vokabularien: nur echte Äquivalenzpaare
- > bei strukturungleichen Vokabularien: echte Äquivalenzpaare und viele Paare, die nur sehr schlecht zusammenpassen

Strukturell verschiedene Modelle zur Erlangung von Interoperabilität zw. Vokabularien *Fortsetzung*

2. Ein Vokabular dient als Drehscheibe („hub“); die Begriffe der verschiedenen Vokabularien werden jeweils nur mit den Begriffen dieses Vokabulars in Beziehung gesetzt.



- > für strukturungleiche Vokabularien geeignet,
- besonders wenn viele verschiedene Vokabularien miteinander verbunden werden sollen.
 - wenn es ein Vokabular mit herausgehobenem Status gibt.

Strukturungleichheit: Unterschiedliche Äquivalenzpaare

Echte Äquivalenzpaare: mit EQ kennzeichnen

„unechte Äquivalenzen“: unterschiedliche Stufen
der Übereinstimmung

„one-to-many-mappings“: - Aufzählung von einzelnen, engeren
Deskriptoren
- Kombination von mehreren
Deskriptoren

Einfluss der Anwendung auf das „mapping“

Bei der Entscheidung, nach welchem Modell die Äquivalenzpaarbildung verschiedener Vokabularien erfolgen soll, soll auch die geplante Anwendung in Betracht gezogen werden.

Dieser richtige Gedanke wird in der Norm z.Z. mit Anweisungen zur Erschließung von Dokumenten mit verschiedenen „gemappten“ Dokumentationssprachen mittels automatischer oder halbautomatischer Verfahren vermischt.

Interesse der DNB an ISO 29564-2

Pflege von Normdateien („name authority files“):

PND (Personennamendatei) <http://www.d-nb.de/standardisierung/normdateien/pnd.htm>

GKD (Körperschaftsdatei) <http://www.d-nb.de/standardisierung/normdateien/gkd.htm>

SWD (Schlagwortnormdatei) <http://www.d-nb.de/standardisierung/normdateien/swd.htm>

SWD: hybrid - Standardliste für Eigennamen
- Thesaurus für Sachschlagwörter

Projekte mit der SWD:

- Verlinkung mit anderen Vokabularien

Projekte: CARMEN <http://www.d-nb.de/wir/projekte/carmen.htm#hauptmenu>;

Crosskonkordanz STW / SWD <http://www.d-nb.de/wir/projekte/crosskonkordanz.htm>;

MACS <http://www.d-nb.de/wir/projekte/macs.htm>

- Verlinkung mit einer Klassifikation (DDC)

CrissCross-Projekt: <http://www.d-nb.de/wir/projekte/crisscross.htm> ; <http://linux2.fbi.fh-koeln.de/crisscross/index.html>

DNB überarbeitet Abschnitte für „name authority lists“ und „classification schemes“.

Einfluss der Anwendung auf das „mapping“

SWD und DDC:

nicht nur reine Begriffssystematisierung, sondern erwarteter „recall“ eines SWD-Descriptors wird mit der zu erwartenden DDC-Notation für das Dokument verlinkt.

800 |s|Kichererbse

808 |a|B 1986 (Quelle)

810 24.3;32.4 **SWD-Systematik: Botanik + Gartenbau**

816 583.74#2# [2007-01-01] **spez. Botanik**

816 633.37#2# [2007-01-01] **Feldfrucht**

816 635.657#3# [2007-01-01] **im Gartenbau**

816 641.35657#3# [2007-01-01] **Lebensmittelprodukt**

816 641.65657#1# [2007-01-01] **als Zutat beim Kochen**

830 |s|Cicer arietinum (Synonymieverweisung)

800 |s|Tomographie

808 |a|Reallex. Med., abc-Messtechn.

810 27.6;31.1c **SWD-Systematik: Medizin. Radiologie,
Nuklearmedizin; Mess-, Steuerungs- u. Regeltechnik**

816 616.0757#2# [2007-01-01] **Radiolog. Diagnostik**

816 636.08960757#2# [2007-01-01] **R.D. bei Tieren**

830 |s|Planigraphie

830 |s|Laminographie

830 |s|Röntgenschichtaufnahme

830 |s|Schichtaufnahmeverfahren

830 |s|Schichtbildtechnik

830 |s|Schichtröntgentechnik

830 |s|Tomografie

850 |s|Radiologische Diagnostik

850 |s|Schnittbilddiagnostik

Keine DDC-Notation aus den Ingenieurwissenschaften, weil der erwartete Literaturanfall („recall“) zu gering ist.

Diskussionspunkte:

- ein gemeinsames Modell für alle Vokabulartypen?
- jeweils ein Modell für einen Vokabulartyp?
- eine Syntax um die „mappings“ zwischen den Vokabularpaaren zu spezifizieren?

Die Arbeitsgruppe für Datenmodellierung freut sich über Unterstützung.

Interesse an Mitarbeit?

Wir geben gerne Auskunft

Jutta Lindenthal: jutta.lindenthal@googlemail.com

Esther Scheven: e.scheven@d-nb.de