

Anmerkungen zu einzelnen Folien der Präsentation

„Ein Wortschatz für den Kulturschatz“

Aufbau und Anwendung von Thesauri in der Museumsdokumentation

13. MusIS-Nutzertreffen, Mannheim – Jutta Lindenthal, 18.06.2012

Folie 4

Die Suche nach „fibel“ in Google-Bildern ergab im April 2009 rund 87.900 Treffer. Schätzungsweise die Hälfte davon ist Informationsballast, weil die Google-Suchmaschine nicht zwischen verschiedenen Bedeutungen der Bezeichnung "Fibel" unterscheiden kann. Dass die vertippte "Gewandfibel" dennoch im Ergebnis der Freitextsuche auftaucht, deutet darauf hin, dass die Suchmaschine häufige Falschschreibungen in die Suche mit einbezieht.

Folie 7

Wodurch kommt es zu Informationsballast und Informationsverlust? Verschiedene Nutzer haben oft recht unterschiedliche Vorstellungen von dem Objekt, das sie meinen und ...

Bezeichnungen der natürlichen Sprache sind meist nicht eindeutig. Der Nutzer weiß häufig nicht, mit welcher Bezeichnung er suchen soll.

Folie 9

Homonyme oder Polyseme, d.h. verschiedene Begriffe mit derselben Bezeichnung, führen zu Ballast beim Retrieval. Es werden auch Dokumente bzw. Objekte gefunden, die nicht gesucht waren.

Folie 10

Beim Retrieval führen Synonyme, die nicht kontrolliert und zu einem Begriff zusammengeführt wurden, zu Verlust von Information.

Folie 12

Die Aufnahme eines Deskriptors in einen Thesaurus sollte gerechtfertigt werden durch Belege und Verwendungsnachweise. Im Englischen wird dafür der Begriff „warrant“ gebraucht.

Literary warrant

Nachweis des Begriffes bzw. der Bezeichnung in der einschlägigen Literatur. Der Art & Architecture Thesaurus z.B. verlangt Belege aus drei einschlägigen Lexika oder Enzyklopädiën. Wikipedia gilt dabei nicht als autoritative Quelle.

User warrant

Beleg des Gebrauchs durch den Nutzer, z.B. durch Logfile-Analysen

Structural warrant

Notwendigkeit der Aufnahme eines Schlagwortes oder Node Labels zur hierarchischen Strukturierung des Vokabulars

Folie 13

Die terminologische Kontrolle allein genügt jedoch nicht, um Ballast zu vermeiden und die Präzision der Suche zu verbessern. Auch die Indexierung sollte möglichst spezifisch sein, d.h. es sollten immer die spezifischsten Begriffe verwendet werden, die der Thesaurus oder das kontrollierte Vokabular bereitstellt, um eine gute Retrievalqualität zu erreichen.

Im Beispiel "Distelfibel" ist der Typ "Fibel" zu unspezifisch, "Kunstgewerbe" ganz unpassend.

Folie 14

Der spezifischste Begriff des Thesaurus lässt sich mit Hilfe der Hierarchiebeziehung in Thesauri finden, z.B. durch die "Suche nach hierarchisch untergeordneten Begriffen" in der OGND.

Die Hierarchiebeziehung in Thesauri ist eine Beziehung zwischen allgemeineren Oberbegriffen und spezifischeren Unterbegriffen. Gemäß Thesaurusstandards sollte es immer eine Beziehung der logischen Über- und Unterordnung sein.

Folie 15

Die Wortungetüme sind nicht nur nicht wieder auffindbar, sondern sie sind oft sogar mehrdeutig, weil die Bezüge der einzelnen Wortbestandteile zueinander nicht erkennbar sind. Ist beim "Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz" ein Übertragungsgesetz gemeint oder geht es um Aufgabenübertragung, ...?

Folie 16

Wenn ein Thesaurus für die Indexierung benutzt wird, kann die Suche entweder automatisch oder durch Auswahl auf Unterbegriffe bzw. Oberbegriffe ausgeweitet werden. Bei dem auswahlbasierten Verfahren werden in die Suche alle Dokumente einbezogen, die mit den ausgewählten Schlagwörtern indexiert sind.

Folie 17

Ein Begriff kann auch zwei oder mehrere logische Oberbegriffe haben, je nach Aspekt, unter dem der Begriff betrachtet wird.

Folie 18

Mit Hilfe von „Node Label“ können Geschwisterbegriffe (Begriffe mit einem gemeinsamen Oberbegriff) unter bestimmten Aspekten spezifischer gruppiert werden. Node Label sind Nichtdeskriptoren, die in der Hierarchieleiter zwischen einem Oberbegriff (in diesem Beispiel "Fibel") und seinen Unterbegriffen eingefügt sind. Sie drücken das Merkmal aus, das der Gruppierung der Geschwisterbegriffe zugrunde liegt.

Node Label können auch für eine facettierte Suche verwendet werden, in der die Suchergebnisse sukzessive verfeinert werden können.

Folie 19

Die Freitextsuche nach „schmetterling“ hat im Portal Museen Nord neben Tassen, Vasen, Mineralien, einer Frisör-Preisliste u.v.a. auch diesen Löffel gefunden, weil ein Schmetterling eingraviert ist. Bei einer Suche nach dem Objekt "Schmetterling" wäre dieser Treffer Ballast. Bei der Erschließung müsste also zwischen einem Schlagwort, das das Objekt selbst beschreibt (Isness) und den Schlagwörtern, die über Inhalte oder Dargestelltes (Ofness, Aboutness) Auskunft geben, unterschieden werden.

"Hochzeitsgeschenk" ist die Verwendung oder Rolle des Objektes und müsste ggf. in einem entsprechenden Metadatenelement verzeichnet werden.

Die Deskriptoren eines Thesaurus sind Werte für die Metadatenelemente einer Datenbank. Je spezifischer die Datenbank strukturiert ist und je konsistenter die Werte aus den Vokabularen verwendet werden, desto besser sind die Retrievalergebnisse.