

Informationserschließung und Wissensorganisation [DIS 02]

Tutorial

Oktober 2022

Modulinhalte

Das Modul „Informationserschließung und Wissensorganisation (DIS 02)“ vermittelt durch theoretische Inhalte und eine praktische Aufgabenstellung die Grundlagen der Wissensorganisation und Wissensrepräsentation.

In der praktischen Aufgabenstellung wird eine strukturierte Datenbank zur Erfassung und Erschließung von Bildern angelegt. Die Bilder werden im Rahmen eines Erschließungskonzepts formal erfasst, durch Datensätze in einer Datenbankumgebung repräsentiert und inhaltlich durch aspektdifferenzierte Thesauri erschlossen. Dabei werden die Grundsätze zur Erstellung von Erschließungskonzepten, des begrifflichen Strukturierens und der Gestaltung von Dokumentationssprachen beachtet. In die Vorgehensweise der intellektuellen Erschließung durch Inhaltsanalyse und Zuteilung normierten Vokabulars auf Grundlage von Indexierungsprinzipien wird sowohl theoretisch als auch mit Blick auf die konkreten Beispiele eingeführt. Die erschlossenen Daten werden mit einer Retrievalumgebung suchbar gemacht.

Dieses Tutorial orientiert sich an der praktischen Aufgabenstellung für das Modul und dient als Basis für ein Selbststudium der Lehrinhalte.

Das Inhaltsverzeichnis auf der folgenden Seite enthält die Teilaufgaben des Programms, die in der genannten Reihenfolge zu bearbeiten sind.

Inhaltsverzeichnis

1	Einführung in die Thematik	3
2	Einrichten der Bilddatenbank „bild-db.dbm“ (Midos 6)	3
2.1	Arbeitsumgebung einrichten	3
2.2	Midos 6 installieren	3
2.3	Datenbank „bild-db.dbm“ öffnen	4
2.4	Arbeitsergebnis	4
3	Erfassen der formalen Dokumentdaten für die 50 Bilder (Midos 6)	4
3.1	Erweiterung der Datenbeschreibung	5
3.2	Erfassen der formalen Daten für die Bilder	5
3.3	Arbeitsergebnis	6
4	Inhaltsanalyse der Bilder	6
4.1	Erstellen der inhaltlichen Kurzcharakterisierungen	6
4.2	Arbeitsergebnis	6
5	Erstellen aspektorientierter Thesauri (MidosThesaurus)	6
5.1	Aspektdifferenzierte Thesauri	7
5.2	Thesauruserstellung mit MidosThesaurus	7
6	Inhaltliches Erschließen der Bilder durch Wortlisten und Thesauri (Midos 6, MidosThesaurus)	8
6.1	Erschließen durch Wortlisten	9
6.2	Erschließung durch Thesauri	9
6.3	Arbeitsergebnis	10
7	Erstellen einer Retrievalanwendung	10
7.1	Arbeitsergebnis	12
8	Musteraufgabe für die Modulprüfung 1020	12

1 Einführung in die Thematik

Lektüre Lesen Sie zur Einführung in die Thematik das [Kapitel 1](#) und die Abschnitte 2.1-2.4 des [Kapitels 2](#) des Lehrbuchs [Informationserschließung und Automatisches Indexieren](#).¹

Arbeitsaufwand: 6 Std.

Beachten Sie: Im Verlauf des Tutorials werden in den „Lektüre“-Abschnitten immer wieder Hinweise auf Abschnitte des Lehrbuchs gegeben, die für den entsprechenden Teil des Arbeitsprogramms relevant sind. Dies kann dazu führen, dass spätere Abschnitte des [Kapitels 2](#) zu einem eher frühen Zeitpunkt vorgeschlagen werden. Sollten Ihnen in diesen Fällen Fachbegriffe noch unbekannt sein, können Sie das „Sachverzeichnis“ auf den S. 427ff. verwenden, das für die im Buch verwendeten Fachbegriffe die Stellen nachweist, an denen diese behandelt werden. Zusätzlich kann für die begriffliche Klärung von Fachterminologie der [Thesaurus Informationserschließung](#) herangezogen werden.

2 Einrichten der Bilddatenbank „bild-db.dbm“ (Midos 6)

Für die Arbeit an der praktischen Aufgabenstellung wird eine bereits eingerichtete Datenbank im Minimalumfang zur Verfügung gestellt. Diese ist herunterzuladen und mit *Midos* einzurichten für die nachfolgenden Aufgaben.

Arbeitsaufwand: 2 Std.

Lektüre Lesen Sie zur Einführung in die Thematik den Abschnitt 8.1 des [Kapitels 8](#).

2.1 Arbeitsumgebung einrichten

Laden Sie den Ordner [bild-db.zip](#) herunter. Entpacken Sie die Archivdatei. Speichern Sie den Ordner in einem Verzeichnis, dessen *vollständiger Pfadname keine Leerzeichen und keine Sonderzeichen* enthält (beispielsweise auf der Hauptebene der Festplatte unter: „c:\bild-db“).

2.2 Midos 6 installieren

Laden Sie das Datenbanksystem *Midos 6* herunter und installieren es auf Ihrem Rechner: Entpacken Sie dazu die Archivdatei und führen Sie „setup.exe“ aus.

¹Das Lehrbuch kann als PDF innerhalb des Hochschulnetzes (bzw. von außerhalb über eine VPN-Verbindung) kostenlos heruntergeladen werden.

3.1 Erweiterung der Datenbeschreibung

Im Ordner „bilder“ im Datenbankverzeichnis „bild-db“ befinden sich 50 Bilddateien. Diese entsprechen den Bildern auf der Seite indexierung-retrieval.de. Die meisten der Bilder dort sind mit ihrer Quelle verlinkt, auf die Sie durch einen Klick auf das Bild gelangen (das klappt nicht immer, weil manche Quellen nicht mehr existieren). Dort können Sie ggf. weitere Informationen zum Bild finden. Technische Daten zu den Bildern erhält man im *Windows-Dateimanager* durch „Rechtsklick auf Dateinamen – Eigenschaften – Allgemein bzw. Details“.

Die Bilddatenbank soll Datensätze für alle 50 Bilder enthalten. Dazu ist es zunächst nötig, das Erfassungsschema für die Datenbank – *Midos 6* bezeichnet das Erfassungsschema als „Datenbeschreibung“ – um zusätzliche Kategorien zu erweitern. Den Dialog für die Erweiterung der Datenbeschreibung erreichen Sie über „Bearbeiten – Datenbeschreibung“. Dort entspricht jede Zeile der Tabelle einer Kategorie bzw. einem Feld für die Erfassung von Daten. Die Datenbeschreibung muss also für jedes Merkmal eines Bildes, das Bestandteil eines Datensatzes werden soll, eine Felddefinition enthalten.

Erweitern Sie die bereits vorhandene Datenbeschreibung um mindestens vier Kategorien für formale Merkmale von Bildern. Bitte bedenken Sie, dass Felder, die Sie anlegen, später auch von Ihnen sinnvoll befüllt werden müssen. Die Übernahme von Feldern aus Vorlagen ist gestattet, allerdings sollte Ihnen für die übernommenen Felder auch klar sein, welche Inhalte dort jeweils zu erfassen sind. Als Hilfestellung eignen sich Abschnitt 2.2 des [Kapitels 2](#) und die Folien 32 und 33 des [Vorlesungsskripts](#).

3.2 Erfassen der formalen Daten für die Bilder

Erfassen Sie mit dem *Datenbankeditor* von *Midos 6* die formalen Daten zu den 50 Bildern. Sie können für die Erfassung auch den sog. *Tabelleneditor* verwenden. Der Einsatz des *Tabelleneditors* ist allerdings nur für die formalen Daten möglich, da die spätere inhaltliche Erschließung eine darüber hinausgehende Funktionalität benötigt.

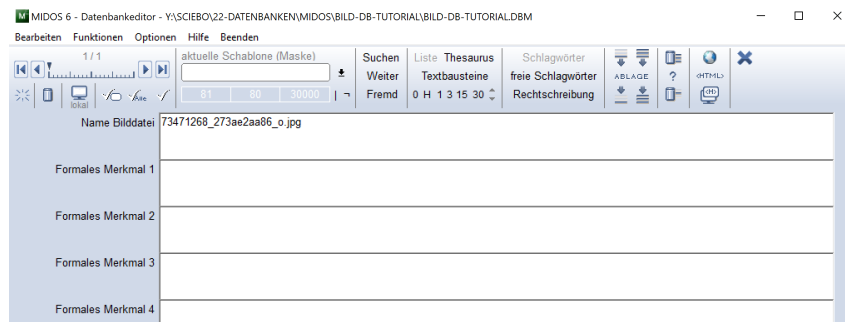


Abbildung 2: *Midos 6* – Datenbankeditor

Beachten Sie: Die Erfassung von Daten in einer Datenbank folgt immer dem Prinzip der Einheitlichkeit. Auch für vergleichsweise einfache Daten, wie beispielsweise die Dateigröße einer Bilddatei (Dateigröße: „345 kb“ vs. „345kb“ vs. „345KB“), ist es nötig, eine Form der

Eingabe verbindlich festzulegen und konsequent beizubehalten (vgl. die Folien 35 und 36 des [Vorlesungsskripts](#)). Ein nützliches Hilfsmittel zur Unterstützung einer einheitlichen Eingabe sind die sog. *Wortlisten*, die sich im *Datenbankeditor* von *Midos 6* vereinbaren lassen. Sie ermöglichen die Übernahme von Merkmalen aus einer Liste, sodass keine Schreibfehler bei der Eingabe mehr entstehen (vgl. Folie 36 des [Vorlesungsskripts](#)) und S. 79-80 des [Kapitels 2](#)). Zweckmäßig ist das Anlegen von Wortlisten v. a. für Merkmale einer insgesamt eher kleinen Merkmalsmenge, z. B. Dateiformate (jpg, tiff, png etc.).

3.3 Arbeitsergebnis

Eine Datenbank „bild-db.dbm“ mit insgesamt 50 Datensätzen mit formalen Merkmalen.

4 Inhaltsanalyse der Bilder

Die Inhalte der Bilder sollen analysiert und in einer sog. „inhaltlichen Kurzcharakterisierung“ zusammenfassend beschrieben werden.

Arbeitsaufwand: 4 Std.

Lektüre Lesen Sie zur Einführung in die Thematik den Abschnitt 2.9 des [Kapitels 2](#).

4.1 Erstellen der inhaltlichen Kurzcharakterisierungen

Ergänzen Sie die Datenbeschreibung der Datenbank „bild-db.dbm“ um ein Feld für die Erfassung einer „inhaltlichen Kurzcharakterisierung“, die den wesentlichen Inhalt des jeweiligen Bildes in einem Satz beschreibt. Auch dies kann in einheitlicher Form erfolgen, beispielsweise durch die Verwendung eines festen Satzanfangs wie „Das Bild zeigt ...“, z. B.: „Das Bild zeigt die Golden Gate Bridge und einen Teil der Bucht von San Francisco“.

Erfassen Sie für jedes der 50 Bilder eine inhaltliche Kurzcharakterisierung.

4.2 Arbeitsergebnis

Eine Bilddatenbank mit 50 Datensätzen, die formale Erschließungsmerkmale und eine freie inhaltliche Erschließung enthalten.

5 Erstellen aspektorientierter Thesauri (MidosThesaurus)

Für die spätere inhaltliche Erschließung der Bilder sollen drei Thesauri angelegt werden.

Lektüre Lesen Sie zur Einführung in die Thematik die Abschnitte 2.5 bis 2.8 des [Kapitels 2](#). Bemühen Sie sich insbesondere um ein gutes Verständnis für die in Tabelle 2.4 (S. 54) zusammengefassten Arbeitsschritte für die Erstellung eines Thesaurus und die darin verwendeten Fachbegriffe.

Arbeitsaufwand: 16 Std.

5.1 Aspekt-differenzierte Thesauri

Das aspekt-differenzierte Thesaurus-Modell sieht für die inhaltliche Erschließung der Bilder drei Thesauri vor, die zuvor angelegt werden müssen. Dazu ist es zunächst erforderlich, einen Überblick über die Inhalte der Bilder zu bekommen und die Terminologie zu sammeln, die für die Erschließung der Bildinhalte benötigt wird. Die so gesammelte Terminologie ist dann auf die drei Aspekte zu verteilen und in normierter Form in den drei Thesauri zu erfassen.

Die drei für die Erschließung der Bildinhalte verwendeten Aspekte sind:

- Allgemeine Sachdeskriptoren für die abgebildeten Objekte;
- Geografische Deskriptoren für den Ort, an dem sich ein abgebildetes Objekt befindet, falls dieser ersichtlich ist;
- Individualnamen für die Eigennamen der Objekte, falls sie einen haben.

Für das Beispiel „Golden Gate Bridge“ könnten dann beispielsweise diese Deskriptoren geeignet sein:

- Hängebrücke (als allgemeiner Sachdeskriptor);
- San Francisco (als geografischer Deskriptor);
- Golden Gate Bridge (als Individualname).

Enthält ein Bild mehrere Objekte, können natürlich auch mehrere allgemeine Sachdeskriptoren und möglicherweise mehrere Individualnamen für eine angemessene Bilderschließung nötig sein (Orte sollte es auf einem Bild in der Regel nur einen geben).

5.2 Thesauruserstellung mit MidosThesaurus

Für die Bilddatenbank sind die drei benötigten Thesauri bereits angelegt. Sie befinden sich im Ordner der Datenbank und haben folgende Dateinamen:

- 020.mth – Thesaurus für Allgemeine Sachdeskriptoren;
- 025.mth – Thesaurus für Geografika;
- 030.mth – Thesaurus für Individualnamen.

Öffnen Sie mit dem Programm *MidosThesaurus* über „Thesaurus – Öffnen“ den Thesaurus für Allgemeine Sachdeskriptoren (Datei „020.mth“ im Ordner „bild-db“). Dass Sie nach dem Öffnen ein leeres graues Programmfenster sehen, ist kein Fehler! Der Thesaurus ist geöffnet, wenn in der Statuszeile des Programms die Daten zu diesem Thesaurus angezeigt werden.

In den Thesauruseditor von *MidosThesaurus* gelangen Sie entweder mit „Bearbeiten – Thesauruseditor“ oder mit dem „Füllhalter“-Symbol. Der *Thesauruseditor* hat ein dreigeteiltes Fenster, in dem sich rechts eine vollständige Liste aller im Thesaurus enthaltenen Terme befindet (Deskriptoren und Synonyme, alphabetisch sortiert), links die Hierarchie zum gerade aktiven Term und in der Mitte das eigentliche Eingabefeld für einen Datensatz im Thesaurus, dem sog. Deskriptorsatz. Dabei beziehen sich alle Einträge unterhalb von „Relationen“ immer auf den Eintrag im Feld „Deskriptor“. Eine Einführung in die Arbeit mit *MidosThesaurus* finden Sie in Abschnitt 8.4 des [Kapitels 8](#).

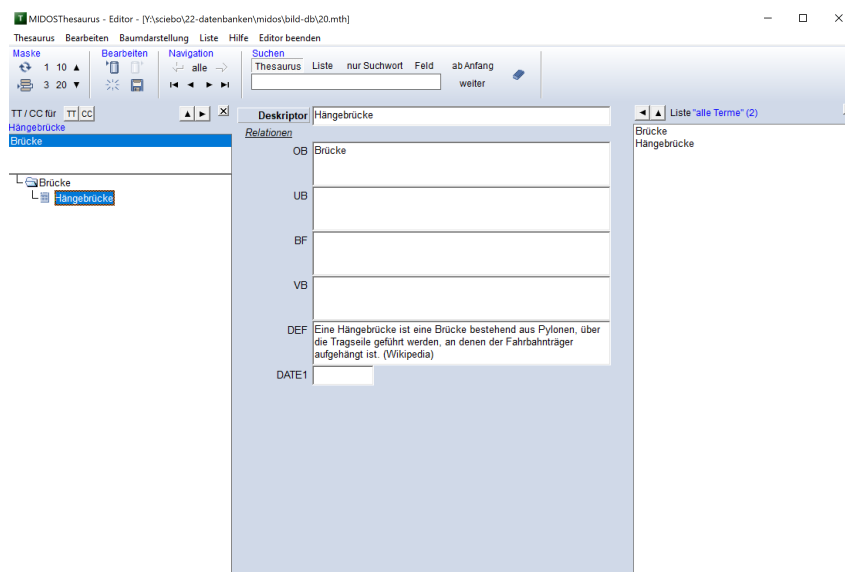


Abbildung 3: *MidosThesaurus* – Programmfenster

Erfassen Sie alle in der Inhaltsanalyse (vgl. Abschnitt 4) erkannten Objekte, Orte und Individualnamen als Deskriptoren in den drei entsprechenden Thesauri. Beachten Sie dabei die folgenden Hinweise:

- Ergänzen Sie den Thesaurus für „Allgemeine Sachdeskriptoren“ um Synonyme und einige Definitionen. Deklarieren Sie hierarchische Beziehungen für alle Deskriptoren und Verwandtschaftsbeziehungen für einige Deskriptoren.
- Erstellen Sie für die „Geografischen Deskriptoren“ eine einfache Hierarchie der Orte. Definitionen, Synonyme und Verwandte Begriffe sind hier verzichtbar.
- Erfassen Sie die Individualnamen in einer einfachen Liste, ohne eine Hierarchie. Legen Sie für die Individualnamen einheitliche Vorzugsbenennungen fest, z. B. Schreibweise von Personennamen in invertierter Form (Nachname, Vorname) oder Namen von Bauwerken in Landessprache (Tour Eiffel vs. Eiffelturm). Legen Sie für alternative Benennungen Synonyme an. Definitionen und Verwandte Begriffe sind verzichtbar.
- Bemühen Sie sich um Vollständigkeit, d. h. die drei Thesauri sollten alle Merkmale (Deskriptoren) enthalten, die für die Beschreibung aller Bildinhalte benötigt werden. Spätere Korrekturen und Ergänzungen an den Thesauri sind natürlich möglich, machen den eigentlichen Erschließungsprozess aber beschwerlicher, wenn sie zu häufig werden.

6 Inhaltliches Erschließen der Bilder durch Wortlisten und Thesauri (Midos 6, *MidosThesaurus*)

Die 50 Bilder der Bilddatenbank sollen auf der Grundlage eines aspektdifferenzierten Erschließungskonzepts mit Merkmalen aus Wortlisten und Deskriptoren aus drei Thesauri inhaltlich

erschlossen werden.

Arbeitsaufwand: 12 Std.

Lektüre Lesen Sie zur Einführung in die Thematik die Abschnitte 2.9.1 und 2.9.2 des [Kapitels 2](#).

6.1 Erschließen durch Wortlisten

Für bestimmte inhaltliche Erschließungsmerkmale von Bildern gibt es vergleichsweise wenige Möglichkeiten der Eingabe – vergleichbar manchen formalen Merkmalen wie etwa den Bildformaten. Derartige Merkmale könnten beispielsweise sein: Perspektive, Tageszeit, Wetter oder Stimmung des Bildes. Für solche Aspekte lohnt die Erstellung eines Thesaurus nicht, stattdessen sollten diese in Wortlisten erfasst werden.

Bestimmen Sie zwei inhaltliche Aspekte, für die Sie eine Erschließung mit Wortlisten vornehmen wollen.

Erweitern Sie die Datenbeschreibung der Bilddatenbank um zwei Felder für diese inhaltlichen Aspekte.

Erschließen Sie die 50 Bilder mit den Merkmalen der beiden ausgewählten Aspekte.

Beachten Sie: Nicht vorhandene Aspekte werden nicht erschlossen. Die Ermittlung inhaltlicher Merkmale basiert auf den erkennbar vorhandenen Bildinhalten, nicht auf vermuteten.

6.2 Erschließung durch Thesauri

Für die inhaltliche Erschließung der Bilder mit den selbst erstellten Thesauri ist es erforderlich, eine Verknüpfung zwischen der Bilddatenbank und den Thesauri herzustellen. Dazu ist zunächst die Datenbeschreibung der Bilddatenbank um Felder für die drei Thesauri zu erweitern, wobei der jeweilige Feldname durch die vorgegebenen Dateinamen der Thesauri bereits festgelegt ist.

Erweitern Sie die Datenbeschreibung der Bilddatenbank um die Felder 020 (für den Thesaurus für Allgemeine Sachdeskriptoren), 025 (für den geografischen Thesaurus) und 030 (für den Thesaurus mit Individualnamen).

Gehen Sie in den Datenbankeditor. Gehen Sie über „Optionen – Thesaurus“ in den Dialog „Einstellungen zur Thesaurusübernahme“ und aktivieren dort die Option „Verwende feldbezogene Thesaurusdateien“ (vgl. S. 80 des [Kapitels 2](#)). Bitte machen Sie keine weiteren Einträge in diesem Dialog.

Im *Datenbankeditor* von *Midos 6* gibt es nach der erfolgreichen Verknüpfung der Thesauri einen Button „Thesaurus“. Wenn man sich in der Erfassungsmaske mit dem Cursor in einem Thesaurusfeld befindet, führt dessen Aktivierung zum Aufruf des jeweiligen Thesaurus im *ThesaurusViewer*. Im *ThesaurusViewer* können nun die für das jeweilige Bild benötigten Deskriptoren ausgewählt und in die sog. *Sammelbox* übertragen werden. Über die Funktion „Export“ werden die Deskriptoren aus der *Sammelbox* in das Feld des *Datenbankeditors* übertragen (vgl. S. 80-82 des [Kapitels 2](#)).

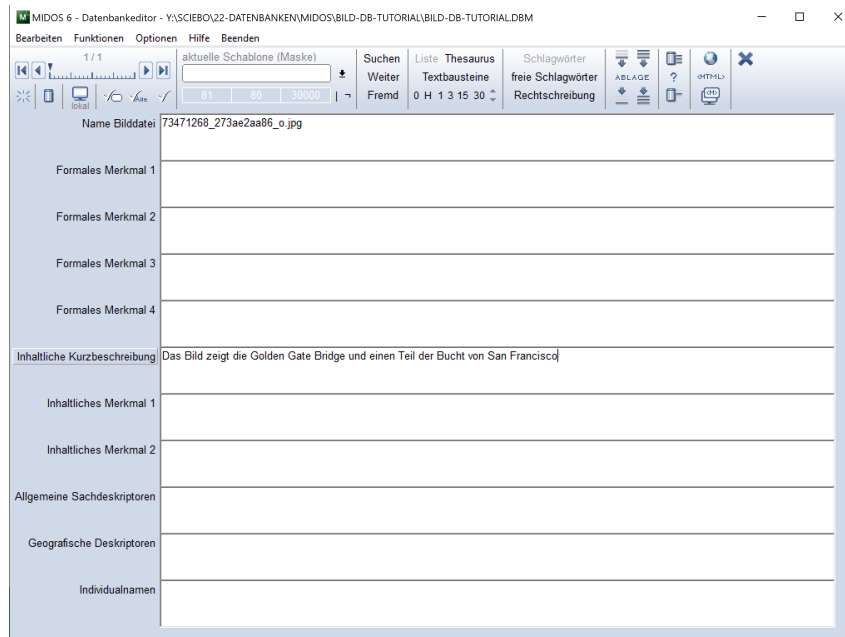


Abbildung 4: Midos 6 – Datenbankeditor – Thesaurusverknüpfung

Erschließen Sie die 50 Bilder mit den drei bereits mit der Datenbank verknüpften Thesauri.

Beachten Sie: Jedem Dokument ist mindestens ein *Allgemeiner Sachdeskriptor* aus dem Thesaurus „020.mth“ zuzuweisen. Orte und Individualnamen werden nur für einen Teil der Bilder von Bedeutung sein. Orientieren Sie sich bei der Zuteilung der Deskriptoren am Konzept der „koextensiven Erschließung“, d. h. teilen Sie jeweils nur die spezifischsten Deskriptoren zu: z. B. „Hängebrücke“ nicht „Bauwerk, Brücke, Hängebrücke“ bzw. „San Francisco“, nicht „Vereinigte Staaten von Amerika, Kalifornien, San Francisco“ (vgl. Folie 63 des [Vorlesungsskripts](#) und S. 75-76 des [Kapitels 2](#)).

6.3 Arbeitsergebnis

Eine Bilddatenbank mit 50 formal und inhaltlich erschlossenen Datensätzen.

7 Erstellen einer Retrievalanwendung

Für die Bilddatenbank soll eine sog. *Retrievalanwendung* erstellt werden. Dabei handelt es sich um ein eigenständiges Informationsprodukt, das auch auf Rechnern ausgeführt werden kann, auf denen kein *Midos 6* installiert ist. Die *Retrievalanwendung* erlaubt eine umfangreiche Konfiguration der Suchmöglichkeiten und – dies ist ein entscheidender Unterschied zu vergleichbaren Softwareprodukten – die Einbindung der Thesauri in die Suche.

Arbeitsaufwand: 12 Std.

Lektüre Lesen Sie zur Einführung in die Thematik den Abschnitt 2.10 des [Kapitels 2](#) und den Abschnitt 8.8 des [Kapitels 8](#). Die Funktionalitäten und Konfigurationsmöglichkeiten der Retrievalanwendung finden Sie mit zahlreichen Screenshots auf den Folien 73-86 des [Vorlesungsskripts](#).

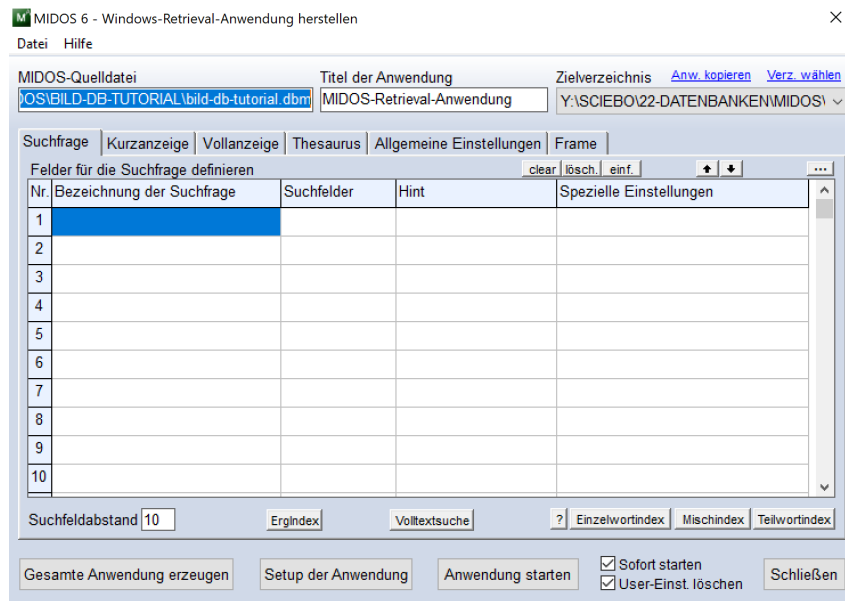


Abbildung 5: *Midos 6* – Dialog zum Erstellen einer Retrievalanwendung

Erstellen Sie mit *Midos 6* eine *Windows-Retrievalanwendung* mit folgenden Eigenschaften:

- Startseite mit Hilfestellungen zur Durchführung inhaltlicher Suchen (Beispiele, Thesaurussuche);
- Navigation über Frame am rechten oder linken Rand;
- Wortlisten im Frame;
- Suchmaske mit Volltextsuche und Suchen für die wichtigsten Merkmale;
- mouse-over-hints zu den einzelnen Suchmöglichkeiten;
- Einbindung aller Thesauri in die Suche;
- Ausgabeformat mit Bildanzeige und Verlinkung geeigneter Feldinhalte für die direkte Navigation in der Vollanzeige;
- ein weiteres Ausgabeformat als kategorisiertes Vollformat mit allen Feldern der Datenbank.

Durch das Erstellen der *Retrievalanwendung* mit „Gesamte Anwendung erzeugen“ werden alle für die Funktion der *Retrievalanwendung* benötigten Dateien in das Unterverzeichnis „m2retanw“ kopiert (inkl. Bilderordner, Thesauri und Startseite). Der Ordner „m2retanw“

kann auf andere Rechner kopiert werden und die *Retrievalanwendung* dort – ohne dass eine *Midos 6*-Installation auf diesem Rechner vorhanden ist – mit „start.exe“ gestartet werden.

Führen Sie mit der *Retrievalanwendung* Testrecherchen durch, um die Funktionalität der unterschiedlichen Suchen zu überprüfen, insb. die Funktionsweise der Suche über eine Thesaurusverknüpfung. Beispiele für derartige Recherchen könnten sein:

- Suche nach allen Bildern mit Giraffen;
- Suche nach allen Bildern mit Säugetieren;
- Suche nach allen Bildern mit Tieren;
- Suche nach Bildern mit dem Eiffelturm;
- Suche nach allen Bildern mit Türmen;
- Suche nach Bildern mit Motiven in Frankreich;
- etc.

7.1 Arbeitsergebnis

Eine *Retrievalanwendung*, die formale Suchen und durch die Einbindung der Thesauri unterschiedliche inhaltliche Suchen nach den Bildern der Datenbank ermöglicht.

8 Musteraufgabe für die Modulprüfung 1020

Die [Musteraufgabe](#) entspricht den Anforderungen der Modulprüfung 1020 (DIS 02). Die Inhalte weisen eine große Übereinstimmung mit dem Programm der praktischen Aufgabenstellung im Modul DIS 02 auf. Lediglich die zu bearbeitenden Bilder unterscheiden sich.

Nach Abschluss des Vorlesungsbetriebs (in der letzten Vorlesungswoche) werde ich die vollständige Aufgabenstellung für die Modulprüfung per Mail versenden. Ab diesem Zeitpunkt, d. h. ab Beginn des Bearbeitungszeitraums, ist es nicht mehr möglich, Fragen zur Modulprüfung zu stellen.