



DigiBib – Die Digitale Bibliothek

Arbeitsunterlagen

Allgemeine Hinweise zur Einbindung von Datenbanken in die Metasuche

Version: 1.0

Autor: Heiko Jansen

Zuletzt aktualisiert: 22.11.06

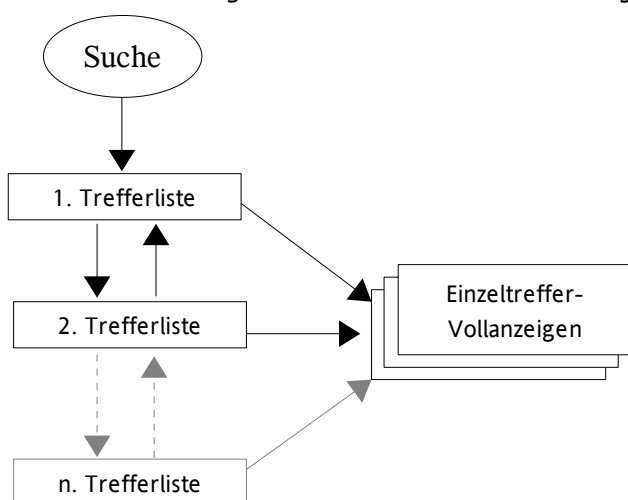


Inhaltsverzeichnis

1 Grundlegende Abläufe.....	2
2 Suche.....	2
3 Ergebnisformat.....	3
4 Sonstiges.....	4

1 Grundlegende Abläufe

Die Suche in der DigiBib wird durch einen mehrstufigen Prozess abgebildet.



- 5 Das Ergebnis einer Suche ist zunächst eine Trefferliste mit den wichtigsten Angaben zu den gefundenen Datensätzen (jeweils Autor/Körperschaft, Titel, Erscheinungsdatum, Volltext-Link).

Die Länge der Trefferlisten ist im Idealfall schon auf Seiten der Datenbank variabel. Notfalls holt aber die DigiBib solange weitere Trefferlisten von der durchsuchten Datenbank ab, bis die in der DigiBib spezifizierte Länge der Liste erreicht ist.

- 10 Von einer angezeigten Liste kann solange zur nächsten weitergeschaltet werden, bis alle Ergebnisse geholt wurden. Einmal geholte Listen werden von der DigiBib zwischengespeichert, so dass die Navigation zu vorherigen Listen keine Ansprache der Zieldatenbank erfordert.

Zu jedem Treffer in den Listen kann der DigiBib-Nutzer eine Vollanzeige aufrufen. I.d.R. wird dazu wieder eine Anfrage an die Zieldatenbank abgesetzt, um die vollständigen Daten zu holen. Einmal geholte Vollanzeigen werden zwischengespeichert.

- 15 Es ist möglich, dass eine Datenbank schon für die Listenanzeige alle verfügbaren Daten zu den Treffern liefert. In diesem Fall werden in der Liste nur ausgewählte Information angezeigt, die vollständigen Informationen werden aber zwischengespeichert und beim Aufruf der Langanzeige wird diese aus dem Zwischenspeicher bedient, ohne dass die Zieldatenbank erneut angesprochen wird.

2 Suche

- 20 Die Suchmaske der DigiBib verfügt derzeit über 9 Eingabefelder, die mit den Operatoren "UND", "ODER", "UND NICHT" verknüpft werden können:

- Titel
- Autor
- Körperschaft
- 25 ● Schlagwort
- Verlag
- ISBN
- ISSN
- Erscheinungsjahr

30 Zusätzlich wäre folgende Suchoptionen hilfreich:

- ◆ Suche nach Datensatz-ID mit dem betreffenden Datensatz als Ergebnis sowie - wenn derartige Strukturen in der Datenbank vorhanden sind -
- ◆ Suche nach Datensatz-ID mit allen diesem Datensatz untergeordneten Datensätzen als Ergebnis (also z.B. Suche nach ID einer Serie und als Ergebnis alle Stücktitel der Serie).

35 Übergeben werden pro Feld entweder eine Phrase oder ein oder mehrere Stichworte. Die Zieldatenbank sollte beides unterstützen.

Die DigiBib unterstützt außerdem eine Rechtstrunkierung (gesondert je Stichwort bzw. einmalig für die Phrase).

40 Falls die Datenbank erweiterte Funktionalitäten anbietet, die eine Obermenge der genannten Optionen darstellen (z.B. Suche nach Bereichen von Erscheinungsjahren), stellt dies i.d.R. kein Problem dar.

Nach Abschicken einer Suchanfrage über die DigiBib werden die Werte der o.g. Felder ausgelesen und in die Suchsyntax der jeweiligen Zieldatenbank übersetzt, z.B.:

45 http://www.ub.uni-blabla.de:8080/cgi-bin/suche.pl?base=Katalog&job=Freie+Suche&PER=Autor&PER_STI=AND&STI=Titel&STI_KOR=AND&...&start=1&num=10&page=1&Format=digibib

Dazu werden folgende Informationen benötigt:

- ◆ Über welche URL wird die Datenbank angesprochen? z.B.: <http://www.ub.uni-blabla.de:8080/cgi-bin/suche.pl?start=1&base=Katalog&job=Freie+Suche>
- 50 ◆ Welche Felder sind in der Datenbank suchbar und wie werden sie benannt? z.B.: Autor (PER), Titel (STI), Körperschaft (KOR) etc.
- ◆ Welche Verknüpfungsoperatoren existieren und welche Werte können diese annehmen? z.B.: Autor-Titel (STI_PER) etc. ; mögliche Werte: AND, OR, NOT
- ◆ Welche Parameter legen den Bereich der anzuzeigenden Treffer fest? z.B.: " &start=11&numbers=10 " für die Treffer 11 bis 20
- 55 ◆ Welches Zeichen wird für die (Rechts-) Trunkierung verwendet? z.B.: * oder ?

Die vorstehenden Daten sind als Hinweise zu verstehen. Die DigiBib ist nicht auf die Kommunikation via HTTP GET-Requests beschränkt. Genauso kommen Z39.50, SOAP, SRU usw. in Frage. Welche Schnittstelle jeweils gewählt wird, bedarf der Absprache zwischen Datenbank-Host und hbz.

60 3 Ergebnisformat

Die DigiBib verwendet ein eigenes, proprietäres Format (XML-basiert) für die interne Datenhaltung. Es ist möglich, dass von einer Datenbank bereits dieses Format geliefert wird. Dies bedarf zusätzlicher Absprachen.

In den meisten Fällen wird die Datenbank allerdings über ihr Such-Interface ein oder mehrere

- 65 anderweitig spezifizierte Ergebnisformate anbieten. Die DigiBib ist auch hier nicht auf bestimmte Formate festgelegt. Wichtig ist lediglich, dass die Formate maschinell vernünftig zu verarbeiten sind, d.h. unstrukturierte oder nur bedingt strukturierte (z.B. ISBD) Formate sind nicht nutzbar.
- Bevorzugt werden solche Formate, die in offiziellen Standards definiert sind (z.B. MAB, MARC).
- Da auf der Basis der Daten weitere Dienstleistungen aufbauen, sollten die Strukturen immer so atomar wie möglich ausgelegt sein. Z.B. sollten bei Artikeln Seitenzahl, Heftnummer und Jahrgang in getrennten Feldern stehen.
- 70 Ob ein Format geeignet ist, kann das hzb i.d.R. anhand kurzer Tests klären.
- Die DigiBib kann Daten in vielen verschiedenen Zeichensätzen verarbeiten. Optimal wäre allerdings die Lieferung der Ergebnisse in einem UNICODE-Format.
- 75 Neben den reinen Trefferdaten müssen die Ergebnisse für die Listen-Anzeigen jeweils um bestimmte Metadaten ergänzt werden:
- URL bzw. Aufrufparameter für die Vollanzeige des jeweiligen Datensatzes
 - URL bzw. Aufrufparameter für die Anzeige der folgenden Trefferliste
 - Anzahl der Treffer für die gesamte Suchanfrage
- 80 Wie bereits unter 1) beschrieben, können die Trefferlisten auch bereits die vollständigen Informationen zu den einzelnen Datensätzen enthalten. In diesem Fall entfällt der Aufruf der Zieldatenbank und die DigiBib füllt die interne Vollanzeige mit den Daten aus der Liste.
- Bei Datenbank-Anbindungen über HTTP ist die Verwendung von Cookies durch die Datenbank-Schnittstelle (für Session-Handling o.ä.) zu vermeiden.
- 85 Die Schnittstelle sollte sich insofern konsistent verhalten, als dass bei Suchanfragen mit nur einem Treffer ebenfalls eine Listenstruktur zurückgeliefert wird und nicht abweichend zum sonstigen Verhalten in diesen Fällen direkt eine Langanzeige generieren.

4 Sonstiges

- 90 Die Schnittstelle der Datenbank sollte entweder ohne ein Session-Konzept arbeiten oder eine notwendige Session (oder Anfrage-ID) implizit über eine eingehende neue Suchanfrage (ohne Session-ID/Anfrage-ID) erzeugen. Zusätzliche vorgeschaltete Requests zur Erzeugung einer Session sind zu vermeiden. Die DigiBib startet für jede neue Suchanfrage eine neue Session. Es ist nicht möglich, dass Sessions für mehrere Anfragen des gleichen (oder unterschiedlicher) DigiBib-Benutzer(s) weiterverwendet werden.
- 95 Sofern Sessions verwendet werden, wird die DigiBib i.d.R. die gleiche Session für das Navigieren in den Trefferlisten oder Vollanzeigen verwenden. Das Timeout der Sessions sollte auf Seiten der Datenbank daher nicht zu knapp eingestellt sein. Mindestens 5 – 10 Minuten sind sinnvoll.