

Tools zur semantischen Suchunterstützung in Datenbanken Entwicklungen beim Deutschen Bildungsserver und beim InfoWeb Weiterbildung

23.03.2011, BMBF, Bonn

Doris Hirschmann, DIPF

Suchunterstützungsformen

Anregung

IWWB-Suche beim Deutschen Bildungsserver – Anregungen zur Suche

<http://www.bildungsserver.de/zeigen.html?seite=3358>

Sortierung

Elixier – Kategorien zur Sortierung einer Treffermenge

<http://www.bildungsserver.de/elixier/>

Ergänzung

Zusätzliche Suchwortvorschläge bei geringer Treffermenge

<http://www.iwwb.de/>

Suchwortunterstützungstool beim IWWB

Bestandteile

Wesentlicher Bestandteil dieses Tool sind Vokabularien, d.h. Wortlisten die im weiteren und im engeren Sinne mit dem Thema Weiterbildung zu tun haben.

In diesem Fall sind das

- Das FIS-Schlagwörterbuch: Vokabular des Fachinformationssystems Bildung beim DIPF mit vielfachen, semantischen Zuführungen, das laufend gepflegt und aktualisiert wird
- IWWB Suchwörterliste (wird monatlich erfasst)
- Berufeliste der Bundesagentur für Arbeit

Die Listen müssen entsprechend aufbereitet (verknüpft) sein, damit ein Zusammenspiel und ein Abgleich möglich werden.

Suchwortunterstützungstool beim IWWB

Zusammenspiel der Vokabularien

- Abgleich mit einem eingegebenen Suchwort: Errechnung der Beziehungen in den Schlagwortfeldern: d.h. wie oft kommen prozentual welche Schlagwörter, in wie viel Datensätzen, in welcher Häufigkeit zueinander vor.
- Die dabei entstehenden Prozentzahlen werden als Empfehlung für das Ranking der zum Suchwort vorgeschlagenen Alternativsuchworte genommen bzw. dann dargestellt.

Suchwortunterstützungstool beim IWWB

Ablauf einer Abfrage

Ein Suchwort wird von einem Nutzer eingegeben und es werden zu diesem Wort weniger als 6 Kursangebote gefunden.

Um nun eine Empfehlung von weiteren Suchworten ausgeben zu können, wird das Suchwort mit dem Vokabular aus dem FIS und dem IWWB-Schlagwortbestand abgeglichen.

Wird im ersten Durchlauf nichts gefunden, gibt es einen weiteren Durchlauf, in dem die Endungen des Suchwortes modifiziert wurden (z.B. mit und ohne Pluralansetzung).

Findet man dann auch noch nichts, dann wird die BA-Berufeliste mit einbezogen und noch einmal abgeglichen.

Diese Abläufe geschehen aufgrund statistischer Berechnungen.

Beispiele

Suchwortvorschläge zu „windows 7“

- Partnerdatenbanken IWWB
- Service für Datenbankanbieter
- Über das IWWB

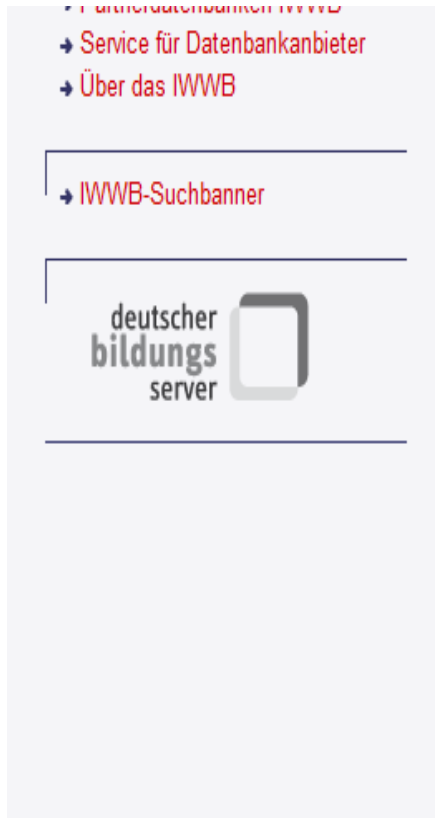
- IWWB-Suchbanner



"Informatik" "Computer" "Datenverarbeitung" "Software" "Computerprogramm"
"Computeranwendung" "Benutzeroberfläche" "Personal Computer"
"Textverarbeitung" "Handbuch" "Betriebssystem" "Grafische Darstellung"
"WORD" "Programm" "Nachschlagewerk" "Applikation" "Anwendung"
"Neue Technologien" "Praktische Informatik" "Praktische Anleitung" "Netzwerk"
"Programmierung" "Grundlagen" "Tabellenkalkulation" "Programmiersprache"
"Datenbanksystem" "Internet" "Informatikunterricht" "Datenübertragung"
"Programmbedienung" "MS-DOS" "Dateiverwaltung" "Beispiel" "Zeichnen"
"Winword" "Vernetztes System" "TURBO-PASCAL" "Nachrichtentechnik"
"Listing" "Kommunikationstechnik" "Informationstechnik" "Excel" "Bild"
"Angewandte Informatik" "Analyse" "Übung" "WORKS" "Tabellenverarbeitung"
"Statistik" "SPSS"


Beispiele

Suchwortvorschläge zu „business english“



→ Service für Datenbankanbieter
→ Über das IWWB

→ IWWB-Suchbanner

deutscher
bildungs
server 

Vielleicht bringt auch eine Suche mit den folgenden "ähnlichen" Wörtern mehr relevante Einträge:

"Management" "Fremdsprachenunterricht" "medizinisch integrierter" "ambulant"
"Verwaltung" "Versorgung"
"Management ambulanter und medizinisch integrierter Versorgung" "International"
"Business Administration" "Baltisch" "Bachelor of Business Management"
"Bachelor of Business Administration Baltisch" "Bachelor of Business Administration"
"Administration" "B1" "A2" "Fortgeschrittene" "A1" "Landeskunde"
"Konversation" "B2" "Anfänger/innen" "Vereinigte Staaten" "USA"
"Wortschatz" "Methodik" "Literatur" "Hochschule" "Grammatik" "Französisch"
"Englischunterricht" "Deutsch" "4.6" "Zertifikat" "Vereinigtes Königreich"
"Textarbeit" "Schreiben" "Motivation" "Lesen" "Kommunikative Kompetenz"
"Hörverständnis" "Großbritannien" "Fremdsprache" "Fortgeschrittenenunterricht"
"Fachsprache" "Erfahrungsbericht" "Didaktik" "Arbeitsuchende" "Übung"
"Wortschatzarbeit"

Forschungsinteresse des DIPF

„...Informationsangebote und Arbeitswerkzeuge bedarfsgerecht zur Verfügung stellen...“

- **Transfer:**
Erkenntnisse und Ergebnisse aus der informationswissenschaftlichen Arbeit am DIPF und beim Deutschen Bildungsserver stehen dem IWWB zur Verfügung.
- **Erprobung:**
Planung von Forschungsvorhaben zur Weiterentwicklung von Informationssystemen, die im Bereich der Weiterbildungsdatenbanken erprobt werden könnten.

Optimierungswege für das Suchworttool

- I. Weiterentwicklung des schon bestehenden statistischen Verfahrens durch Erstellung weiterer Suchwortlisten mit anderen Kriterien und Erprobungen verschiedener Verknüpfungen und Kombinationen, um das Tool weiter zu optimieren.

- II. Erweiterung des Tools um ein thesaurusgestütztes System.
Entwicklung eines gemeinsamen, nach inhaltlichen Kriterien entwickelten Vokabulars oder einer Ontologie speziell für Weiterbildungsdatenbankbetreiber und für Kursanbieter.

Schritte zur Weiterentwicklung des Tools

I. Veränderungen in der „Mixtur“

- Generierung und Aufbereitung weiterer Vokabularien
- Erprobungen verschiedener Verknüpfungen der Vokabularien nach Häufigkeit ihres Vorkommens und nach unterschiedlichen Eigenschaften der Verknüpfung (siehe Folie 8 Carola Carstens, 24.3.2010:
http://projekt.iwwb-files.de/Steuerungsgremium/SemWeb_Carstens.pdf)
- Erprobungen von Veränderungen im Ablaufverfahren

II. Veränderung der „Zutaten“

- Entwicklung von Ontologien, d.h. Vokabularien werden zusätzlich hierarchisch aufgebaut. Das bedeutet man weist den Wörtern nicht nur statistisch eine bestimmte Bedeutung zu (a kommt x oft vor mit b zusammen), sondern auch eine Bedeutung nach Wichtigkeit (kommt vor oder nach x) (Folie 8 + 9, Carola Carstens, 24.3.2011)

Erprobung verschiedener Einsatzmöglichkeiten

Die Vokabularien können sehr unterschiedlich eingesetzt werden

- Suchwortvorschläge (wie schon vorhanden) – präzisieren, ausbauen
- zur Clusterung/Sortierung von Suchergebnissen (Elixier)
- zur Anreicherung von Eingabefeldern für Kursanbieter (Vereinheitlichung der Kursdaten)
- zur Einrichtung von Autocompleteverfahren bei der Begriffseingabe

Dankeschön

Fragen zum Tool beim IWWB:

Wolfgang Plum, E-Mail: wp@iwwb.de

Christian Richter, E-Mail: richter@dipf.de

Conni Poppe, E-Mail: poppe@dipf.de

Fragen zu den Weiterentwicklungsideen des DIPF

Prof. Dr. Marc Rittberger, E-Mail: rittberger@dipf.de

Carola Carstens, E-Mail: carstens@dipf.de

Doris Hirschmann, E-Mail: hirschmann@dipf.de