

# Datenübernahme in Contents-Linking

## Einleitung

Contents-Linking ist Teil des Multisuchsystems E-Connect<sup>1</sup>. Es ist ebenso wie ASEZA ein Suchsystem für elektronische Zeitschriften, darüber hinaus und vor allem ein Such- und Verwaltungssystem für Aufsatzzitate. Beide Systeme zeichnen sich durch folgende Besonderheiten aus:

1. Der Quellcode der Systeme besteht ausschliesslich aus HTML und JavaScript.
2. Der gesamte Datenbankinhalt ist in Webformularen untergebracht, befindet sich also im Computer des Benutzers. Die Rechercheaktionen laufen deshalb besonders schnell ab.
3. Bemerkenswert ist die aussergewöhnlich leichte und schnelle Übernahme von Daten, sowohl von Aufsatz- als auch von Zeitschriftentiteln.

Die Datenübernahme von Zeitschriftentiteln erstreckt sich auf beide Systeme. Contents-Linking greift auf die gleiche Zeitschriftendatenbank zurück wie ASEZA. Die Übernahme von Aufsatzzitaten ist auf das Contents-Linking-System beschränkt, und dieses wird hier vorgestellt.

Die beiden Datenbanken von Contents-Linking für Zeitschriftentitel und für Aufsatzzitate können völlig unabhängig voneinander aufgebaut, verwaltet und abgefragt werden. Ihr Aufbau geschieht im Prinzip durch das gleiche Einarbeitungsverfahren, wobei jeweils vollständige (!) Webseiten von einer Internetquelle in ein Formularfeld eingefügt werden, was, wie gesagt, aussergewöhnlich leicht und effektiv geschieht.

Abb. 1 und 2 zeigen dieses Verfahren für die Übernahme von Aufsatzzitaten, Abb. 3 und 4 für die Übernahme von Zeitschriftentiteln.

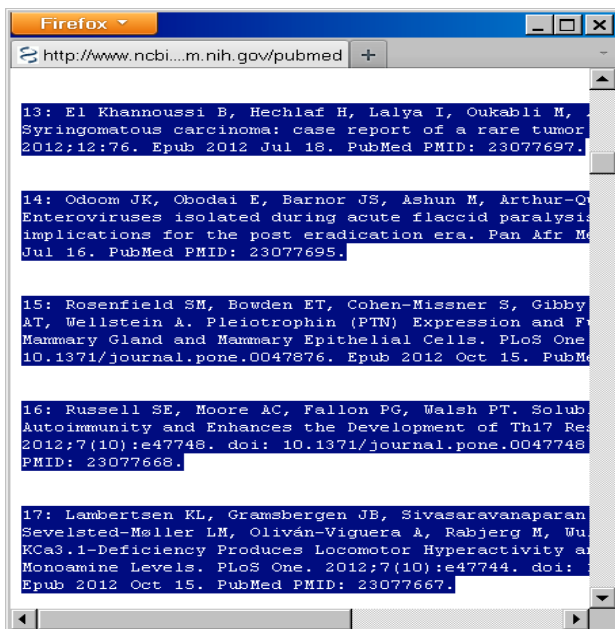


Abb. 1 PubMed Zitate im Textformat

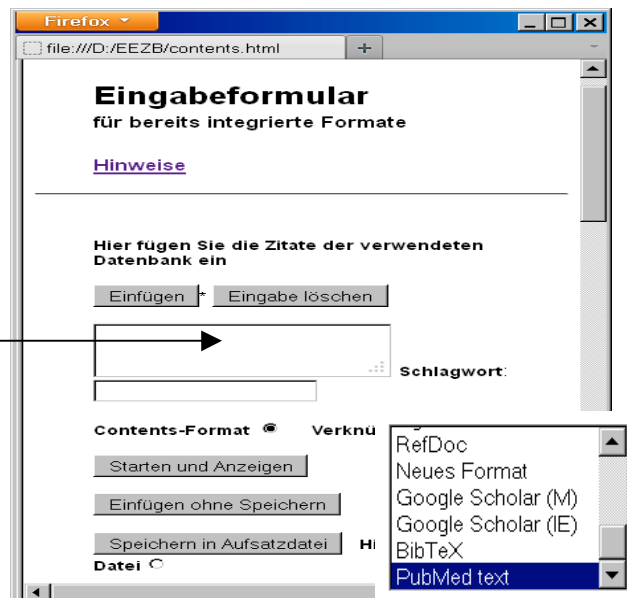


Abb. 2 Eingabe-Formular mit Menü

<sup>1</sup> <http://www.multisuchsystem.de/>

Contents-Linking vereinigt die bisher im Multisuchsystem entwickelten Systeme in einem einzigen System!

Das Eingabeformular enthält ein Auswahlménü, das hier separat abgebildet ist. Aus diesem muss stets die verwendete Datenbank und ggf. auch das Format ausgewählt werden. Hier ist *PubMed text* bereits voreingestellt.

Über den Schalter *Starten und Anzeigen* werden die eingefügten Zitate in das spezielle Contents-Linking-Format umgewandelt und sofort zur Kontrolle in einem Fenster angezeigt. Über den Schalter darunter werden die Zitate in das System temporär übernommen. In einer lokal installierten Version können die Zitate über den Schalter *Speichern in Aufsatzdatei* fest eingespeichert werden und zwar wahlweise als neue Datei oder als Hinzufügung zur bereits bestehenden Datei.

Beachtenswert ist, dass gleichzeitig auch mehrere Tausend Zitate (!) von PubMed auf diese Weise übernommen werden können und dass ähnlich effektiv über dasselbe Formular auch die Zitate einer Reihe anderer Fachdatenbanken erfasst werden können. Bei einigen Datenbanken muss dabei auch der verwendete Browser (IE oder Mozilla) berücksichtigt werden. Darüber hinaus ist eine Option für die Erfassung eines neuen Formates bzw. Datenbank eingerichtet. Diese basiert auf einem Erfassungsschema, das sich bereits bei der Erstellung der Abfrage-Codes für die in Contents-Linking integrierten Datenbanken bestens bewährt hat.

## Übernahme der Daten von Zeitschriftentitellisten

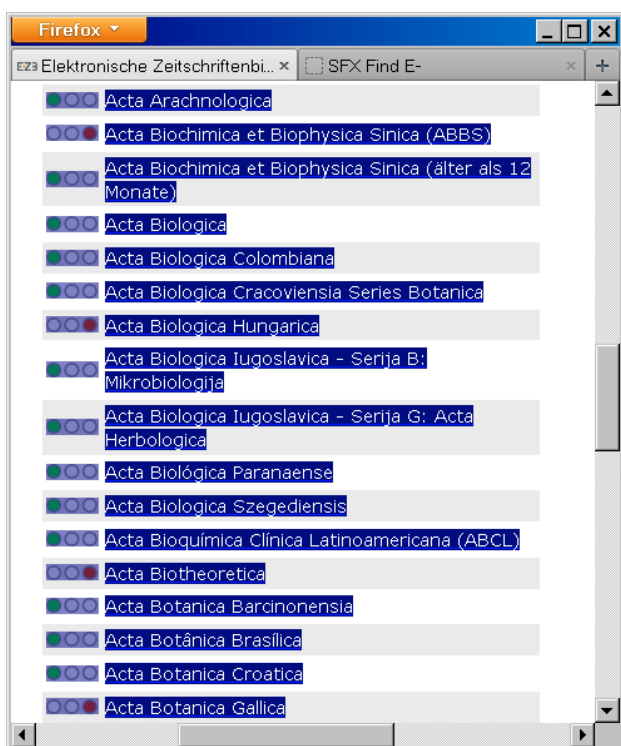


Abb. 3. EZB-Zeitschriftenliste markiert



Abb. 4. Einarbeitungsprogramm

Wie das obige Beispiel zeigt, geschieht die Übernahme von Zeitschriftentiteldaten in der gleichen Weise wie die Übernahme von Aufsatzzitaten und zwar ebenso leicht und effektiv.

Abb. 3 zeigt links eine markierte Zeitschriftenliste der EZB, die durch Copy und Paste in das Formularfeld des Einarbeitungsprogramms rechts eingefügt wird. Auch hier ist es bemerkenswert, dass die EZB-Liste bis zu mehreren Tausend Titel enthalten, also z.B. die Titel eines gesamten Fachgebietes auf einmal umfassen oder, was noch mehr ist, den gesamten Titelbestand (!) einer mittelgrossen Bibliothek enthalten kann, den das Programm auch auf einmal zu erfassen und zu bearbeiten vermag. Das Einarbeitungsprogramm ist zwar für die Übernahme von EZB-Listen optimiert, kann aber, nach entsprechender Vorbereitung, auch Listen aus anderen Quellen

übernehmen.<sup>2</sup> Das Einarbeitungsprogramm arbeitet vollständig nur mit dem IE. Für Mozilla Firefox kann auf eine Demo-Version zugegriffen werden.

In der Internetversion wird nur eine Demonstration des Einarbeitungsvorgangs angeboten, während das Einspeichern dieser Liste in das Datenbanksystem nur in einer lokal installierten Version mit dem IE möglich ist. Die Übernahme der Titel geschieht automatisch durch das Drücken des Start- und der beiden Fortsetzungsschalter, wobei jeweils Kontrollmeldungen eingeblendet werden. Vor der endgültigen Speicherfunktion wird die gesamte (!) Zeitschriftendatenbank mit allen Neueinträgen und Ergänzungen angezeigt. Dann erscheint in einer lokalen Version der Warnhinweis auf die ActiveX-Funktion, der zu akzeptieren ist. In der Internetversion dagegen wird das Programm mit einer (nichtsichtbaren) Fehlermeldung beendet.

Zeitschriftentitel, Titel mit Schlagwortangaben und Zeitschriftentitel mit Bestandsangaben werden auf die gleiche Weise übernommen. Entsprechend müssen vorher die Felder für das Bibliothekskennzeichen und/oder für das Thema ausgefüllt werden. In Abb. 4 ist z.B. als Kennzeichen *Frei* für frei verfügbare Titel und *chem* für das Fachgebiet Chemie eingetragen.

Die folgende Abbildung zeigt, wie frei verfügbare Titel für das Fach Chemie in der EZB aufgeführt werden.

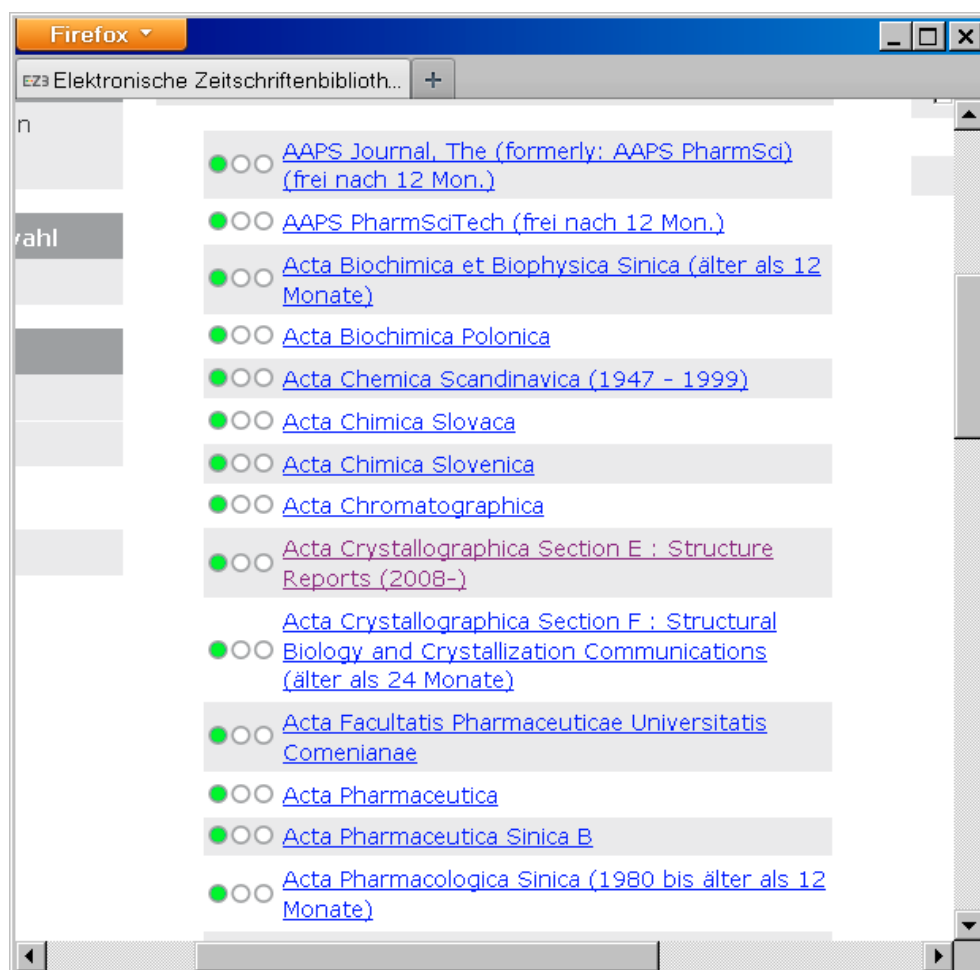


Abb.5. EZB-Zeitschriftentitelanzeige für frei zugängliche Titel der Chemie

<sup>2</sup> z.B. Thomson Reuters [http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jloptions.cgi?PC=master\\_T](http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jloptions.cgi?PC=master_T)  
Harvard Libraries [http://sfx.hul.harvard.edu/sfx\\_local/az/](http://sfx.hul.harvard.edu/sfx_local/az/)

Die folgende Auflistung zeigt, wie die oben eingegebenen Titel mit den zusätzlichen Angaben zur Verfügbarkeit und zum Thema in die Datenbank integriert werden. Zugleich wird die Struktur erkennbar, mit der die Titel in die Zeitschriftendatenbank eingeordnet sind.

So folgt nach dem Trennungszeichen > der Zeitschriftentitel in Grossbuchstaben, abgeschlossen durch =, danach folgen die Schlagwortnotationen (=pharma=) in Kleinbuchstaben, eingeschlossen durch Gleichheitszeichen. Falls vorhanden, folgt die ISSN und ggf. ein Verlagshinweis. Schliesslich, eingeleitet durch Sternchen, folgen die Notationen der Bibliotheken, die die betreffende Zeitschrift elektronisch abonniert haben oder der Hinweis auf die freie Verfügbarkeit, evtl. mit zeitlichen Beschränkungen.

```
>AAPS J=pharma=chem=pharm=15507416=SPRING*TIB*TUBS*TUCB*TUDA*FREI(NACH 12 MON.)!!*G
>AAPS
PHARMSCITECH=pharma=chem=pharm=15309932=SPRING*G*JE*TIB*TUBS*TUCB*TUDA*ULBH*FREI(
NACH 12 MON.)!!
.....
>ACTA BIOCHIMICA ET BIOPHYSICA
SINICA=bioch=chem=phys=physc=biol=16729145*AA*BHU*ERF*G*JE*RE*ROS*SLUB*SULB*TIB*TUBS*TU
CB*TUDA*TUM*UBPO*ULBH*UBEN*FREI(ALTER ALS 12 MONATE)!!
>ACTA BIOCHIMICA POLONICA=bioch=biol=chem=0001527X*FREI!!
.....
>ACTA CHEMICA SCANDINAVICA=chem=*FREI(1947-1999)!!
.....
>ACTA CHIMICA SLOVACA=chem=*FREI!!
>ACTA CHIMICA SLOVENICA=chem=13180207*SLUB*SULB*FREI!
.....
>ACTA CHROMATOGRAPHICA=chemana=chem=12332356*FREI!
.....
>ACTA CRYSTALLOGRAPHICA SECTION E=crys=geo=chem=phys=16005368=ING(2004-2004)*AA(2001-
2007)*AA(2004-2004)*BHU(2001-2007)*BTU(2001-2007)*FHM(2001-2007)*FHR(2001-2007)*G(2001-
2007)*RE(2001-2007)*ROS(2001-2007)*TIB(2001-2007)*TUBS(2001-2007)*TUCB(2001-2007)*TUDA(2001-
2007)*TUM(2001-2007)*ULBH(2001-2007)*FREI(2008-)!*UBEN(2001-2007)
>ACTA CRYSTALLOGRAPHICA SECTION
F=crys=geo=phys=bioc=biol=chem=crys*BHU*BTU*FHM*G*JE*RE*SLUB*SULB*TIB*TUBS*TUDA*TUM*UB
PO*ULBH*FREI(ALTER ALS 24 MONATE)!!*UBEN
.....
>ACTA FACULTATIS PHARMACEUTICAE UNIVERSITATIS COMENIANAE=pharma=chem=*FREI+
.....
>ACTA PHARMACEUTICA SINICA B=pharma=chem=??*FREI+
>ACTA PHARMACEUTICA=pharma=chem=pharm=13300075 *FREI!
.....
>ACTA PHARMACOLOGICA SINICA=pharmak=pharma=pharm=chem=med=16714083*BHU*G*JE(2005-
)*RE*SLUB(2005-)*SULB(2005-)*TIB*TUBS*TUCB*TUDA*TUM(2005-)*UBPO(2005-)*ULBH(2005-)*FREI(1980
BIS ALTER ALS 12 MONATE)!!*AA*UBEN
```

++ = neuer Titel; + = neuer Standort; !! = Angaben ergänzt oder Zeitvermerk geändert; ! = Angaben bestätigt

Dieser Ausschnitt zeigt, wie die einzelnen Einträge durch zusätzliche Zeichen gekennzeichnet werden. Die Zeichen + und ! zeigen neue Titel bzw. neue oder veränderte Angaben an. So wurde bei einigen Titeln die Zeitbeschränkung ergänzt oder verändert, was durch ein doppeltes Ausrufungszeichen markiert ist. Fragezeichen weisen auf gleichlautende Titel hin, ein spezielles Problem, das weiter unten näher erläutert wird. Alles in allem ein Hinweis auf die ungewöhnliche, eigentlich nicht zu erwartende Genauigkeit bei dieser Art von Datenübernahme, die auf Metadaten wie die ISSN verzichtet und sich auf die Erfassung der auf der Weboberfläche angezeigten Daten beschränkt.

Bei der Datenübernahme werden also, wie aus der Auflistung zu ersehen ist, Titel in der Schreibweise verändert, Zusätze zum Titel und interne Hinweise weggelassen, zeitliche Angaben zur Verfügbarkeit dagegen in meist abgeänderter Form übernommen.

Wie das Beispiel zeigt, sind die einzelnen Titel der Datenbank mit äußerst knappen Daten ausgeschrieben. Die Datenbank kann mit geringem Aufwand, im Prinzip ohne Metadaten, aufgebaut, erweitert und verwaltet werden. Als einzige Metadaten werden die ISSN verwendet, die aber lediglich einen geringen Vorteil für die interne Verwaltung des Titelmaterials bieten, um z.B. doppelte Erfassungen zu vermeiden. Für die Recherche selbst und für die Verknüpfungen mit Google Scholar und Google sind diese aber nicht verwendbar. Diese werden lediglich über die vorliegenden Titel der Zeitschriften hergestellt. Auch die Verknüpfungen mit der EZB und der ZDB können problemlos über die Titel erfolgen, wenn die ISSN nicht gegeben ist. Nur bei Titeln, die zu viele Treffer liefern, kann dies von Nachteil sein.

Im Prinzip könnte die Datenbank also aus den vorliegenden Zeitschriftentiteln ohne eine identifizierende ISSN aufgebaut werden, natürlich ohne die nicht zum originalen Titel gehörenden zusätzlichen Vermerke der EZB oder herausgebenden Institution. Die Verknüpfung mit Google Scholar erfordert aber eine gewisse Normierung, da Google Scholar die Titel der Zeitschriften normalerweise ohne die auf bestimmte Zeichen folgenden ergänzenden Zusätze erfasst und ebenso den hinter der Serienkennung folgenden Zusatz weglässt. Beispiele:

```
>PALLAS: EINE ZEITSCHRIFT FÜR STAATS- UND KRIEGS-KUNST
>HUAFEI-GONGYE = J OF THE CHEMICAL FERTILIZER INDUSTRY
>HYLE - AN INTERNATIONAL J FOR THE PHILOSOPHY OF CHEMISTRY
>ARCHIVE OF APPLIED MECHANICS / INGENIEUR ARCHIV

>ENVIRONMENTAL POLLUTION SERIES A: ECOLOGICAL AND BIOLOGICAL
=ENVIRONMENTAL POLLUTION SERIES A
>ENVIRONMENTAL POLLUTION SERIES B: CHEMICAL AND PHYSICAL
=ENVIRONMENTAL POLLUTION SERIES B
```

In diesen Beispielen werden die hinter den trennenden Zeichen . : / – stehenden Zusätze weggelassen. Es gibt aber Fälle, wo die hinter Zeichen stehenden Ergänzungen wesentliche, d.h. unterscheidende Zusätze sind und wo diese nicht weggelassen werden können, wobei auch Google Scholar in diesem Fall die Zusätze, aber ohne trennende Zeichen, erfasst und anzeigt:

```
>ADVANCES IN HUMAN FACTORS/ERGONOMICS
= ADVANCES IN HUMAN FACTORS ERGONOMICS
>ARCHIVES OF DISEASE IN CHILDHOOD / EDUCATION AND PRACTICE
= ARCHIVES OF DISEASE IN CHILDHOOD EDUCATION AND PRACTICE
>ARCHIVES OF DISEASE IN CHILDHOOD / FETAL AND NEONATAL
= ARCHIVES OF DISEASE IN CHILDHOOD FETAL AND NEONATAL
>JAHRESBERICHT / UMWELTBUNDESAMT
= JAHRESBERICHT UMWELTBUNDESAMT
>CURRENT MEDICINAL CHEMISTRY. ANTI-CANCER AGENTS
= CURRENT MEDICINAL CHEMISTRY ANTI-CANCER AGENTS
>CURRENT MEDICINAL CHEMISTRY. ANTI-INFECTIVE AGENTS
= CURRENT MEDICINAL CHEMISTRY ANTI-INFECTIVE AGENTS
```

### **Angleichung der Schreibweise durch das Einarbeitungsprogramm**

Die in der Zeitschriftentiteldatenbank durchgeführte Schreibweise entspricht im wesentlichen der der EZB mit einigen Vereinfachungen. Diese wiederum richten sich vielfach nach der Schreibweise der *Master Journal List* von Thomson Reuters<sup>3</sup>. Die Datenübernahme zielt

<sup>3</sup> <http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jlresults.cgi?PC=MASTER>

schliesslich und vor allem darauf, die Zeitschriftentitel in Contents-Linking so zu übernehmen, dass sie mit Google Scholar verknüpft werden können. Eine andere Art der Verknüpfung, etwa über die ISSN, ist, wie schon gesagt, hier nicht möglich.

Im Folgenden wird die durch das Einarbeitungsprogramm erreichte Angleichung gezeigt. Auf die in der Schreibweise der EZB-Liste aufgeführten eingegebenen Titel folgen jeweils die durch das Einarbeitungsprogramm ausgeschriebenen Titel in der Schreibweise der Datenbank, unter Beachtung der o.g. Ausnahmen, bei denen Zusätze nicht weggelassen werden dürfen:

In runden und eckigen Klammern Stehendes wird entfernt:

ACM TRANSACTIONS ON DATABASE SYSTEMS (ACM TODS)  
ACM TRANSACTIONS ON DATABASE SYSTEMS  
J OF SYNCHROTRON RADIATION [SYNCHROTRON RADIATION ONLINE]  
J OF SYNCHROTRON RADIATION

Auf Doppelpunkt, Bindestrich oder Schrägstrich folgende Zusätze zum Titel werden entfernt:

METABOLISM - CLINICAL AND EXPERIMENTAL  
METABOLISM 00260495

BJU INTERNATIONAL : BRITISH J OF UROLOGY  
BJU INTERNATIONAL 14644096

ARCHIVE OF APPLIED MECHANICS / INGENIEUR ARCHIV  
ARCHIVE OF APPLIED MECHANICS 09391533

Serienvermerke werden im allgemeinen ohne Zusätze nach den aufzählenden Zeichen oder Buchstaben (I,II..., A, B C ...) abgeschlossen, die folgenden Zusätze weggelassen. *SERIES*, *SECTION*, *PART* wird im allgemeinen aber beibehalten. Übersetzte Titel (z.B. die russischen nach C/C..) werden entfernt. Umlaute und sprachbezogene Zeichen werden beibehalten, müssen für die Google Scholar-Verknüpfung allerdings umgewandelt werden.

### **Gleichlautende Titel und gleichlautende Titel mit unterschiedlichen Zusätzen**

Gleichlautende Titel und gleichlautende Titel mit unterschiedlichen Zusätzen stellen ein nicht geringes Problem dar. Den in der Bestandsliste enthaltenen gleichlautenden Titeln können falsche ISSN und falsche fachliche Angaben zugewiesen werden, Fehler, die beim Einarbeiten von gleichlautenden Titeln ohne ISSN entstehen. Um dies zu vermeiden, berücksichtigt das Einarbeitungsprogramm dieses Problem, indem es alle erfassten Titel mit einer Liste gleichlautender Titel vergleicht und identifizierte Titel mit einem ? kennzeichnet. Das Einarbeitungsprogramm zeigt diese Titel über einen Schalter in einem separaten Fenster. Ebenfalls zur Kontrolle zeigt das Einarbeitungsprogramm alle insgesamt in der Datenbank gefundenen und nicht gefundenen einzuarbeitenden Titel an.

Viel häufiger als gleichlautende Titel im strengen Sinn treten serienartige Titel auf oder gleichlautende Titel, bei denen nur die Zusätze zum Titel voneinander abweichen. Gleichlautende Titel mit unterschiedlichen Zusätzen sind relativ zahlreich. Oben wurden einige Beispiele für solche Titel aufgeführt. Ausserdem gibt es gleichlautende Titel, bei denen nur jeweils der eine Titel mit einem Zusatz erscheint. Ein Beispiel dafür:

In der EZB wird neben *Journal of Hydrology* auch *Journal of Hydrology (New Zealand)* aufgeführt, das frei verfügbar ist. Bei der Einarbeitung wurde der Zusatz des zweiten Titels in Paranthese fälschlicherweise entfernt, fälschlich, weil er dadurch nicht mehr vom ersten zu

unterschieden ist. Das Einarbeitungsprogramm hat stattdessen beide Titel in dem einen Titel zusammengefasst und dabei auch die freie Verfügbarkeit zu den Bestandsangaben hinzugefügt.

Zur Behebung dieser Fehler habe ich ein Programm entwickelt, das die gesamte Titelliste der EZB auf solche gleichlautenden Fälle sowie auf alle serien- und serienähnlichen Titel untersucht.<sup>4</sup> In diesem Programm werden die dieser Schreibweise entsprechenden Titel ausgeschrieben. Sie können dann in eine Datei des Einarbeitungsprogramms integriert und bei jeder Einarbeitungs-prozedur berücksichtigt werden.

Der folgende Ausschnitt ist dem 2. Teil dieser Datei (such1g.htm) entnommen, der serienartige Titel bzw. Titel mit unterschiedlichem oder keinem Zusatz enthält:

```
>BOLETÍN / CENTRO DE ESTUDIOS MONETARIOS LATINOAMERICANOS=  
>BOLETÍN / INSTITUTO ESPAÑOL DE OCEANOGRAFÍA=  
>BOLETÍN / OBSERVATORIO DE LOS DERECHOS HUMANOS EN COLOMBIA=  
>BOLETÍN / REAL SOCIEDAD GEOGRÁFICA=  
>BOLETÍN ESTADÍSTICO / BANCO CENTRAL DE BOLIVIA=  
>BOLETÍN ESTADÍSTICO / BANCO CENTRAL DE HONDURAS SUBGERENCIA DE ESTUDIOS  
ECONÓMICOS=  
.....  
>J OF HYDROLOGY NEW ZEALAND=  
>J OF HYDROLOGY=  
>J OF INFORMATION DISPLAY JID=  
>J OF INFORMATION DISPLAY=
```

Die beiden letzten Titel weisen noch auf ein besonderes Problem hin: Enthält der Zusatz zum Titel lediglich die Abkürzung, ist dieser natürlich überflüssig und kann auch nicht in den Suchstring von Google Scholar eingesetzt werden. Google Scholar liefert bei diesen Zusätzen kein Ergebnis. Im Unterschied allerdings zu Google, das hierbei und auch bei sonstigen Abweichungen fast immer ein Ergebnis liefert. Google ist übrigens auch bei Verknüpfungen mit dem Sachtitel eines Zitats öfter erfolgreich als Google Scholar, ist aber bei Verknüpfungen mit dem reinen Zeitschriftentitel weniger vorzuziehen, weil Google hier nur auf die Homepage und nicht wie Google Scholar sofort auf den Inhalt einer Zeitschrift hinführt.

Die folgende Auflistung zeigt, wie die oben aufgelisteten EZB-Titel in Contents-Linking schliesslich mit Hyperlinks angezeigt werden, wobei der Zeitschriftentitel jeweils mit der EZB verknüpft ist.

```
AAPS J \(Google Scholar\)  
AAPS PHARMSCITECH \(Google Scholar\)  
.....  
ACTA BIOCHIMICA ET BIOPHYSICA SINICA \(Google Scholar\)  
ACTA BIOCHIMICA POLONICA \(Google Scholar\)  
ACTA CHEMICA SCANDINAVICA \(Google Scholar\)  
ACTA CHIMICA SLOVACA \(Google Scholar\)  
ACTA CHIMICA SLOVENICA \(Google Scholar\)  
ACTA CHROMATOGRAPHICA \(Google Scholar\)  
ACTA CRYSTALLOGRAPHICA SECTION A \(Google Scholar\)  
ACTA CRYSTALLOGRAPHICA SECTION B \(Google Scholar\)  
ACTA CRYSTALLOGRAPHICA SECTION C \(Google Scholar\)
```

<sup>4</sup> [BearbEZB4.htm](#) (Die hier erwähnten Bearbeitungsdateien sind nur ein kleiner Teil der insgesamt für das System entwickelten Dateien.)

## Übernahme der Daten von Aufsatzzitaten

Die Datenbank der Aufsatzzitate ist im Laufe der Zeit aus sehr verschiedenen Datenbanken und nicht ganz einheitlich aufgebaut worden. So wurden mehrere Tausend Zitate in der Reihenfolge Zeitschrift-Verfasser-Sachtitel erfasst, was vor allem bei Google Scholar zu erfolglosen Verknüpfungen führen kann. Erst allmählich hat sich das jetzt gültige Erfassungsschema für die Übernahme der Zitate entwickelt.<sup>5</sup> Solche Unstimmigkeiten und Ungenauigkeiten können in einer lokal installierten Version vermieden werden. Die Datenbank der Aufsatzzitate ist in der Internetversion ja nur ein Demonstrationsbeispiel. Sinnvoll ist im Grunde nur die nach eigenen Interessen aufgebaute Datenbank, die sehr einheitlich sein kann. Der Benutzer einer lokal installierten Version hat zudem die Möglichkeit, nur wirklich vollständige und strukturierte Daten in diesem oder jenem Format aus zuverlässigen Quellen zu übernehmen.

Nr	Autor(en)	Titel	ISSN
▶ 1	Kenne, L.	Modification and cross-linking parameters in hyaluronic acid	
In:	CARBOHYDRATE POLYMERS	2013, Vol. 91, Nr. 1, S. 410-418	0144-8617
▶ 2	Song, N.	Transglutaminase cross-linking effect on sensory characteris	
In:	FOOD CHEMISTRY	2013, Vol. 136, Nr. 1, S. 144-151	0308-8146
▶ 3	Dovrolis, N.	Semantic Annotation and Linking of Medical Educational Resou	
In:	IFMBE PROCEEDINGS	2012, Vol. 37, T. 2, S. 1400-1403	1680-0737
▶ 4	Rolf, T.	Linking continental drift, plate tectonics and the thermal s	
In:	EARTH AND PLANETARY SCIENCE LETTERS	2012, Vol. 351-352, S. 134-146	0012-821X
▶ 5	Krahn, M.P.	Phosphoinositide lipids and cell polarity: linking the plasm	
In:	ESSAYS IN BIOCHEMISTRY	2012, Vol. 53, S. 15-27	0071-1365
▶ 6	Eita, M.	Optical properties of thin films of zinc oxide quantum dots	
In:	JOURNAL OF COLLOID AND INTERFACE SCIENCE	2012, Vol. 387, Nr. 1, S. 135-140	0021-9797

Abb. 6 JADE-Suchergebnis

Die Aufsatzdatenbank im jetzigen Status enthält Zitate aus einigen Quellen, wo diese Vollständigkeit und Konstanz nicht oder nur teilweise gegeben ist. So wurden viele Zitate aus der JADE-Datenbank<sup>6</sup> übernommen, die Sachtitel und manchmal auch Zeitschriftentitel willkürlich abkürzt. Contents-Linking berücksichtigt dieses. Leider lässt sich bei der Übernahme mit dem IE auch der Verfasser nicht immer eindeutig vom Sachtitel abgrenzen, was auch bei PubMed im Textformat vorkommen kann. Das kann negative Folgen bei der Verknüpfung mit Google Scholar, weniger bei Google haben. Diese Unvollkommenheiten, obwohl diese insgesamt nicht sehr ins Gewicht fallen, lassen sich durch zusätzliche Codes bzw. Vermeidung von ungenauen

<sup>5</sup> mehrere Tausend Zitate älteren Datums sind im Format Zeitschrift-Verfasser-Sachtitel erfasst, was ebenfalls zu erfolglosen Verknüpfungen führen kann

<sup>6</sup> wird nicht mehr aktualisiert, anscheinend eingestellt



Ausgabeformaten vermeiden. Notfalls kann in einem speziellen Bearbeitungsprogramm<sup>7</sup> die Angleichung an das jetzt gültige Format erreicht werden.

Eine besondere Herausforderung stellt die Übernahme von Google Scholar-Zitaten dar. In diesen sind die Zeitschriftentitel oft stark abgeschnitten. Trotzdem habe ich mich für die Übernahme von solchen Zitaten entschieden, wenn sie nicht am Anfang, sondern am Ende abgekürzt sind. Obwohl Zitate von Google Scholar extrem wenig strukturiert erscheinen, sind sie im allgemeinen trotz abgekürzter Titel erstaunlich gut zu erfassen und zu übernehmen. Allerdings sind auch hier Verfasser und Sachtitel manchmal nicht zu trennen, was zu einer nicht erfolgreichen Hyperlinkverbindung führen kann.

## Programmablauf

Es mag vielleicht von Interesse sein, wie diese Vereinheitlichung von Zitaten verschiedener Formate aus ganz unterschiedlichen Quellen im Programmablauf zustande kommt.

Zunächst wird mit Hilfe des Menüs der Abfragecode für die betreffende Datenbank ermittelt. Dieser ist für JADE (IE):

```
ta="pv=1>ps=2>pz=3>pe=4>v1=^>s1=\w+\.|-\|s-\|s\|s>z1=\|*In:>e1=\|s\|d\|d\|d\|d>e2=";
```

Mit pv, ps, pz, und pe werden die Positionen für Verfasser, Sachtitel, Zeitschriftentitel und Erscheinungsvermerk erfasst. Dahinter werden die Zeichenfolgen am Anfang (1) ggf. auch am Ende (2) von Verfasser (v), Sachtitel (s), Zeitschrift (z) und Erscheinungsvermerk (e) angegeben. In diesem Fall beginnt der Verfasser mit ^, d.h. steht am Anfang des Zitats, der Sachtitel beginnt nach den Zeichenfolgen `\w+\.|-\|s-\|s\|s`, die üblicherweise bei Verfassernamen auftreten. Die Zeitschrift beginnt nach `=\|*In:>`, der Erscheinungsvermerk nach `\s\|d\|d\|d\|d`, d.h. der Jahreszahl.

Dieser für die ausgewählte Datenbank spezifische Abfragecode wird in einem einheitlichen Programmteil bearbeitet und auf die eingegebenen Suchergebnisse angewandt. Hier ein Ausschnitt aus dem Code:

```
A=Tx2.replace(/\{\|\}\|"|SFX/g,""); // der String eines Zitats in A
p1=A.search(q1); // q1 ist der erste Suchstring für Verfasser oder Sachtitel
sq1=A.match(q1); // sq1 ist die ermittelte Zeichenkette dieses Suchstrings
if(sq1)
sq1l=sq1[0].length; // sq1l ist die Länge dieser Zeichenkette
if(!sq1)
sq1l=1;
A1=A.slice(p1+sq1l); // A1 = der erste Abschnitt des Zitats
p2=A1.search(q2); // q2 = der zweite Suchstring für Verfasser oder Sachtitel
sq2=A1.match(q2);
if(sq2)
A2=A1.slice(p2+sq2[0].length); // A2 = der zweite Abschnitt des Zitats
if(!sq2)
A2=A1.slice(p2);
if(A2.length<2)
A2=A1;
p3=A2.search(q3); // q3 = der dritte Suchstring für Verfasser oder Sachtitel
sq3=A2.match(q3);
if(sq3)
```

---

<sup>7</sup> Bearbscidat.htm

```

A3=A2.slice(p3+sq3[0].length); // A3 = der dritte Abschnitt des Zitats
if(!sq3)
A3=A2.slice(p3+1);
if(q1==v1a) // falls q1 für Verfasser
Vb=A1.substring(0,p2); // Vb = Verfasser
if(q1==s1a) // falls q1 für Sachtitel
Sb=A1.substring(0,p2); // Sb = Sachtitel
if(q2==s1a) // falls q2 für Sachtitel
Sb=A2.substring(0,p3); // Sb = Sachtitel
if(q2==v1a) // falls q2 für Verfasser
Vb=A2.substring(0,p3); // Vb = Verfasser
p4=A3.search(q4); // q4 = 4. Suchstring für Erscheinungsvermerk
Zb=A3.substring(0,p4); // Zb = Zeitschrift
Eb=A3.slice(p4+1);

```

### Übernahme eines Suchergebnisses von Aufsätzen

Abb. 6 zeigt einen Ausschnitt eines Suchergebnisses der JADE-Datenbank. Der Suchbegriff war hier *linking*. Auch hier wird die gesamte Webseite markiert, kopiert und in das in der Abbildung gezeigte Formular eingefügt. Aus dem Menü der Datenbanken wird JADE und ausserdem der verwendete Browser (in diesem Fall der IE) gewählt. Über den Schalter *Starten und Anzeigen* werden in einem Fenster die eingegebenen und übernommenen Zitate in der Form ausgeschrieben, wie sie das System weiter verarbeiten kann. Wohlgermerkt, werden die Zitate in den verschiedenen Formaten der betreffenden Datenbanken immer in derselben Form, im sogen. Contents-Linking-Format, ausgeschrieben, wie die folgende Auflistung zeigt.

```

>z=CARBOHYDRATE POLYMERS* Modification and cross-linking parameters in hyaluronic
acid *Kenne, L.*2013 Vol. 91, Nr. 1, S. 410-418 0144-8617 *Sw: linking
>z=FOOD CHEMISTRY* Transglutaminase cross-linking effect on sensory characteris *Song,
N.*2013 Vol. 136, Nr. 1, S. 144-151 0308-8146 *Sw: linking
>z=IFMBE PROCEEDINGS* Semantic Annotation and Linking of Medical Educational Resou
*Dovrolis, N.*2012 Vol. 37, T. 2, S. 1400-1403 1680-0737 *Sw: linking
>z=EARTH AND PLANETARY SCIENCE LETTERS* Linking continental drift, plate tectonics
and the thermal s *Rolf, T.*2012 Vol. 351-352, S. 134-146 0012-821X *Sw: linking
>z=ESSAYS IN BIOCHEMISTRY* Phosphoinositide lipids and cell polarity linking the plasm
*Krahn, M.P.*2012 Vol. 53, S. 15-27 0071-1365 *Sw: linking
>z=J OF COLLOID AND INTERFACE SCIENCE* Optical properties of thin films of zinc oxide
quantum dots *Eita, M.*2012 Vol. 387, Nr. 1, S. 135-140 0021-9797 *Sw: linking
>z=CITY* Un-linking the rings cities and the Olympic Games Introduct *Gibbons, A.*2012 Vol.
16, Nr. 4, S. 439-445 1360-4813 *Sw: linking

```

Die einzelnen Zitate werden auch hier wie die Zeitschriftentitel durch > begrenzt. Hinter z= folgt der Zeitschriftentitel in Grossbuchstaben, nach einem Sternchen folgt der Sachtitel, nach einem weiteren Sternchen der Verfasser und am Schluss, wieder nach einem Sternchen, der Erscheinungsvermerk. Ganz am Ende kann hinter Sw: das Suchwort folgen, hier *linking*.

### Darstellung der Zitate in Contents-Linking

Die eingegebenen Zitate werden in Contents-Linking schliesslich in einheitlicher Form dargestellt, wie die folgende Darstellung zeigt. Abb. 7 zeigt das Zitat 3 in der Vollanzeige. Ausgeschrieben sind oben die im Sachtitel enthaltenen Fachwörter als Hyperlinks. Beim Anklicken wird eine entsprechende Suche nach diesen Wörtern in der Aufsatz- und Zeitschriftentiteldatenbank

gestartet. Der Zeitschriftentitel ist mit der EZB verknüpft, die Verknüpfung mit der ZDB folgt dahinter. Manchmal folgt eine Verknüpfung mit ScienceDirect oder Springer, wenn entsprechende Hinweise bei einem Titel angegeben sind. Darunter folgen die Google- und GoogleScholar-Verknüpfungen, die Verknüpfung zur summarischen Bestandsanzeige und schliesslich hinter SW: die ermittelten Schlagwörter oder Fachgebiete. Diese sind nicht zu verwechseln mit den hinter Sw: angezeigten Wörtern, die fest mit den Zitaten verbunden sind und nicht wie die Schlagwörter und Bestandsangaben dynamisch durch einen Suchvorgang aus der Zeitschriftendatenbank gewonnen werden.

**1**

## CARBOHYDRATE POLYMERS

Modification and cross-linking parameters in hyaluronic acid

Kenne, L.

2013 Vol. 91, Nr. 1, S. 410-418 0144-8617

Sw: linking

---

**2**

## FOOD CHEMISTRY

Transglutaminase cross-linking effect on sensory characteris

Song, N.

2013 Vol. 136, Nr. 1, S. 144-151 0308-8146

Sw: linking

---

**3**

## IFMBE PROCEEDINGS

Semantic Annotation and Linking of Medical Educational Resou

Dovrolis, N.

2012 Vol. 37, T. 2, S. 1400-1403 1680-0737

Sw: linking

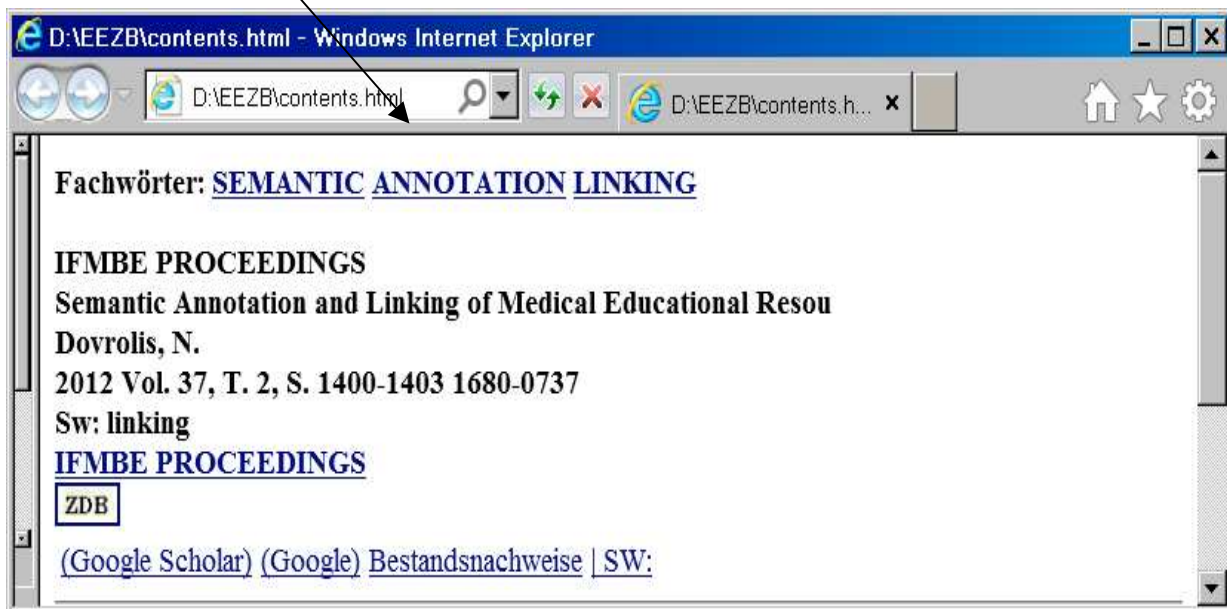


Abb. 7. Vollanzeige des Zitats mit Verknüpfungen

## Schlussbemerkung

Die hier dargestellte Technologie der Datenübernahme<sup>8</sup> ist die Basis dieses Systems und eine wesentliche Voraussetzung für dessen Effizienz, d.h. für die Leichtigkeit und Schnelligkeit, mit der Daten aus allen möglichen Quellen übernommen werden können. Das lässt sich zwar vollständig nur in einem lokal installierten System verwirklichen, weil ja nur in diesem die Speicherung der übernommenen Daten möglich ist. Gleichwohl kann auch in der Internetversion die Übernahme von Daten überzeugend demonstriert werden. Vor dem potentiellen Einspeichern von Zeitschriftentiteln, Bestandsnachweisen und Schlagwörtern schreibt das Einarbeitungsprogramm das Ergebnis, d.h. die gesamte Zeitschriftendatenbank mit allen Zuträgungen and Änderungen (!), in einem separaten Fenster aus. Über das Eingabeformular für Aufsatzzitate wiederum können temporär die eingegebenen Zitate im standardisierten Contents-Linking-Format wie bei einem Linkresolver durch Verknüpfungen zu Bestandsnachweisen und zum Volltext verwertet werden.

Insgesamt zeichnet sich diese in Contents-Linking praktizierte Datenübernahme trotz der oben erwähnten Einwände durch grosse Genauigkeit aus, was umso erstaunlicher ist, als bisher auch sehr unstrukturierte Formate erfasst wurden. Durch konsequente Beschränkung auf sehr genaue Formate mit ausreichenden Metadaten liesse sich in einer lokal installierten Version natürlich noch eine weit grössere Genauigkeit erreichen. Es gibt schliesslich *nichts in der Welt, was nicht noch genauer sein könnte* (Nikolaus von Cues).

---

<sup>8</sup> Eine weitere Darstellung wird die Such- und Verknüpfungstechnologie des Systems ausführlich beschreiben.