

Schnelles Auswerten von Literaturverzeichnissen: Anleitung für den «NBCI Batch Citation Matcher»

Marco Strehler (marcostrehler@yahoo.de). Version vom 25.08.2013. Die aktuelle Version dieser Anleitung findet sich unter <http://de.scribd.com>, Dokument-Nr. 152201179.

Intro

Ein interessantes Tool auf der NBCI-Website ist der *Batch Citation Matcher* (Direkter Link: <http://www.ncbi.nlm.nih.gov/entrez/getids.cgi>). Damit kann in der Literaturdatenbank mit einem Rutsch nach zahlreichen Einträgen gesucht werden. Das Prinzip: Minimalinformationen werden zeilenweise in ein File geschrieben, dieses hochgeladen und man erhält die *PubMed Identifiers* (PMID), mit dem die entsprechenden PubMed-Eintrag eindeutig identifiziert sind.

Schritt-für-Schritt

1. Mit einem beliebigen Text-Editor eine neue Datei erstellen.
2. Zu jedem gesuchten Artikel eine Zeile mit folgendem Format eintippen:





```
journal|jahr|volume|erste_seite|autor|schluessel|
```

Das Feld *journal* ist obligat, die anderen können ausgelassen werden, allerdings müssen bei weggelassenen Feldern die entsprechenden Feldtrenner (|) verbleiben¹. Ein weiterer vertikaler Strich wird zusätzlich als Zeilenabschluss benötigt. Als Beispiel eines verkürzten Eintrages mit nur dem Journal-Titel, dem Jahrgang (Volume) und dem Erstautor:

```
j consult clin psychol||51||baum||
```

Mit dem *schlüssel* kann ein beliebiger Wert übergeben werden, der bei der Suche ignoriert wird.

3. Textfile abspeichern
4. Auf der Seite des *Batch Citation Matcher* (URL siehe oben) die Datenbank (für unsere Zwecke «PubMed») einstellen, eigene E-Mail-Adresse angeben und das erzeugte Textfile auf der Festplatte auswählen (Knopf «Durchsuchen»).
5. «Go» anklicken

¹Das mit dem | ist auch wieder eine Sache für sich. Es gibt nämlich zwei verschiedene vertikale Striche. Einer wird (auf der Schweizerdeutschen Tastatur) mit  -  und der andere mit  -  erzeugt. Der erste ist der richtige.

6. Man erhält (nach einiger Zeit, ging bei mir bis zu einer Stunde) eine E-Mail mit dem Output. Der jeweilige PMID ist an der Zeile angehängt. Wenn das Suchkriterium nicht eindeutig war (z.B. wenn eine Autor im selben Jahr mehrere Artikel in der gleichen Zeitschrift veröffentlicht hat), kommen mehrere passende PMID zurück (siehe untenstehende Beispiele). Der BCM funktioniert nur bei 1:1-Suche. Wenn die Suchkriterien mehrdeutig sind, werden nur drei Treffer geliefert. Bei mehr als drei Fundstellen wird keine mehr angegeben. Auch ist der BCM extrem pingelig bez. Zeitschriften-Abkürzungen.

Beispiel eines Input-Files und für die Rückgabe des BCM

Inhalt des Input-Text-Files:

```
j consult clin psychol||51||baum||  
gerontologist||18||bell||  
int j aging hum dev||16||bolin||  
int j aging hum dev||26||bolin||  
gerontologist||21|8|borup||  
gerontologist||2||bourestom||  
arch gen psychiatry||48||breslau||
```

Output (erhalten per E-Mail), jeweils mit dem 7-stelligen, angehängten PMID am Ende der Zeilen.

```
j consult clin psychol||51||baum||6619364  
gerontologist||18||bell||750293  
int j aging hum dev||16||bolin||7184870  
int j aging hum dev||26||bolin||3338865  
gerontologist||21|8||NOT_FOUND  
gerontologist||2||NOT_FOUND  
arch gen psychiatry||48||breslau||AMBIGUOUS 1845224,1996917
```

Bequemer das Input-File erzeugen mit Hilfe von OpenOffice Calc

Eine solche Input-Liste anzulegen ist «von Hand» mit einem beliebigen Text-Editor möglich. Es geht aber auch komfortabler; und zwar bietet sich bei solchen strukturierten Eingaben eine Tabellenkalkulation an. Im folgenden am Beispiel von OpenOffice Calc dargestellt.

Zunächst wird in Calc eine neue Tabelle angelegt. In die Spalten A – E kommen (streng in dieser Reihenfolge) Journal, Jahr, Volume, erste Seite und Erstautor. Wie oben beschrieben ist nur das Journal obligatorisch, die anderen Felder dürfen leer gelassen werden (solange

	A	B	C	D	E	F
1	j consult clin psychol	1983	51	565	baum	
2	gerontologist		18		bell	
3	int j aging hum dev		16		bolin	
4	int j aging hum dev		26		bolin	
5	gerontologist		21	8	borup	
6	gerontologist		2		bourestom	
7	arch gen psychiatry		48		breslau	
8	int j aging hum dev		26		bolin	
9	am j psychiatry		102		carmago	
10	gerontologist		23		coffman	
11						

Abbildung 1: Beispiel, wie die Tabelle in OpenOffice Calc ausgefüllt werden kann. Das Zeichen | in der Spalte F ist wichtig und dient dazu, dass beim Abspeichern der Datei im Textformat – wie vom BCM verlangt – auch am Zeilenende ein vertikaler Strich steht.

die Publikation damit eindeutig identifiziert werden kann). In die Spalte F kommt der senkrechte Strich (|). Zur Erinnerung: auf der Schweizerdeutschen Tastatur dargestellt mit - . Siehe dazu Abbildung , Seite 3.

Fertig ausgefüllt wird die Tabelle mit «Speichern unter ...» auf der Festplatte abgelegt. Und zwar unter dem Format «Text CSV (.csv)». Im Speichern-Dialog die Optionen «Automatische Dateinamenerweiterung» **deaktivieren** und «Filtereinstellung bearbeiten» **aktivieren**. Als Dateinamen wählt man mit Vorteil ein aussagekräftiger Name. Wichtig ist auch die Extension «txt».

Es erscheint ein Dialog, in dem die Trenner eingestellt werden. Dabei werden die Feld- und Texttrenner gemäss der Abbildung , Seite 4) eingestellt.

PMID und weiter?

Die PMID kann verwendet werden, um den Eintrag in PubMed direkt zu finden. Dazu kopiert man die Nummer einfach ins Suchfeld. Es kann auch nach mehreren PMID gesucht werden, indem sie mit einem Leerzeichen getrennt eingegeben werden. Wenn also – ausgehend von obigem Beispiel – nach folgendem gesucht wird (siehe letzte Zeile im Output):

1845224 1996917

...erhält man die zwei entsprechenden PubMed-Einträge (Der Autor *Breslau* hat in jenem Jahr zwei Artikel im *Arch Gen Psychiatry* veröffentlicht.

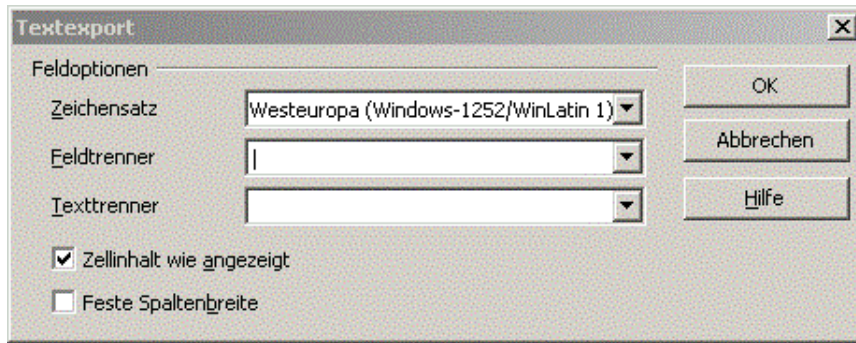


Abbildung 2: Die Filtereinstellung. Feld- und Texttrenner können per Menü eingestellt werden, die Voreinstellung muss für unsere Zwecke aber überschrieben werden. Im Feldtrenner kommt der bekannte vertikale Strich und der Texttrenner bleibt leer.

So kann aber auch direkt aus der entsprechenden Literaturdatenbank (z.B. JabRef) heraus gesucht werden. Mehrfachtreffer und das entsprechende Herausklauen des gesuchten Eintrages entfallen.