

Historische Literaturdatenbanken von Chadwyck-Healey/ProQuest und ihre Verwendung als Korpora

Titel	Daten	Texte	Wörter	KWIC	Rohdaten
Early English Prose Fiction	1518 – 1700	211	10 Mio	✓	✓
Eighteenth-Century Fiction	1700 – 1780	96 (-3)	10 Mio	✓	✓
Nineteenth-Century Fiction	1782 – 1903	250	40 Mio	✓	✓
Early American Fiction	1789 – 1875	567	18 Mio		✓
English Prose Drama	Erstauff. 1540 – 1922 Erstdruck 1540 – 1965	1651	30 Mio	✓	✓
American Drama (Vers + Prosa)	Erstauff. 1762 – 1929 Erstdruck 1714 – 1967	> 1100	> 30 Mio		

Die Datenbanken sind von allen Internet-fähigen Computern auf dem Campus zugänglich, und zwar im Netz der UB (<http://www.uni-bamberg.de/ub/>) unter "DBIS (Datenbank-Infosystem)" => "Anglistik/Amerikanistik".

I. Durchführung einer Suche mit der integrierten Software

1. Im *Standard Search*-Fenster unter Keyword das Suchwort eingeben.

Man kann, wenn man Varianten des Wortes einschließen will, durch die Pfeiltaste rechts von der Eingabezeile die *Keyword Browse*-Liste bekommen und in ihr mehrere Suchwörter markieren.

Um 2 Wörter in bestimmtem Abstand voneinander zu suchen, kann man in die Eingabezeile eintippen:

[Suchwort 1] within n words of/after/before [Suchwort 2]

z.B.: question within 3 words before if
if within 9 words after question

Nicht suchbare Wörter sind *after, and, before, cont, containing, directly, in, inside, like, not, of, or, with, within, word, words*; diese können aber gesucht werden, indem man sie in Anführungszeichen einschließt,

z.B. person "of" quality oder
"person of quality"

Eckige Klammern können benutzt werden, um Graphievarianten zu suchen,

z.B. v[ie]rtue findet virtue und vertue
p[iy]racy findet piracy und pyracy

Die Wildcard ? steht für ein beliebiges Zeichen,

z.B. s?nk findet sink, sank, sunk, etc.

Die Wildcard * steht für eine beliebige Anzahl von Zeichen oder, eingerahmt von Leerschritten, für ein beliebiges Wort,

z.B. person* findet person, persons, personal, personally, personification, etc.
person * quality findet person of quality, person and quality, etc.

Komplexere Suchabfragen (*Command line searches*) können auch durchgeführt werden, bieten aber keine wesentlich weiter gehenden Suchmöglichkeiten.

2. Je nach Datenbank können weitere Einschränkungen getroffen werden, z.B. nach Titel, Autor, Publikationsjahr, Geschlecht, Nationalität und Ethnie des Autors/der Autorin.
3. Zum Starten der Suche *OK* anklicken. Es erscheint das Fenster *Generating Summary of Matches*. Nach Beendigung des Suchvorganges erscheint das Fenster (*Brief*) *Summary of Matches*.

Leider erlauben nur die in der obigen Tabelle mit KWIC (= Key Word In Context) markierten (britischen) Datenbanken es, alle Treffer in einem Fenster aufzulisten. Als weiteres Verfahren ist in diesem Fall Folgendes am Praktikabelsten (da man die Ergebnisse im Programm selbst nicht/nur sehr schlecht bearbeiten kann):

4. Man markiert alle Autoren und Werke (indem man mit Hilfe der Bildlaufleiste ans Ende der Liste fährt und bei gedrückter Shift-Taste mit der Maus den letzten Eintrag anklickt; alle Einträge werden daraufhin blau unterlegt).
5. Man klickt die Schaltfläche *View Context* (unterhalb der Ergebnisliste) an. Im nun erscheinenden Fenster sieht man blau unterlegt die gesuchten Beispielwörter und deren unmittelbaren Kontext. Will man den weiteren Kontext sehen, bekommt man ihn per Doppelklick auf das betreffende Beispiel oder durch Markieren des Beispiels und Anklicken von *View Text* zu Gesicht.
6. Man kann nun entweder alle Beispiele speichern oder sie sofort aussortieren und nur die ausgewählten speichern.
 - a) Um alle zu speichern, markiert man im Fenster *Context of Matches* alle aufgelisteten Zitate (indem man bei gedrückter Shift-Taste mit der Maus den letzten Eintrag anklickt). Im Menü *File* auf *Save...* klicken. Dort *Context of Matches* eingestellt lassen, unter *Drives* das Ziellaufwerk und ggf. den Zielordner auswählen und einen Dateinamen eingeben. *OK*.
 - b) Um nur die relevanten Beispiele zu speichern, kann man im Fenster *Context of Matches* mit Hilfe der Maus, der <Strg> (bzw. <Ctrl>)- und der Shift-Tasten die gewünschten Beispiele markieren (Vorsicht, nicht schon markierte wieder löschen!). Dann geht man im Menü *File* auf *Save* und ändert *Context of Matches* in *Selected Matches* um, wählt das Ziellaufwerk und ggf. den Zielordner aus, gibt einen Dateinamen ein und klickt auf *OK*.

7. Man kann die Ergebnisse jetzt mit *Wordsmith* weiterbearbeiten oder die eben erstellte Datei mit einem Textverarbeitungsprogramm aufrufen und ausdrucken oder bearbeiten.

Bei den nicht mit KWIC markierten (amerikanischen) Datenbanken muss ein umständlicheres Verfahren gewählt werden:

4. Die einfachste Möglichkeit, alle Treffer zu sichten, besteht darin, aus dem *Summary of Matches* durch Doppelklick auf den ersten Titel in den ersten Volltext zu springen, und anschließend durch wiederholtes Klicken auf das Symbol *Next hit* von einem zum anderen Treffer zu springen. Sinnvolles Speichern ist bei dieser Softwareversion leider nicht möglich.

II. Durchführung einer Suche mit *Wordsmith*

Die Rohdaten der in der obigen Tabelle markierten Datenbanken liegen uns vor und sollen in nächster Zeit auf dem Lehrstuhl-Server zugänglich gemacht werden. Diese bieten den Vorteil, dass sie mit *Wordsmith* zu bearbeiten sind. *Wordsmith* bietet im Gegensatz zur integrierten Software bessere Möglichkeiten, den Suchausdruck zu definieren und die Konkordanz auszusortieren und zu kategorisieren.

Von besonderer Bedeutung bei der Auswahl der Texte ist der Aufbau des Dateinamens, der folgende Informationen kodiert:

z.B.	<input type="text" value="515"/> <input type="text" value="01"/> <input type="text" value="01"/> <input type="text" value="m"/> <input type="text" value="584"/>
bedeutet	Geburtsjahr des Autors/der Autorin: 1 <input type="text" value="515"/> laufende Nummer unter den Autor(inn)en mit diesem Geburtsjahr: <input type="text" value="01"/> laufende Nummer des Werks dieses Autors/dieser Autorin: <input type="text" value="01"/> Geschlecht des Autors/der Autorin: <input type="text" value="m"/> Publikationsjahr des Werkes: 1 <input type="text" value="584"/>
dies ist	William Baldwin (c.1515-1563): <i>Beware the cat</i> (1584).
z.B.	<input type="text" value="fl"/> <input type="text" value="01"/> <input type="text" value="m"/> <input type="text" value="656"/>
bedeutet	Geburtsjahr des Autors/der Autorin: unbekannt laufende Nummer des Werks dieses Autors/dieser Autorin: <input type="text" value="01"/> Geschlecht des Autors: <input type="text" value="m"/> Publikationsjahr des Werkes: 1 <input type="text" value="656"/>
dies ist	Samuel Holland (?): <i>Don Zara Del Fogo</i> (1656).
z.B.	<input type="text" value="an"/> <input type="text" value="03"/> <input type="text" value="683"/>
bedeutet	Autors/Autorin: <input type="text" value="an"/> onym laufende Nummer des Werks mit diesem Publikationsjahr: <input type="text" value="03"/> Publikationsjahr des Werkes: 1 <input type="text" value="683"/>
dies ist	Anonymous: <i>The London Jilt</i> (1683).

Die volle bibliographische Information über die Werke kann durch Öffnen der Datei mit einem Textverarbeitungsprogramm erschlossen werden. Sie findet sich jeweils am Anfang der Datei.

Wordsmith erlaubt nun bei der Auswahl der zu durchsuchenden Texte entweder die Sortierung nach Dateinamen (= nach Geburtsjahr des Autors/der Autorin) oder nach Dateinamenerweiterung (Type = nach Publikationsjahr des Werkes).