ZHAW Zürcher Hochschule für Angewandte Wissenschaften Departement Angewandte Linguistik IUED Institut für Übersetzen und Dolmetschen

Benutzerdokumentation CQPweb



Funktionsweise und Nutzung des Korpusanalyse-Tools CQPweb

Impressum

Dokumentversion: 1.0 **Datum**: 7. Mai 2018

Alle Informationen in der Benutzerdokumentation beruhen auf folgendem Stand: 31. März 2018.

Kontakt:

korpuslinguistik@zhaw.ch

Verfasser:

Borkovic Aleksandar Geissbühler Tim Kratter Romina Marty Andrea Mathis Ramun Müller Gabriel Rettig Simon Ryter Susanne Steiger Désirée

Im Auftrag von:

Martin Schuler shmt@zhaw.ch

Julia Krasselt krss@zhaw.ch

Inhaltsverzeichnis

1	Ein	Einführung			
2	Pro	oduktbeschreibung	3		
3	Ber	nutzeroberfläche	4		
	3.1	Korpusfunktionen-Menü	4		
	3.2	Benutzerseite	6		
4	Nut	tzung	8		
	4.1	Benutzeraccount erstellen	8		
	4.2	Vorinstalliertes Korpus wählen	9		
	4.3	Benutzeroberfläche anpassen	10		
	4.4	Korpusinformationen anzeigen	12		
	4.5	Suchanfrage erstellen	15		
	4.6	Suchanfrage speichern	23		
	4.7	Darstellung der Suchergebnisse verändern	26		
	4.8	Subkorpora erstellen	29		
	4.9	Subkorpora verwalten	35		
	4.10	Texte in Subkorpora verwalten	37		
	4.11	Korpus untersuchen	40		
	4.12	Word lookup: Wortformen suchen	48		
	4.13	Frequency lists: Frequenzlisten erstellen	51		
	4.14	Keywords: Typisches Vokabular berechnen	52		
	4.15	Ergebnisse herunterladen	55		
5	Glo	ossar	63		
	5.1	Fachbegriffe	63		
	5.2	Metadaten			
	5.3	Tagset	65		
6	FA	Q	67		
7		iterführende Informationen			
		Literatur			
	7.1 7.2	Links			
	1.2	LIDKS	n9		

1 Einführung

Dieses Kapitel stellt eine Einführung in die Benutzerdokumentation von *CQPweb* dar. Zuerst werden der Zweck und die Zielgruppe der Benutzerdokumentation genannt, danach werden der Aufbau der Benutzerdokumentation und die Verwendung der Zeichen erläutert.

Zweck der Benutzerdokumentation

Die vorliegende Benutzerdokumentation beschreibt die Arbeit mit dem Korpusanalyse-Tool *CQPweb*. Es existieren bereits einige Ansätze zur Dokumentation von *CQPweb*, jedoch ist noch kein ausführliches Dokument erstellt worden, das zur grundlegenden Einführung oder bei Problemen konsultiert werden kann. Die Mitarbeitenden des Departements Angewandte Linguistik, die für die Verwaltung von *CQPweb* zuständig sind, erhalten einige Nutzeranfragen. Aus diesen Gründen wurden im Rahmen des Seminars "Technische Dokumentation: Projekt" die Studierenden der Vertiefung Technikkommunikation des Studienjahrgangs AS15 damit beauftragt, eine Benutzerdokumentation zu *CQPweb* zu erstellen. Diese Benutzerdokumentation soll die Dokumentationslücke schliessen und dem Problem der Nutzeranfragen Abhilfe schaffen.

Zielgruppe

CQPweb wurde 2012 von Andrew Hardie von der Lancaster University als Open-Source-Software entwickelt. Am Departement Angewandte Linguistik der ZHAW wird CQPweb dafür genutzt, um im Korpus Swiss-AL (Swiss-Angewandte-Linguistik) zu suchen. Sowohl das Korpus als auch CQPweb stehen den Angehörigen des Departements Angewandte Linguistik für Forschungszwecke frei zur Verfügung. Diese Benutzerdokumentation richtet sich demnach an alle Mitarbeitende, Dozierende und Studierende, die CQPweb bereits nutzen oder nutzen möchten, unabhängig davon, ob sie über tiefergehende technische Kenntnisse verfügen oder nicht.

Aufbau

Im nächsten Kapitel wird *CQPweb* kurz beschrieben und seine Funktionsweise erklärt. Darauf folgt eine Übersicht der Benutzeroberfläche von *CQPweb*, die dem Nutzer den Einstieg in die Arbeit erleichtern soll. Das vierte Kapitel stellt den Hauptteil der Benutzerdokumentation dar: Darin ist beschrieben, wie *CQPweb* genutzt wird und wie einzelne Funktionen ausgeführt werden. Im Zentrum stehen dabei das Erstellen von Suchanfragen sowie die Analysemöglichkeiten der Suchergebnisse. Die Beschreibungen der jeweiligen Funktionen wurden zur besseren Veranschaulichung mit Screenshots und, wo sinnvoll, mit beispielhaften Suchszenarien ergänzt.

Zeichenverwendung

In der nachfolgenden Tabelle sind die verwendeten Zeichen sowie ihre Bedeutung aufgelistet:

Zeichen	Verwendung
†× ×ò	Alternative
*G	Glossareintrag
•	Hinweis
	Tipp
	Voraussetzung
A	Warnung
→	Ziel
>	Zwischenhinweis (in Handlungsanweisung)
→	Zwischenziel (in Handlungsanweisung)

2 Produktbeschreibung

Dieses Kapitel beschreibt *CQPweb*. Es zeigt auf, wofür und wie *CQPweb* genutzt werden kann.

Korpus Swiss-AL

Zurzeit entsteht an der ZHAW am Departement Angewandte Linguistik das Korpus Swiss-AL (Swiss-Angewandte-Linguistik). Das Korpus enthält Texte von Schweizer Webseiten in den Sprachen Deutsch, Französisch, Italienisch und Englisch und dient als Grundlage für diskurslinguistische Projekte an der ZHAW sowie für die Wahl des "Wort des Jahres Schweiz".¹

CQPweb als Suchtool

Für die Suche in diesem Korpus steht den Nutzern *CQPweb* zur Verfügung. Ursprünglich existierte *CQP* (*Corpus Query Processor*) nur in einer kommandozeilenbasierten Form. Inzwischen gibt es aber auch eine browserbasierte Benutzeroberfläche, *CQPweb*, welche die Suche einem breiteren Publikum zugänglich macht (Hardie 2012). Zum Zeitpunkt der Erstellung der vorliegenden Benutzerdokumentation ist am Departement Angewandte Linguistik die Version 3.1.16 von *CQPweb* in Betrieb.

Zugriff

Um im Korpus Swiss-AL zu suchen, wurde *CQPweb* auf einem ZHAW-internen Server installiert. *CQPweb* kann innerhalb des ZHAW-Netzes in einem beliebigen Browser über die Adresse http://160.85.180.39/CQPweb/ erreicht werden. Für den Zugriff auf das ZHAW-Netz ausserhalb der Schulumgebung, z. B. von zuhause aus, wird eine VPN-Verbindung benötigt. Wie der entsprechende VPN-Client heruntergeladen und genutzt wird, ist in der Anleitung unter https://intra.zhaw.ch/ict beschrieben. Bei Problemen kann entweder das IT-Büro im Gebäude SM aufgesucht oder auf https://servicedesk.zhaw.ch/ ein Ticket erfasst werden.

Funktionen

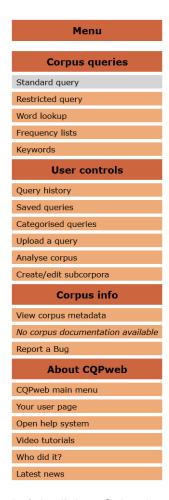
Nebst der grundlegenden Möglichkeit, in *CQPweb* nach Wortformen oder Wortabfolgen zu suchen, können auch weiterführende Analysen durchgeführt werden. Dazu zählen unter anderem das Ermitteln von Kollokationen und das Berechnen von Keywords. Für Suchanfragen stehen zwei Abfragesprachen zur Verfügung: die *Simple Query Language**^G und die *CQP Query Language**^G, wobei letztere für erfolgreiche Suchergebnisse empfohlen wird.

¹ https://www.zhaw.ch/de/linguistik/forschung/swiss-al/ [zit. 01.03.2018]

3 Benutzeroberfläche

Dieses Kapitel beschreibt die zwei grundlegenden Benutzeroberflächen von *CQPweb*: das Korpusfunktionen-Menü und die Benutzerseite.

3.1 Korpusfunktionen-Menü



Auf der linken Seite des Bildschirms befinden sich die Verknüpfungen zu den Funktionen von *CQPweb*.

Corpus queries

Unter *Corpus queries* sind die verschiedenen Funktionen für Korpusanfragen aufgelistet:

Standard query: Hier können normale Suchanfragen erstellt werden (siehe Kap. 4.5 Suchanfrage erstellen, S. 15).

Restricted query: Hier können eingeschränkte Suchanfragen erstellt werden (siehe Kap. 4.5 Suchanfrage erstellen, S. 15).

Word lookup: Hier können Wortformen gesucht werden (siehe Kap. 4.12 Word lookup: Wortformen suchen, S. 48).

Frequency lists: Hier können Frequenzlisten erstellt werden (siehe Kap. 4.13 Frequency lists: Frequenzlisten , S. 51).

Keywords: Hier können Keywords berechnet werden (siehe Kap. 4.14 Keywords, S. 52).

User controls

Unter *User controls* sind Funktionen aufgeführt, die unter dem Benutzeraccount abgespeicherte Daten verwenden.

Query history: Hier können alle vom Nutzer erstellten Suchanfragen angezeigt werden.

Saved queries: Hier können die vom Nutzer gespeicherten Suchanfragen angezeigt werden (siehe Kap. 4.6 Suchanfrage speichern, S. 23).

Categorised queries: Hier können kategorisierte Suchanfragen angezeigt, separiert, erweitert oder gelöscht werden (siehe Kap. 4.11 Korpus untersuchen, S. 40).

Upload a query: Hier kann eine Suchanfrage hochgeladen werden.



ACHTUNG: Die multivariante Analyse eines Korpus ist komplex und eignet sich nur für fortgeschrittene Nutzer. Des Weiteren wurde die Funktion nie fertig entwickelt, die Ergebnisse sind daher mit Vorsicht zu verwenden.

Analyse corpus: Hier kann ein Korpus multivariant analysiert werden.

Create/edit subcorpora: Hier können Subkorpora erstellt oder editiert werden (siehe folgende Kapitel: Kap. 4.8 Subkorpora erstellen, S. 29 / Kap. 4.9 Subkorpora verwalten, S. 35 / Kap. 4.10 Texte in Subkorpora verwalten, S. 37).

Corpus info

Unter *Corpus info* sind Funktionen aufgeführt, die Informationen über das Korpus anzeigen.

View corpus metadata: Hier können die Metadaten des Korpus angezeigt werden (siehe Kap. 4.4 Korpusinformationen anzeigen, S. 12).

Corpus documentation: Hier kann die Korpusdokumentation angezeigt werden (siehe Kap. 4.4 Korpusinformationen anzeigen, S. 12). Diese Funktion ist im Korpus September 2017 – DE nicht verfügbar.

Report a bug: Hier wird versucht, einen fehlerhaften Link zu öffnen.

About CQPweb

Unter *About CQPweb* sind Funktionen aufgeführt, die Informationen über *CQPweb* anzeigen.

CQPweb main menu: Hier wird die Startseite von CQPweb angezeigt.

Your user page: Hier werden die Einstellungen des Benutzeraccounts angezeigt (siehe Kap. 3.2 Benutzerseite, S. 6).

Open help system: Hier wird das Hilfemenü von *CQPweb* geöffnet.

Video tutorials: Hier werden die Tutorials zu den verschiedenen Funktionen von *CQPweb* angezeigt.

Who did it?: Hier werden die Urheber von CQPweb angezeigt.

Latest news: Hier werden die Veränderungen durch die letzten Aktualisierungen angezeigt.

3.2 Benutzerseite

Menu	
Your account	
Overview	
Interface settings	
User macros	
Corpus permissions	
Your files and corpora	
Manage files	
Upload a file	
View corpora	
Create corpus	
Account actions	
Account details	
Change password	
Log out of CQPweb	
About CQPweb	
CQPweb main menu	
Your user page	
Open help system	
Video tutorials	
Who did it?	
Latest news	

Auf der linken Seite des Bildschirms befinden sich die Verknüpfungen zu den Funktionen und Einstellungen der Benutzerseite von *CQPweb*.

Your Account

Unter **Your Account** sind benutzerdefinierte Einstellungen und Berechtigungen einsehbar.

Overview: Hier gelangen Sie zur Übersicht der Benutzerseite.

Interface Settings: Hier können Standardeinstellungen für Suchanfragen und andere Funktionen vorgenommen werden (siehe Kap. 4.3 Benutzeroberfläche anpassen, S. 10).

User Macros: Hier können Makros erstellt werden.



Um Makros erstellen zu können, werden Programmierkenntnisse benötigt.

Corpus permissions: Hier können Sie direkt auf die verschiedenen Korpora zugreifen, die für Ihren Benutzeraccount freigegeben sind.

Your Files and Corpora

Die Funktionen unter diesem Menüpunkt sind in der aktuellen Version (3.1.16) nicht verfügbar.

Account Actions

Unter *Account Actions* können Sie die Angaben zu Ihrem Benutzeraccount verwalten und sich ausloggen.

Account Details: Hier können die Angaben Ihres Benutzeraccounts geändert werden.

Change Password: Hier können Sie Ihr Passwort ändern.

Log out of CQPweb: Hier können Sie sich ausloggen.

About CQPweb

Unter *About CQPweb* sind Funktionen aufgeführt, die Informationen über *CQPweb* anzeigen.

CQPweb main menu: Hier wird die Startseite von CQPweb angezeigt.

Your user page: Hier gelangen Sie zurück zur Übersicht.

Open help system: Hier würde das Hilfemenü von *CQPweb* geöffnet werden, der Link ist aber fehlerhaft.

Video tutorials: Hier werden die Tutorials zu den verschiedenen Funktionen von *CQPweb* angezeigt.

Who did it?: Hier werden die Urheber von CQPweb angezeigt.

Latest news: Hier werden die Veränderungen durch die letzten Aktualisierungen angezeigt.

4 Nutzung

Das vorliegende Kapitel beschreibt die Nutzung von *CQPweb*, indem die wichtigsten Funktionen erläutert werden.

4.1 Benutzeraccount erstellen

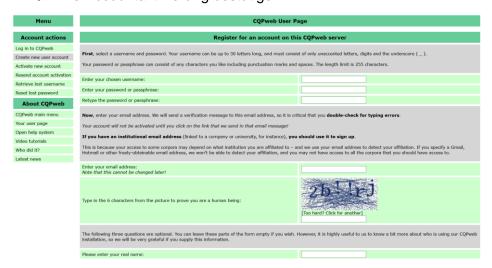
Für die Nutzung von *CQPweb* (http://160.85.180.39/CQPweb/) ist ein Benutzeraccount nötig. In diesem Kapitel wird beschrieben, wie dieser erstellt wird.



Voraussetzung: Der Computer ist direkt oder per VPN mit dem ZHAW-Netz verbunden.

Führen Sie die nachfolgenden Schritte aus, um einen Benutzeraccount zu erstellen:

- 1. Die Startseite von *CQPweb* in einem beliebigen Browser öffnen.
- 2. Auf Create account klicken.
- > Verwenden Sie zwingend eine ZHAW-E-Mail-Adresse. Wählen Sie ein Passwort nach den üblichen Sicherheitsstandards.
 - 3. Die Eingabefelder ausfüllen.
 - 4. Dem Link in der Bestätigungsmail folgen.
 - 5. Die Accountaktivierung bestätigen.





Ihr Account ist nun erstellt und Sie können mit der Nutzung von *CQPweb* beginnen.

4.2 Vorinstalliertes Korpus wählen

Auf der Startseite von *CQPweb* sind die vorinstallierten Korpora ersichtlich. In diesem Kapitel wird beschrieben, wie ein Korpus geöffnet wird. Nicht alle Korpora können geöffnet werden, da eine Lizenz benötigt wird.



Voraussetzung: Sie sind eingeloggt.

Führen Sie den nachfolgenden Schritt aus, um ein vorinstalliertes Korpus zu wählen:

- 1. Die Startseite von CQPweb in einem beliebigen Browser öffnen.
- 2. Unter **Corpora available on this server** auf das gewünschte Korpus klicken.

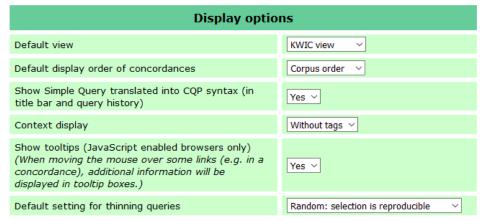


Sie befinden sich nun auf *CQPweb* auf der Startseite des gewählten Korpus und können mit Ihrer Suche beginnen.

4.3 Benutzeroberfläche anpassen

In diesem Kapitel werden sämtliche Einstellungen beschrieben, die für die Auswahl und die Darstellung der Suchergebnisse relevant sind. Die Werte, die auf der Seite *User interface settings* angezeigt werden, werden als Standard übernommen, können aber bei der jeweiligen Suche noch verändert werden.

Display Options



Hier sind die Einstellungen für die Benutzeroberfläche aufgeführt, die angepasst werden können.

Default view: Die Darstellung der Suchergebnisse kann zwischen **KWIC View***^G und **Line View***^G gewechselt werden.

- KWIC (Keyword in Context) View: Kontexte links und rechts vom Suchergebnis werden in separaten Spalten dargestellt.
- Line View: Suchergebnis und Kontext werden in einer gemeinsamen Spalte dargestellt.

Default display order of concordances: Die Reihenfolge der Ergebnisse kann entweder chronologisch absteigend **(Corpus Order)** oder zufällig **(Random Order)** angezeigt werden.

Show Simple Query translated into CQP syntax: Die verwendeten Suchbegriffe können entweder einfach oder mit der CQP Syntax angezeigt werden (für Informationen zur CQP Syntax siehe Kap. 4.5 Suchanfrage erstellen, S. 15).

Context display: Der bei den Ergebnissen angezeigte Kontext kann entweder mit oder ohne Annotation angezeigt werden.

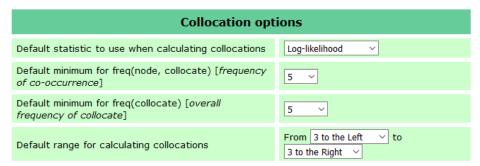
Show tooltips: Tooltipps mit Zusatzinformationen erscheinen, wenn bei der Anzeige der Ergebnisse der Mauszeiger über einen Link gehalten wird. Das Anzeigen von Tooltipps kann aktiviert oder deaktiviert werden.



Tooltipps können nur auf JavaScript-fähigen Browsern angezeigt werden.

Default Setting for thinning queries: Die Anzahl der angezeigten Ergebnisse kann auf eine zufällige Auswahl reduziert werden. Die Art der zufälligen Auswahl kann entweder als reproduzierbar (selection is reproducible) – falls die Ergebnisse später wieder so reduziert werden sollen können – oder als nicht reproduzierbar (selection is not reproducible) festgelegt werden.

Collocation options



Hier sind die Einstellungen aufgeführt, die für das Berechnen von Kollokationen relevant sind.

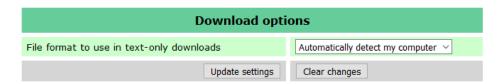
Default statistic to use when calculating collocations: Die Art der Statistik, die beim Berechnen von Kollokationen verwendet werden soll, kann hier ausgewählt werden.

Default minimum for freq (node, collocate): Die benötigte Frequenz von Kombinationen aus Knoten und Kollokat für das Aufführen des Ergebnisses kann hier bestimmt werden.

Default minimum for freq (collocate): Die minimale Frequenz eines Wortes, um als Kollokat eines Knotens in Frage zu kommen, kann hier bestimmt werden.

Default range for calculating collocations: Die maximale Distanz zwischen Knoten und Kollokat, die für Kollokationen berücksichtigt werden soll, kann hier bestimmt werden.

Download options



File format to use in text-only downloads: Das verwendete Betriebssystem kann entweder manuell definiert werden oder es kann von CQPweb automatisch erkannt werden (Automatically detect my computer).

4.4 Korpusinformationen anzeigen

In diesem Kapitel wird beschrieben, wie Korpusinformationen *(Corpus info)* angezeigt werden können. Es gibt zwei Arten von Korpusinformationen: Korpuseigenschaften *(Corpus metadata)* und Korpusdokumentation *(Corpus documentation).*



Voraussetzung: Die Zugriffsrechte zum Korpus sind im Benutzeraccount enthalten.

Korpuseigenschaften anzeigen



Führen Sie den nachfolgenden Schritt aus, um die Korpuseigenschaften anzuzeigen:

1. Unter Corpus info auf View corpus metadata klicken.



Im neuen Fenster werden drei Kategorien von Korpuseigenschaften angezeigt.



Die Art und Zahl der Korpuseigenschaften, die angezeigt werden, variieren je nach Korpus. Ausserdem enthält nicht jedes Korpus eine Korpusdokumentation. Nachfolgen werden die Metadaten am Beispiel des Korpus «September 2017 – DE Only» erläutert.



Eine Liste allgemeingültiger Metadaten befindet sich im Glossar (siehe Kap. 5.2 Metadaten, S. 64).

Metadata for...

Unter *Metadata for* sind die verschiedenen Metadaten aufgelistet, die das Korpus betreffen:

Corpus Name: Name des Korpus.

CQPweb's short handles for this corpus: Kurzversion des Namens des Korpus.

Total number of corpus texts: Anzahl der im Korpus enthaltenen Texte.

Total words in all corpus texts: Anzahl der im Korpus enthaltenen Wörter (gleiche Wörter werden mehrmals gezählt).

Word types in the corpus: Anzahl der im Korpus enthaltenen einzigartigen Wörter (gleiche Wörter werden einfach gezählt).

Type:Token ratio: Verhältnis zwischen der Anzahl der im Korpus enthaltenen Wörter *(type)* und der Anzahl der enthaltenen einzigartigen Wörter *(token)*.

Text metadata

Unter *Text metadata* sind die verschiedenen Metadaten aufgelistet, die den Korpustext betreffen und nach welchen die Texte bei Suchanfragen sortiert werden können:



Die folgenden Metadaten beziehen sich ausschliesslich auf die Texte des Korpus «September 2017 – DE Only». Texte von anderen Korpora können weniger oder andere Metadaten enthalten.

Class: Akteursklasse, zu welcher der Text gehört (z. B. pbv).

Date Published: Datum, an dem der Text veröffentlicht wurde.

Date Crawled: Datum, an dem der Text ins Korpus aufgenommen wurde.

Day: Tag im Monat, an dem der Text veröffentlicht wurde.

German Sentences: Prozentualer Anteil der Sätze in deutscher Sprache im Text.

English Sentences: Prozentualer Anteil der Sätze in englischer Sprache im Text.

Filetype: Dateiformat, in dem der Text vorliegt

French Sentences: Prozentualer Anteil der Sätze in französischer Sprache im Text.

Italian Sentences: Prozentualer Anteil der Sätze in italienischer Sprache im Text.



ACHTUNG: Die Funktionen *German Sentences*, *English Sentences* etc. berücksichtigen das Dateiformat, in welchem der Text abgespeichert wurde. So kann es z. B. bei Texten im HTML-Format vorkommen, dass Begriffe aus dem Quellcode oder aus den User-Kommentaren unterhalb des eigenen Artikels ungewünscht berücksichtigt werden.

Month: Monat, in dem der Text veröffentlicht wurde.

Other Languages: Prozentualer Anteil der Sätze in nicht deutscher, nicht englischer, nicht französischer oder nicht italienischer Sprache im Text.

Rawdate: Link zum Quelltext der Textquelle (z. B. HTML-Text).

Source: (Publikations-)Quelle der Texte (z. B. 20min).

Subclass: Art der (Publikations-)Quelle der Texte (z. B. Tageszeitung).

URL: Link zur Textquelle.

Year: Jahr, in dem der Text veröffentlicht wurde.

The primary classification of texts is based on: Voreingestelltes Kriterium für die Zuordnung der Texte.

Word-level annotation

Unter *Word-level annotation* sind die Metadaten aufgelistet, welche die Wörter im Korpus betreffen.



Die folgenden Metadaten beziehen sich ausschliesslich auf die Wörter des Korpus «September 2017 – DE Only». Wörter von anderen Korpora können weniger oder andere Metadaten enthalten.

Words in this corpus are annotated with: Metadatenkategorien, mit denen die Wörter annotiert sind (z. B. Lemma, Eigenname etc.).

Named Entity Type*G: Eigennamen-Kategorie des Wortes (z. B. *Alexander der Grosse*, Personenname).

Lemma*G: Grundform eines Wortes (wie im Wörterbuch).

Part of Speech*G: Wortart, grammatische Kategorie eines Wortes.

The primary tagging scheme is: Part of Speech: Voreingestelltes Kriterium, nach dem die Wörter mit Annotation versehen wurde.

Further information about this corpus is available on the web at: Weiterführende Informationen zum Korpus.

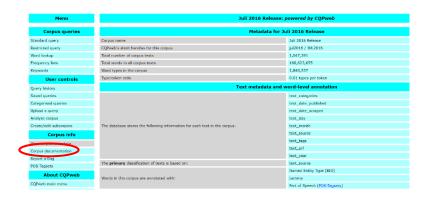


ACHTUNG: Durch das Klicken auf den unter *Further information* about this corpus is available on the web at angezeigten Link verlassen sie *CQPweb*.



Öffnen Sie Links per Rechtsklick in einem separaten Tab. So haben Sie gleichzeitig Zugriff auf *CQPweb* und das neu geöffnete Dokument.

Korpusdokumentation anzeigen



Führen Sie den nachfolgenden Schritt aus, um die Korpusdokumentation anzuzeigen:



ACHTUNG: Durch das Ausführen des nachfolgenden Handlungsschrittes verlassen Sie *CQPweb*.

1. Unter Corpus Info auf View corpus documentation klicken.



Im neuen Fenster wird die Korpusdokumentation angezeigt.



Öffnen Sie *View corpus documentation* per Rechtsklick in einem separaten Tab. So haben Sie gleichzeitig Zugriff auf *CQPweb* und die entsprechende Korpusdokumentation.

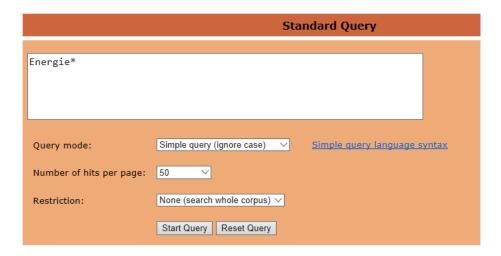
4.5 Suchanfrage erstellen

In diesem Kapitel wird beschrieben, wie eine Suchanfrage erstellt werden kann.

Es gibt zwei Arten von Suchanfragen: Die normale Suchanfrage (Standard Query) und die eingeschränkte Suchanfrage (Restricted Query).

Ausserdem gibt es zwei Syntaxen, wie Suchbegriffe geschrieben werden können: Die **Simple Query Syntax***^G und die **CQP Syntax***^G.

Standard Query: Normale Suchanfrage erstellen



↑X xò Führen Sie die nachfolgenden Schritte aus, um eine normale Suchanfrage zu erstellen:

- 1. Auf Standard Query in der Menüleiste klicken.
- Query mode auswählen.Es gibt drei Arten von Query modes:
- CQP syntax: Suchbegriff wird mit CQP Syntax*^G eingegeben. Gross-/Kleinschreibung wird berücksichtigt, kann aber mit Hilfe des regulären Ausdrucks %c ignoriert werden (siehe Abschnitt CQP Syntax, S. 19).
- **Simple query (ignore case):** Suchbegriff wird mit Simple Query Syntax*^G eingegeben. Gross-/Kleinschreibung wird ignoriert (siehe Abschnitt Simple Query Syntax, S. 18).
- **Simple query (case-sensitive):** Suchbegriff wird mit Simple Query Syntax*^G eingegeben. Gross-/Kleinschreibung wird berücksichtigt (siehe Abschnitt Simple Query Syntax, S. 18).
- 3. Suchbegriff im Suchfeld eingeben.
- 4. Bei *Number of hits per page* auswählen, wie viele Suchtreffer pro Seite angezeigt werden sollen.
- 5. Bei **Restriction** auswählen, ob die Suchanfrage auf einzelne Quellen oder auf ein selbsterstelltes Subkorpus eingeschränkt werden soll (siehe Kap. 4.8 Subkorpora erstellen, S. 29).
- Auf Start Query klicken. Die Suche kann, je nach Komplexität der Anfrage und gerade gleichzeitig arbeitender Nutzer, einige Zeit beanspruchen.



Nach der Suche werden alle Suchergebnisse mit der jeweiligen Quelle angezeigt.

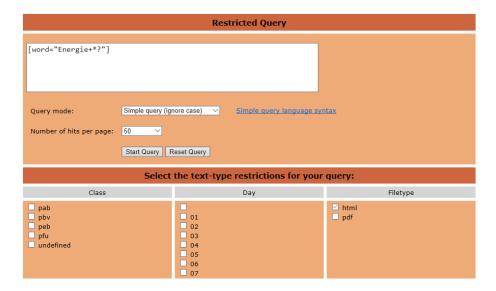


Wenn Sie die Suchanfrage noch weiter einschränken wollen, erstellen Sie eine eingeschränkte Suchanfrage oder ein eigenes Subkorpus (siehe Kap. 4.8 Subkorpora erstellen, S. 29).



Im Auswahlmenü oben rechts können Sie eine neue Suche starten oder die Suchergebnisse weiter analysieren (siehe Kap. 4.11 Korpus untersuchen, S. 40).

Restricted Query: Eingeschränkte Suchanfrage erstellen





Führen Sie die nachfolgenden Schritte aus, um eine eingeschränkte Suchanfrage zu erstellen:

- 1. Auf Restricted Query in der Menüleiste klicken.
- Query mode auswählen.Es gibt drei Arten von Query modes:
- CQP syntax: Suchbegriff wird mit CQP Syntax*^G eingegeben. Gross-/Kleinschreibung wird berücksichtigt, kann aber mit Hilfe des regulären Ausdrucks %c ignoriert werden (siehe Abschnitt CQP Syntax, S. 19).
- **Simple query (ignore case):** Suchbegriff wird mit Simple Query Syntax*⁶ eingegeben. Gross-/Kleinschreibung wird ignoriert (siehe Abschnitt Simple Query Syntax, S. 18).
- **Simple query (case-sensitive):** Suchbegriff wird mit Simple Query Syntax*^G eingegeben. Gross-/Kleinschreibung wird berücksichtigt (siehe Abschnitt Simple Query Syntax, S. 18).
- 3. Suchbegriff im Suchfeld eingeben.
- 4. Bei *Number of hits per page* auswählen, wie viele Suchtreffer pro Seite angezeigt werden sollen.
- In den unteren Feldern die Einschränkungen für die Suchanfrage festlegen (siehe Kap. 5.2 Metadaten, S. 64 für die Erklärung der Begriffe der Einschränkung).
- 6. Auf Start Query klicken.



Nach der Suche werden alle Suchergebnisse mit der jeweiligen Quelle angezeigt.



Die Suche kann, je nach Komplexität der Anfrage und gerade gleichzeitig arbeitender Nutzer, einige Zeit beanspruchen.



Im Auswahlmenü oben rechts können Sie eine neue Suche starten oder die Suchergebnisse weiter analysieren (siehe Kap. 4.11 Korpus untersuchen, S. 40).

Simple Query Syntax

Beim Suchen mit der Simple Query Syntax*^G wird ein Suchbegriff grundsätzlich normal, z. B. wie in einer Suchmaschine eingegeben.

Zum Beispiel:

Um das Wort *Energie* im Korpus zu suchen, Energie im Suchfeld eingeben.

Um die Suche zu erweitern, können zusätzlich zum Suchbegriff sogenannte Platzhaltersymbole benutzt werden.



Die Platzhaltersymbole in der Simple Query Syntax haben zum Teil eine andere Bedeutung als die regulären Ausdrücke in der CQP Syntax. Zum Beispiel hat das Symbol «?» in beiden Systemen eine andere Funktion.



Für einige Funktionen brauchen Sie sogenannte Tags (siehe Kap. 5.3 Tagset, S. 65).

In der folgenden Tabelle sind die wichtigsten Platzhaltersymbole aufgeführt:

Symbol	Bedeutung
?	Ein beliebiges Zeichen: erneuerbare? → erneuerbaren, erneuerbarem, erneuerbarer,
+	Ein beliebiges Zeichen oder mehrere beliebige Zeichen: Energie+ → Energien, Energiestrategie, aber nicht einfach Energie
*	Keine weiteren Zeichen oder mehrere beliebige Zeichen: Energie* → Energie, Energien, Energiestrategie,
[x,y]	Findet separat jedes Element in den eckigen Klammern: Energie[strategie, system, - Zukunft] → Energiestrategie, Energiesystem, Energie-Zukunft (und nichts anderes). Kann auch «nichts» finden als Ergebnis: Energie[n,] → Energie, Energien
Wort_PO S	Findet ein spezifisches Wort mit einer bestimmten Wortart: energ*_ADJA → energiepolitische, energieintensivem, energetischer,

POS	finde alle Wörter einer Wortart: $\texttt{ADJA} \rightarrow \textit{junge},$ schnelle, brasilianische,
X«3«Y	Findet das Wort X in einem Abstand von 0 bis 3 Wörtern rechts vom Wort Y.
X»3»Y	Findet das Wort X in einem Abstand von 0 bis 3 Wörtern links vom Wort Y.
X«3»Y	Findet das Wort X in einem Abstand von 0 bis 3 Wörtern rechts und links vom Wort Y.
\	Buchstäbliche Interpretation von Zeichen, die sonst <i>CQPweb</i> -Operatoren wären: Erneuerbare\? → Findet buchstäblich <i>Erneuerbare?</i>
X(Y){0,3}	Findet das Wort X mit 0 bis 3 Wiederholungen des Elements Y: $Energie(\.) \{1,3\} \rightarrow Energie.,$ Energie
{X}	Findet alle Wörter, deren Lemma X ist: {Haus} → Haus, Häuser, Häusern,

CQP Syntax

Beim Suchen mit der CQP Syntax*^G wird ein Suchbegriff in eckigen Klammern und mit Deklarierung der gewünschten Annotationsebene eingegeben.

Zum Beispiel:

- Um die Wortform «Energien» im Korpus zu suchen, [word="Energien"] im Suchfeld eingeben.
- Um alle Wörter mit dem Lemma «Energie» im Korpus zu suchen, [lemma="Energie"] im Suchfeld eingeben.
- Um nach einer spezifischen Wortart (z. B. allen Substantiven) im Korpus zu suchen, [pos="NN"] im Suchfeld eingeben (POS – part of speech*^G).
- Um nach Eigennamen, z.B. denen von Personen zu suchen, [entitytype=".*PER"] im Suchfeld eingeben.

Um die Suche zu erweitern, können zusätzlich zum Suchbegriff sogenannte reguläre Ausdrücke*^G (RegEx) benutzt werden.



Die regulären Ausdrücke in der CQP Syntax haben zum Teil eine andere Bedeutung als die Platzhaltersymbole in der Simple Query Syntax. Zum Beispiel hat das Symbol «?» in beiden Systemen eine andere Funktion.



Für die Suche nach Eigennamen und Wortarten brauchen Sie sogenannte Tags (siehe Kap. 5.3 Tagset, S. 65). In der folgenden Tabelle sind die wichtigsten regulären Ausdrücke aufgeführt:

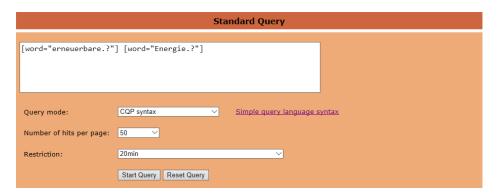
RegEx	Bedeutung
	Ein weiteres, beliebiges Zeichen: [word="erneuerbare"] → erneuerbaren, erneuerbarem, erneuerbarer,
+	Wiederholungsoperator: Das vorangehende Zeichen muss mindestens einmal vorkommen, darf aber auch mehrmals vorkommen (bis zur Wortgrenze); z.B. sinnvoll in der Kombination mit «.»: [word="Energie.+"] → Energien, Energiegesetz, Energiestrategien,
*	Wiederholungsoperator: Das vorangehende Zeichen darf beliebig oft, d.h. auch keinmal, vorkommen (bis zur Wortgrenze); z.B. sinnvoll in der Kombination mit «.»: [word="Energie.*"] → Energie, Energien, Energiegesetz, Energiestrategien,
?	Wiederholungsoperator: Das vorangehende Zeichen kann, muss aber nicht vorhanden sein: [word="energie-?effizient"] → energie- effizient, energieeffizient,
(x y)	Findet separat jedes Element in den runden Klammern. [word="(E e)nergie- ?(Z z)ukunft"] → Energiezukunft, energiezukunft, Energie-zukunft, Energie-Zukunft,
x{0,3}	Findet 0 bis 3 Wiederholungen des Elements davor. Mehrere Elemente können in eckigen Klammern eingegeben werden. → [xyz] {0,3}
0	Zusätzliches unbestimmtes Wort zwischen Suchbegriffen.: [word="die"] [] [word="Energie"] → die graue Energie, die saubere Energie,
[^]	Die Zeichen in den Klammern nach ^ dürfen nicht vorkommen.: [word=".*?[^eEÉé]nergie"] → Minergie, Synergie, aber nicht Energie
!	Der gesamte Ausdruck nach dem Ausrufezeichen darf nicht vorkommen: [pos!= "NN"] → Als Ergebnis kommen alle Wortarten ausser Nomina (NN).

&	Kombiniert die pos-, word-, lemma- und Eigennamen-Eigenschaften eines Suchwortes: [word="richtig" & pos="ADV"] → Okkurrenzen von <i>richtig</i> nur als Adverb, nicht z. B. als Adjektiv.
\	Buchstäbliche Interpretation von Zeichen, die sonst CQP-Platzhalter wären: [word="Erneuerbare"] [word="\?"] → Findet buchstäblich Erneuerbare?
%c	Gross-/Kleinschreibung wird nicht berücksichtigt: [word="energie"%c] → Energie, energie

Szenario: Erneuerbare Energien

Sie fragen sich, ob Schweizer beim Zeitunglesen mit dem Thema» «erneuerbare Energien» in Berührung kommen. Eine Tageszeitung, die von vielen Schweizern gelesen wird, ist 20 Minuten. Um sich ein Bild zu machen, ob dort je von erneuerbaren Energien die Rede war, wollen Sie eine kurze Suchanfrage erstellen. Dazu führen Sie nachfolgende Schritte aus:

- 1. Auf Standard Query in der Menüleiste klicken.
- 2. Query Mode CQP Syntax auswählen.
- 3. Im Suchfeld [word="erneuerbare.?"] [word="Energie.?"] eingeben.
- > Die regulären Ausdrücke*G «.» und «?» in den eckigen Klammern bedeuten, dass ein weiteres beliebiges Zeichen dort stehen darf (und nicht muss). Kasusformen werden folglich auch gesucht.
 - 4. Bei Number of hits per page 50 auswählen.
 - 5. Bei *Restriction* 20min auswählen.



6. Auf Start Query klicken.



Die Suchergebnisse werden angezeigt.





In der ersten Zeile wird angezeigt, wie viele Suchergebnisse (matches) in wie vielen Texten gefunden wurden.



Im Auswahlmenü oben rechts können Sie eine neue Suche starten oder die Suchergebnisse weiter analysieren (siehe Kap. 4.11 Korpus untersuchen, S. 40).

4.6 Suchanfrage speichern

In diesem Kapitel wird beschrieben, wie eine Suchanfrage innerhalb von *CQPweb* gespeichert und zu einem späteren Zeitpunkt aufgerufen werden kann. Dies ist auch nach einem erneuten Einloggen möglich.

Das Abspeichern einer Suchanfrage ist zu empfehlen, wenn die ursprüngliche Suchanfrage eingeschränkt wird. Das Speichern der Suchanfrage ohne zusätzliche Einschränkung ist nicht empfehlenswert, da es einfacher ist, für ein erneutes Aufrufen der Suchanfrage nochmals über die Suchfunktion nach dem Begriff zu suchen.

Suchanfrage speichern



Führen Sie die nachfolgenden Schritte aus, um die Suchanfrage zu speichern:

- Suchanfrage erstellen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15).
- 2. Im Auswahlmenü oben rechts **Save current set of hits** auswählen und daneben mit **Go!** bestätigen.
- → Das Fenster mit einem Feld für die Namensgebung der Suchanfrage erscheint.
 - 3. Gewünschten Speichernamen in das Feld eingeben.
 - Auf Save the query rechts neben dem Feld für die Namensgebung klicken.





Die Suchanfrage wurde gespeichert.



Der Name darf nur aus Buchstaben, Ziffern und Unterstrichen bestehen. Sind andere Sonderzeichen enthalten, kann die Suchanfrage nicht gespeichert werden.

Suchanfrage aufrufen



Führen Sie die nachfolgenden Schritte aus, um eine gespeicherte Suchanfrage aufzurufen:

- 1. Das Feld **New query** im Auswahlmenü oben rechts auswählen und daneben mit **Go!** bestätigen.
- → Die Startseite für die Suchanfrage öffnet sich.
 - 2. Im Menü links auf Saved queries klicken.
- → Alle gespeicherten Suchanfragen werden unter dem entsprechend gewählten Speichernamen rechts gelistet angezeigt. Angegeben sind ausserdem die Anzahl der Treffer zur Suchanfrage und das Datum, an dem die Suchanfrage ursprünglich abgespeichert wurde.
 - 3. Auf den Namen der benötigten Liste klicken, um die Suchanfrage anzeigen zu lassen.



Die entsprechende Suchanfrage wird angezeigt.

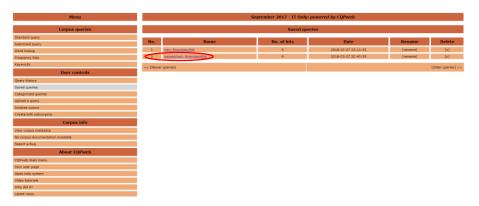
Speichernamen ändern

Führen Sie die nachfolgenden Schritte aus, um den Speichernamen der gespeicherten Suchanfrage zu ändern:

- 1. Das Feld **New query** im Auswahlmenü oben rechts auswählen und daneben mit **Go!** bestätigen.
- → Die Startseite für die Suchanfrage öffnet sich.
 - 3. Bei der entsprechenden gespeicherten Suchanfrage auf das Feld [rename] klicken.
- → Das Fenster mit dem Eingabefeld für den neuen Speichernamen erscheint.
 - 4. Neuen Speichernamen für die Suchanfrage eingeben.



Die gespeicherte Suchanfrage wird unter dem neuen Namen angezeigt.



Suchanfrage löschen

Führen Sie den nachfolgenden Schritt aus, um eine gespeicherte Suchanfrage zu löschen:



ACHTUNG: Durch das Ausführen des nachfolgenden Handlungsschritts wird die Suchanfrage sofort gelöscht, ohne dass ein weiteres Fenster zur Bestätigung des Vorgangs erscheint.

1. In der Liste mit den **Saved queries** bei der zu entfernenden Suchanfrage auf [x] klicken.



Die gespeicherte Suchanfrage wurde gelöscht.

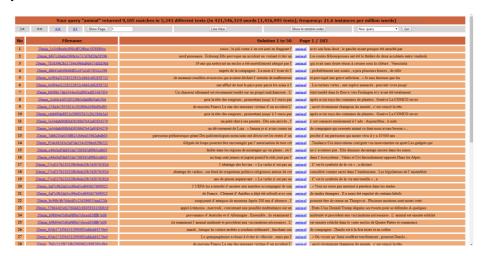
4.7 Darstellung der Suchergebnisse verändern

In diesem Kapitel wird beschrieben, welche Optionen zur Darstellung von Suchergebnissen zur Verfügung stehen. Es kann zwischen zwei verschiedenen Ansichten ausgewählt werden: *KWIC View**^G oder *Line View**^G. Ausserdem können die Suchergebnisse in diesen Ansichten jeweils sortiert nach Medium angezeigt werden oder die Treffer können zufällig sortiert erscheinen.

KWIC-View: Kontextanzeige

In der Standardeinstellung werden die Suchergebnisse in der *KWIC View**^G angezeigt. Die Ergebnisse erscheinen sortiert nach Zielmedium.

Die Trefferliste wird zentriert angezeigt. Dabei befindet sich in der Anzeige in der Mitte der Zeile das gesuchte Wort, links und rechts davon jeweils die kontextuellen Wörter:





Um die Ergebnisse zufällig sortiert anzuzeigen, in der Zeile unterhalb der Frequenzinformationen auf den Button **Show in random order** klicken.

Line View: Zeilenansicht

Führen Sie die nachfolgenden Schritte aus, um die Suchergebnisse in der *Line View**^G anzuzeigen:

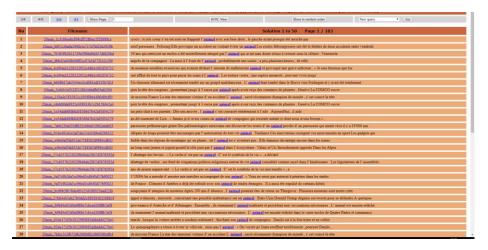
- Suchanfrage erstellen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15).
- 2. In der Zeile unterhalb der Leiste mit den Frequenzinformationen auf *Line View* klicken.



Die Suchergebnisse werden in der *Line View* angezeigt.



Standardmässig werden die Ergebnisse sortiert nach Zielmedium angezeigt. Es gibt keine zentrierte Darstellung der Ergebnisse wie in der Kontextanzeige, sondern die Anzeige ist linksbündig ausgerichtet, das Suchwort ist unterstrichen und in Fettschrift.





Um die Ergebnisse zufällig sortiert anzuzeigen, in der Zeile unterhalb der Frequenzinformationen auf den Button **Show in random order** klicken.

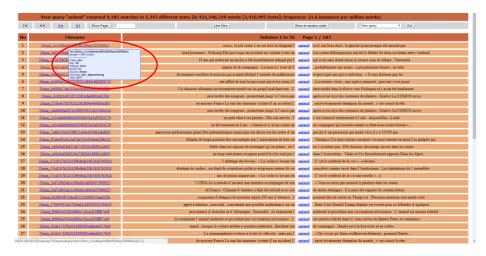
Anzeige der Metadaten

Es besteht die Möglichkeit, die Metadaten zum entsprechenden Quelltext in einer Schnellansicht oder in einer Vollansicht anzeigen zu lassen.

Schnellansicht anzeigen

Führen Sie den nachfolgenden Schritt aus, um die Schnellansicht anzuzeigen:

1. Mit der Maus über den Dateinamen fahren:



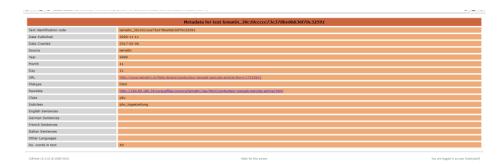


Die Quelle des Textes ist ersichtlich, ausserdem werden die Textklasse und die Unterklasse, das Erscheinungsdatum, die Anzahl Wörter sowie das Datenformat angezeigt.

Vollansicht anzeigen

Führen Sie die nachfolgenden Schritte aus, um die Vollansicht anzuzeigen:

- 1. Mit der Maus über den Dateinamen fahren.
- 2. Auf den verlinken Dateinamen links in der Anzeige klicken.



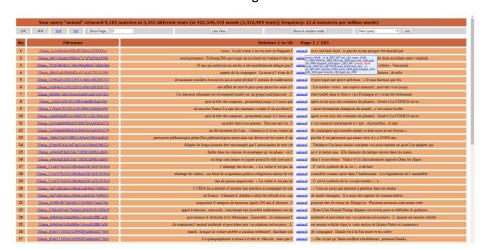


Die Metadaten werden in der Vollansicht angezeigt. Zusätzlich zu den oben erwähnten Daten wird auch das Datum der Aufnahme des Textes in das Tool erwähnt.

Anzeige kontextueller Informationen Es besteht die Möglichkeit, sich inhaltliche und sprachformelle Kontextinformationen zum Suchwort beziehungsweise Suchtext anzeigen zu lassen.

Wortarten der Kontextwörter anzeigen Führen Sie den nachfolgenden Schritt aus, um Informationen bezüglich der Wortarten der Kontextwörter anzuzeigen:

1. Den Cursor über den Suchbegriff halten.





Die Wortarten der entsprechenden Wörter werden angezeigt.

Inhaltlichen Kontext anzeigen

Führen Sie den nachfolgenden Schritt aus, um einen zusammenfassenden inhaltlichen Kontext zur Textquelle anzuzeigen:

1. Auf den Suchbegriff klicken:





Eine inhaltliche Zusammenfassung der Textquelle wird angezeigt.

4.8 Subkorpora erstellen

Das Hauptkorpus auf Swiss-AL ist für alle Nutzer dasselbe. Aus Teilmengen des Swiss-AL können jedoch mittels bestimmten Parametern benutzerspezifische Subkorpora*⁶ erstellt werden. Dies kann mit unterschiedlichen Methoden für jeweils unterschiedliche Zwecke geschehen. In diesem Kapitel werden die möglichen Methoden sowie ihre jeweiligen Anwendungsbereiche beschrieben.

Texts found in a saved query:
Via Suchanfrage
erstellen

Bei dieser Methode wird eine normale oder eingeschränkte Suchanfrage gespeichert, die danach als Subkorpus weiter analysiert werden kann. Diese Methode lässt die meisten benutzerspezifischen Einschränkungen und Einstellungen zu. Sie eignet sich für Subkorpora, die Texte mit bestimmten Stichworten enthalten sollen.



Wird eine eingeschränkte Suche (siehe Kap. 4.5 Suchanfrage erstellen, S. 15) ausgeführt, speichert CQPweb die Suchergebnisse automatisch als Subkorpus mit dem Titel "Last restrictions" ab, das bei jeder durchgeführten eingeschränkten Suche neu überschrieben wird. Das Korpus "Last Restrictions" kann weder angesehen noch durchsucht werden, deshalb ist es für weitere Analysen ungeeignet und wird nicht weiter beschrieben. In diesem Kapitel wird nur die weiterverwendbare Methode beschrieben.

Führen Sie die nachfolgenden Schritte aus, um ein Subkorpus via Suchanfrage zu erstellen:

- Normale oder eingeschränkte Suchanfrage erstellen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15).
- 2. Suchanfrage speichern (siehe Kap. 4.6 Suchanfrage speichern, S. 23).
- 3. In der Menüleiste auf *Create/Edit Subcorpora* klicken.
- 4. Bei *Define new subcorpus via* im Auswahlmenü *Texts found in a saved query* auswählen und mit *Go!* bestätigen.
- → Das Untermenü **Design a new subcorpus** wird angezeigt.
 - 5. Bei *Please enter a name for your new subcorpus* im Eingabefeld einen Namen für das neue Subkorpus eintragen.
- > Ohne Namen kann das Subkorpus nicht erstellt werden. Der Name darf nur aus Buchstaben, Ziffern und Unterstrichen bestehen. Sind andere Sonderzeichen enthalten, kann das Subkorpus nicht erstellt werden.
 - Bei Which Saved Query do you want to use as the basis of the subcorpus? im Auswahlmenü die betreffende gespeicherte Suchanfrage auswählen.



Um das Korpus ohne zusätzliche Selektion von einzelnen Texten zu erstellen:

7. (a) Auf Create Subcorpus from selected query klicken.



Das Subkorpus wird unter *Existing Corpora* angezeigt.



Um vor dem Erstellen des Subkorpus die enthaltenen Texte weiter zu selektieren:

- 7. (b) Auf Get list of texts klicken.
- → Die Liste der ins Subkorpus aufzunehmenden Texte öffnet sich.
 - 8. In der Spalte *Include in Subcorpus* die Texte auswählen, die ins Subkorpus aufgenommen werden sollen.
 - 9. Bei *Add files to subcorpus* im Auswahlmenü *Use specified name for new subcorpus* auswählen.
 - 10. Bei **New subcorpus** im Eingabefeld nochmals den bei Schritt 2 eingegebenen Namen eingeben.
 - 11. Auf Add texts klicken.



Das Subkorpus wird unter Existing Subcorpora angezeigt.

Corpus Metadata: Via eingeschränkte Metadaten erstellen Bei dieser Methode kann in den vorhandenen Metadatenkategorien des Korpus nach jenen Kriterien gefiltert werden, nach denen das Subkorpus erstellt werden soll. Die möglichen Metadatenkategorien (siehe Kap. 5.2 Metadaten, S. 64) sind:

- Day
- Month
- Year
- Class
- Subclass
- Filetype

Die Methode via Korpus-Metadaten eignet sich gut zur Erstellung von quellen-, sprach- und/oder jahresspezifischen Subkorpora. So kann zum Beispiel ein Subkorpus von allen Texten der Quelle «nzz» aus dem Jahr «2015» erstellt werden.

Führen Sie die nachfolgenden Schritte aus, um ein Subkorpus via eingeschränkte Metadaten zu erstellen:

- 1. In der Menüleiste auf *Create/Edit Subcorpora* klicken.
- 2. Bei *Define new subcorpus via* im Auswahlmenü *Corpus Metadata* auswählen und mit *Go!* bestätigen.
- → Das Untermenü **Design a new subcorpus** wird angezeigt.



- Bei Please enter a name for your new subcorpus im Eingabefeld einen Namen für das neue Subkorpus eintragen.
- > Ohne Namen kann das Subkorpus nicht erstellt werden. Der Name darf nur aus Buchstaben, Ziffern und Unterstrichen bestehen. Sind andere Sonderzeichen enthalten, kann das Subkorpus nicht erstellt werden.
 - 4. Bei **Select text-type restrictions for your subcorpus** die gewünschten Kästchen in den Metadatenkategorien mittels Klicken anwählen.



Um das Korpus ohne zusätzliche Selektion von einzelnen Texten zu erstellen:

5. (a) Auf Create Subcorpus from selected categories klicken.



Das Subkorpus wird unter Existing Corpora angezeigt.



Um vor dem Erstellen des Subkorpus die enthaltenen Texte weiter zu selektieren:

- 5. **(b)** Auf *Get list of texts* klicken.
- → Die Liste der ins Subkorpus aufzunehmenden Texte öffnet sich.
 - In der Spalte *Include in Subcorpus* die Texte auswählen, die ins Subkorpus aufgenommen werden sollen. Wenn alle Texte aufgenommen werden sollen, oberhalb der Liste die Option *Include all texts* auswählen.
 - 7. Bei *Add files to subcorpus* im Auswahlmenü *Use specified name for new subcorpus* auswählen.
 - 8. Bei *New subcorpus* im Eingabefeld nochmals den bei Schritt 2 eingegebenen Namen eingeben.
 - 9. Auf Add texts klicken.



Das Subkorpus wird unter *Existing Subcorpora* angezeigt.

Scan text metadata: Via Metadaten-Scan erstellen Bei dieser Methode kann eine einzelne Metadatenkategorie nach einem Suchbegriff durchsucht werden, nach dem das Subkorpus erstellt werden soll. Diese Methode eignet sich für Subkorpora, für die nur eine einzige Metadatenkategorie entscheidend ist, die jedoch noch spezifisch durchfiltert werden muss. So kann z.B. ein Subkorpus erstellt werden, das alle Quellen enthält, die mit "n" anfangen.

Führen Sie die nachfolgenden Schritte aus, um ein Subkorpus via Metadaten-Scan zu erstellen:

- 1. In der Menüleiste auf Create/Edit Subcorpora klicken.
- 2. Bei *Define new subcorpus via* im Auswahlmenü *Scan text-metadata* auswählen und mit *Go!* bestätigen.
- → Das Untermenü **Design a new subcorpus** wird angezeigt.



- Bei Which metadata field do you want to search? im Auswahlmenü die Textmetadatenkategorie auswählen, die durchsucht werden soll (siehe Kap. 5.2 Metadaten, S. 64).
- 4. Im Eingabefeld den gewünschten Suchbegriff eingeben, nach dem die Metadatenkategorie durchsucht werden soll.
- 5. Bei **Search for texts where this metadata field...** im Auswahlmenü auswählen, was gesucht wird:
 - **Starts with:** Eingegebener Suchbegriff steht am Anfang des Metadatentextes.
 - Ends with: Eingegebener Suchbegriff steht am Ende des Metadatentextes.
 - **Contains**: Eigegebener Suchbegrifft steht an beliebiger Stelle im Metadatentextes.
 - Matches exactly: Eingegebener Suchbegriff entspricht dem Metadatentext exakt.
- 6. Auf Get list of texts klicken.
- → Die Liste der ins Subkorpus aufzunehmenden Texte öffnet sich.
 - In der Spalte *Include in Subcorpus* die Texte auswählen, die ins Subkorpus aufgenommen werden sollen. Wenn alle Texte aufgenommen werden sollen, oberhalb der Liste die Option *Include all texts* auswählen.
 - 8. Bei *Add files to subcorpus* im Auswahlmenü *Use specified name for new subcorpus* auswählen.
 - 9. Bei **New subcorpus** im Eingabefeld einen Namen für das neue Subkorpus eingeben.

- > Ohne Namen kann das Subkorpus nicht erstellt werden. Der Name darf nur aus Buchstaben, Ziffern und Unterstrichen bestehen. Sind andere Sonderzeichen enthalten, kann das Subkorpus nicht erstellt werden
 - 10. Auf Add texts klicken.



Das Subkorpus wird unter Existing Corpora angezeigt.

Manual entry of filenames:
Manuell erstellen

Bei der manuellen Erstellung eines Subkorpus werden Texte aus dem Korpus anhand ihrer Dateinamen zusammengestellt. Diese werden vom Nutzer einzeln eingegeben und müssen somit vorangehend bekannt sein.

Die manuelle Erstellung eignet sich üblicherweise nicht für die Zusammenstellung von grossen Subkorpora, da sie einen hohen Aufwand darstellt. Sie eignet sich für kleinere Subkorpora, die sich aufgrund von abweichenden Metadaten nicht automatisch erstellen lassen oder für Fälle, in denen wenige Texte miteinander verglichen werden sollen.

Führen Sie die nachfolgenden Schritte aus, um ein Subkorpus manuell zu erstellen:



Voraussetzung: Die Dateinamen der aufzunehmenden Korpustexte sind bekannt.

- 1. In der Menüleiste auf *Create/Edit Subcorpora* klicken.
- Bei Define new subcorpus via im Auswahlmenü Manual entry of filenames wählen und mit Go! bestätigen.
- → Das Untermenü **Design a new subcorpus** wird angezeigt.
 - Bei Please enter a name for your new subcorpus im Eingabefeld einen Namen für das neue Subkorpus eintragen.
- > Ohne Namen kann das Subkorpus nicht erstellt werden. Der Name darf nur aus Buchstaben, Ziffern und Underline bestehen. Sind andere Sonderzeichen enthalten, kann das Subkorpus nicht erstellt werden.



Bei Enter the filenames... im Eingabefeld nacheinander die Dateinamen der gewünschten Texte eintragen. Die Dateinamen hintereinander auflisten und mittels Komma trennen. Zum Beispiel: 20min_e0618789a16a1d71e955a828f7e748e4, 20min_5417aa9c255645f37b1717bd5e52c13e,

5. Auf Create subcorpus klicken.



Das Subkorpus wird unter *Existing Corpora* angezeigt.

Invert an existing subcorpus: Via Inversion erstellen Die Inversion eines Subkorpus hat zur Folge, dass in einem zweiten Subkorpus sämtliche Texte des Korpus aufgenommen werden, ausser denjenigen, die sich im bereits existierenden Subkorpus befinden. So werden sozusagen zwei Teile des übergeordneten Korpus erstellt.

Ein via Inversion erstelltes Subkorpus eignet sich für Fälle, in denen ein Ausschnitt des Korpus mit dem gesamten Rest verglichen werden soll (in dem der Ausschnitt nicht mehr vorkommt).



Je nach Grösse des Korpus und Subkorpus kann diese Methode einige Zeit beanspruchen.

Führen Sie die nachfolgenden Schritte aus, um ein Subkorpus via Inversion zu erstellen:



Voraussetzung: Ein zu invertierendes Subkorpus ist bereits vorhanden.

- 1. In der Menüleiste auf Create/Edit Subcorpora klicken.
- 2. Bei *Define new subcorpus* im Auswahlmenü *Invert an existing corpus* auswählen und mit *Go!* bestätigen.
- → Das Untermenü **Design a new subcorpus** öffnet sich.
 - 3. Bei *Please enter a name for your new subcorpus* im Eingabefeld einen Namen für das neue Subkorpus eintragen.
- > Ohne Namen kann das Subkorpus nicht erstellt werden. Der Name darf nur aus Buchstaben, Ziffern und Unterstrichen bestehen. Sind andere Sonderzeichen enthalten, kann das Subkorpus nicht erstellt werden
 - 4. Bei **Select** auf das zu invertierende Subkorpus klicken.
 - 5. Auf *Create inverted subcorpus* klicken. Je nach Grösse des Subkorpus, kann die Inversion einige Zeit beanspruchen!



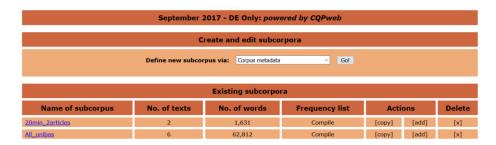
Das Subkorpus wird unter **Existing Corpora** angezeigt.

Create a subcorpus for every text: Subkorpus für jeden Text erstellen Diese Methode erstellt für jeden einzelnen im Korpus aufgenommenen Text ein einzelnes Subkorpus. Bei einem Korpus mit vielen Texten kann das sehr lange dauern. Für jeden einzelnen Text ein eigenes Subkorpus zu erstellen, eignet sich im Falle des Swiss-AL nicht, da das Korpus zu viele Texte enthält. Die Methode wird hier deshalb nicht weiter beschrieben.

4.9 Subkorpora verwalten

In diesem Kapitel wird beschrieben, wie sich bereits erstellte benutzerspezifische Subkorpora verwalten lassen.

Subkorpora anzeigen



Führen Sie den nachfolgenden Schritt aus, um eine Übersicht der bereits erstellten Subkorpora anzuzeigen:

1. In der Menüleiste auf *Create/Edit Subcorpora* klicken.



Im neuen Fenster wird eine Liste mit den verfügbaren Subkorpora angezeigt.



Die Übersicht der Subkorpora finden Sie unter **Existing subcorpora** unten auf der Bildschirmanzeige. Angezeigt werden die (selbst erstellten) Namen, die Anzahl der beinhalteten Texte, die Anzahl der beinhalteten Wörter sowie einige mögliche Aktionen zur Verwaltung der Korpora.

Frequenzlisten kompilieren

Um Frequenzlisten*⁶ (siehe Kap. 4.13 Frequency lists: Frequenzlisten, S. 51) für ein Subkorpus erstellen zu können, müssen sie zuerst kompiliert werden



Bei sehr umfangreichen Subkorpora ist diese Funktion nicht verfügbar.



Führen Sie den nachfolgenden Schritt aus, um die Frequenzliste für ein bestimmtes Subkorpus zu kompilieren:



Voraussetzung: In der Spalte **Frequency list** steht beim betreffenden Subkorpus **Compile**. Steht stattdessen **N/A**, ist diese Funktion für dieses Subkorpus nicht verfügbar.

1. Beim gewünschten Subkorpus in der Spalte *Frequency list* auf *Compile* klicken.



Anstatt **Compile** wird nun **Available** angezeigt. Die Frequenzliste ist nun verfügbar.

Subkorpus durchsuchen

Um ein Subkorpus nach Wörtern oder Wortformen zu durchsuchen, gehen Sie ähnlich vor, wie bei einer normalen Suchanfrage, die auf das Subkorpus eingeschränkt wird.

Führen Sie die nachfolgenden Schritte aus, um ein Subkorpus zu durchsuchen:

- 1. Schritte 1 bis 4 einer normalen Suchanfrage ausführen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15)
- 2. Bei *Restrictions* im Auswahlmenü das zu durchsuchende Subkorpus auswählen (ganz unten).
- 3. Auf Start query klicken.



Nach der Suche werden alle Suchergebnisse im Subkorpus angezeigt.

Subkorpus kopieren

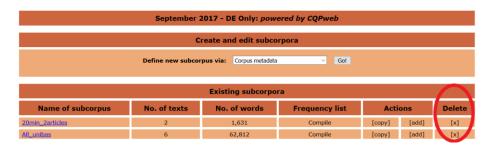
Um zum Beispiel ein Subkorpus zu bearbeiten, ohne das Original zu verlieren, kann eine Kopie des Subkorpus erstellt werden. Führen Sie dazu die nachfolgenden Schritte aus:

- 1. In der Übersicht *Existing subcorpora* beim gewünschten Subkorpus in der Spalte *Actions* auf *copy* klicken.
- → Die Maske Copying subcorpus öffnet sich.
 - 2. Bei *Please enter a name for your new subcorpus* im Eingabefeld einen Namen für das neue Subkorpus eintragen.
- > Ohne Namen kann das Subkorpus nicht erstellt werden. Der Name darf nur aus Buchstaben, Ziffern und Unterstrichen bestehen. Sind andere Sonderzeichen enthalten, kann das Subkorpus nicht erstellt werden.
 - 3. Auf Copy subcorpus klicken.



Die Kopie des Subkorpus wird unter *Existing subcorpora* angezeigt.

Subkorpus löschen



Führen Sie die folgenden Schritte aus, um ein Subkorpus zu löschen:



ACHTUNG: Das Subkorpus wird ohne Nachfrage und unwiderruflich gelöscht!

1. In der Übersicht *Existing subcorpora* beim gewünschten Subkorpus in der Spalte *Delete* auf das x klicken.



Das Subkorpus wird gelöscht.

4.10 Texte in Subkorpora verwalten

In diesem Kapitel wird beschrieben, wie die in ein Subkorpus aufgenommenen Texte nach dem Erstellen weiter verwaltet werden können.

Texte anzeigen

Führen Sie den nachfolgenden Schritt aus, um die Übersicht der in einem Subkorpus enthaltenen Texte anzuzeigen:

In der Übersicht Existing subcorpora unter Name of subcorpus auf das gewünschte Korpus klicken.



Die Liste der im Subkorpus enthaltenen Texte öffnet sich.

Texte zu Subkorpus hinzufügen

Es besteht die Möglichkeit, weitere Texte zu einem Subkorpus hinzuzufügen. Dies bezieht sich nur auf Texte, die sich bereits im Korpus befinden. Es können keine externen Texte zu Swiss-AL hinzugefügt werden. Um Texte zu einem Subkorpus hinzuzufügen, gibt es zwei Methoden:

Texte via Metadaten hinzufügen

Bei dieser Methode gehen Sie ähnlich vor, wie wenn Sie ein Subkorpus anhand von Metadaten erstellen. Sie fügen die Texte dann jedoch zu einem bestehenden Subkorpus hinzu (anstatt ein eigenes Subkorpus zu erstellen).

Führen Sie die nachfolgenden Schritte aus, um Texte anhand ihrer Metadaten zu einem Subkorpus hinzuzufügen:

- 1. Entweder Schritt 1 bis 4 des Kapitels Subkorpus via eingeschränkte Metadaten erstellen oder Schritt 1 bis 5 der Methode Subkorpus via Metadaten-Scan erstellen ausführen (siehe Kap. 4.8 Subkorpora erstellen, Seite 29).
- 2. Auf Get list of texts klicken.
- → Die Liste der im Subkorpus enthaltenen Texte wird angezeigt.
 - In der Reihe *Include in Subcorpus* die Texte auswählen, die ins Subkorpus aufgenommen werden sollen. Wenn alle Texte aufgenommen werden sollen, oberhalb der Liste die Option *Include* all texts auswählen.
 - 4. Bei *Add files to subcorpus* im Auswahlmenü das Subkorpus auswählen, in das die Texte aufgenommen werden sollen.
 - 5. Bei **New subcorpus** im Eingabefeld nochmals den bei Schritt 2 eingegebenen Namen eingeben.
 - 6. Auf Add texts klicken.

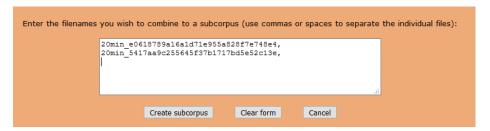
Texte manuell hinzufügen

Führen Sie die nachfolgenden Schritte aus, um Texte manuell zu einem Subkorpus hinzuzufügen:



Voraussetzung: Die Dateinamen der hinzuzufügenden Texte sind bekannt.

- In der Übersicht Existing subcorpora beim gewünschten Korpus in der Spalte Actions auf add klicken.
- → Die Maske *Adding texts to subcorpus* öffnet sich.



 Bei Enter the filenames... im Eingabefeld nacheinander die Identifikationsnummern der gewünschten Texte eintragen. Die Dateinamen hintereinander auflisten und mittels Komma trennen. Zum Beispiel:

```
20min_e0618789a16a1d71e955a828f7e748e4,
20min 5417aa9c255645f37b1717bd5e52c13e,
```

3. Um die Texte endgültig zum Subkorpus hinzuzufügen auf *Add texts to subcorpus* klicken.

Texte löschen

Führen Sie die nachfolgenden Schritte aus, um spezifische Texte aus einem Subkorpus zu löschen:

- 1. Übersicht der Texte im Subkorpus öffnen (siehe Kap. 4.10 Texte in Subkorpora verwalten, S. 37).
- 2. In der Spalte **Delete** die zu löschenden Texte auswählen.
- 3. Auf **Delete marked texts from corpus** im grauen Balken oberhalb der Liste klicken.



Nach dem Löschen von Texten aus einem Subkorpus muss die Frequenzliste dieses Subkorpus neu kompiliert werden.

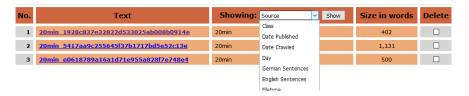
Weitere Metadaten anzeigen

Um die Metadaten der im Subkorpus angezeigten Texte anzuzeigen, gibt es zwei Möglichkeiten:



Führen Sie die nachfolgenden Schritte aus, um eine bestimmte Metadatenkategorie von allen Texten anzuzeigen:

- 1. Übersicht der Texte im Subkorpus öffnen (siehe Kap. 4.10 Texte in Subkorpora verwalten, S. 37).
- 2. In der Spalte **Showing** im Auswahlmenü die anzuzeigende Metadatenklasse auswählen.



Auf Show klicken.



In der Spalte **Showing** wird neu bei jedem Text die gewählte Metadatenkategorie angezeigt.



Führen Sie die nachfolgenden Schritte aus, um *alle* Metadaten eines einzelnen Texts anzuzeigen:

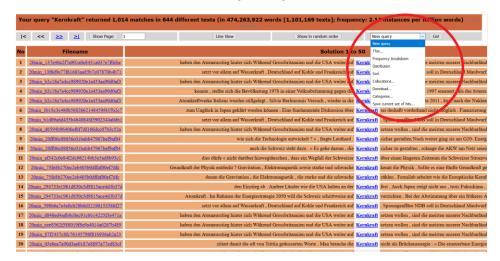
- 1. Übersicht der Texte im Subkorpus öffnen (siehe Kap. 4.10 Texte in Subkorpora verwalten, S. 37).
- 2. Auf den Dateinamen (Text) des gewünschten Texts klicken.



Die Metadaten des gewählten Texts werden in der Vollansicht angezeigt. Zusätzlich zu den Metadaten wird auch das Datum der Aufnahme des Textes in das Tool erwähnt.

4.11 Korpus untersuchen

In diesem Unterkapitel werden die verschiedenen Analysemöglichkeiten von Suchergebnissen beschrieben (siehe Kap. 4.5 Suchanfrage erstellen, S. 15). Diese Analysemöglichkeiten können dabei im Auswahlmenü oben rechts in der Bildschirmanzeige ausgewählt werden.



Suchergebnisse ausdünnen

Bei der Analysemöglichkeit *Thin* kann die Anzahl der angezeigten Suchergebnisse bestimmt werden.

Führen Sie die nachfolgenden Schritte aus, um die Anzahl der angezeigten Suchergebnisse zu bestimmen:

- Thin... im Auswahlmenü oben rechts auswählen und daneben mit Go! bestätigen.
- → Das Optionenmenü der Analysemöglichkeit *Thin* wird angezeigt.
 - 2. Die Art der zufälligen Auswahl bestimmen:
 - random (selection is reproducible) bei reproduzierbarer Auswahl
 - random (selection is not reproducible) bei nicht reproduzierbarer Auswahl
 - 3. Die Anzahl der angezeigten Suchergebnisse bestimmen:
 - für eine bestimmte Anzahl eine Zahl eintippen
 - für einen Prozentsatz eine Prozentzahl eintippen

Zum Beispiel: Um 20 Suchergebnisse zu erhalten, 20 eintippen. Um 20% der ursprünglichen Suchergebnisse zu erhalten, 20% eintippen.



4. Die Auswahl mit *Thin this query* bestätigen, um die tabellarische Ansicht der Suchergebnisse nach der Ausdünnung zu erhalten.



Die ausgedünnten Suchergebnisse werden angezeigt.



Die ausgedünnten Suchergebnisse können erneut mit den Analysemöglichkeiten im Auswahlmenü oben rechts untersucht werden.



Die ausgedünnten Suchergebnisse können heruntergeladen und auf CQPweb gespeichert (siehe Kap. 4.6 Suchanfrage speichern, S. 23) oder kategorisiert werden.

Frequenz der Suchergebnisse untersuchen

Bei der Analysemöglichkeit *Frequency breakdown* (Frequenzanalyse) kann die Häufigkeit von Wortformen oder Annotationen eines Suchbegriffs an einer ausgewählten *Breakdown position* (Position vom Suchbegriff aus) angezeigt werden. Der Suchbegriff wird bei dieser Analysemöglichkeit als *Node* bezeichnet.

Führen Sie die nachfolgenden Schritte aus, um die die Häufigkeit von Wortformen oder Annotationen eines Suchbegriffs an einer ausgewählten **Breakdown position** anzeigen zu lassen:

- Frequency breakdown im Auswahlmenü oben rechts auswählen und daneben mit Go! bestätigen.
- → Das Optionenmenü der Analysemöglichkeit Frequency breakdown wird angezeigt.
 - 2. Die *Breakdown position* bestimmen.
 - 3. Die Funktion bestimmen:
 - Frequency breakdown of words only: Frequenzanalyse von Wortformen
 - Frequency breakdown of annotation only: Frequenzanalyse von Annotationen
 - Frequency breakdown of words and annotation: Frequenzanalyse von Wortformen und Annotationen
 - 4. Die Auswahl mit Go! bestätigen.



Die Suchergebnisse werden nach Frequenz sortiert und in tabellarischer Form angezeigt.





Nachdem die Auswahl bestätigt wird, wechselt die **Breakdown position** für die nächste Anfrage automatisch von links nach rechts oder umgekehrt. Sie behält den Abstand vom Suchbegriff bei.

Verteilung der Suchergebnisse untersuchen

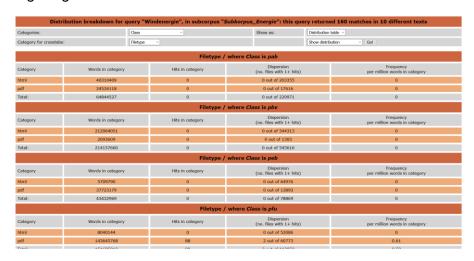
Bei der Analysefunktion *Distribution* kann die Verteilung von Suchergebnissen auf die im Korpus verwendeten Metadatenklassen angezeigt werden.

Führen Sie die nachfolgenden Schritte aus, um die Suchergebnisse gemäss den Metadatenklassen des Korpus anzeigen zu lassen:



ACHTUNG: Wenn ein mittelgrosses oder grosses Korpus verwendet wird, kann das Laden der *Distribution* sehr lange dauern oder der Browser kann abstürzen.

- 1. **Distribution** im Auswahlmenü oben rechts auswählen und daneben mit **Go!** bestätigen.
- → Das Optionenmenü der Analysemöglichkeit *Distribution* wird angezeigt.
 - 2. Die Categories bestimmen:
 - Class
 - Day
 - File type
 - Month
 - Source
 - Subclass
 - Year
 - General Information
 - File-frequency information
 - 3. Die Anzeige der Suchergebnisse bestimmen:
 - **Distribution table**: Tabellenansicht
 - Bar chart: Säulendiagramm
 - 4. Die Category for Crosstabs bestimmen.
- > Bei einer Crosstab wird jeweils die Häufigkeit der bei 4. gewählten Category for Crosstabs bei der Category, die bei 2. gewählt wurde, angezeigt.



5. Auswahl mit Go! bestätigen.



Die Suchergebnisse werden entsprechend der Verteilung sortiert und angezeigt.

Suchergebnisse sortieren

Bei der Analysemöglichkeit **Sort** können die Suchergebnisse alphabetisch sortiert werden.

Führen Sie die nachfolgenden Schritte aus, um die Suchergebnis alphabetisch zu sortieren:

- Sort im Auswahlmenü oben rechts auswählen und mit Go! bestätigen.
- Die *Position* (Position vom Knoten aus) bestimmen, bei der die Sortierung stattfindet: von 5 links bis 5 rechts möglich.
- 3. Die *Tag restriction* (Tag-Einschränkung) bestimmen.
- > Es werden nur Wörter und Zeichen mit dem bestimmten Tag berücksichtigt.
- > Sie können das Feld unter der Auswahl der **Tag restriction** ankreuzen, um stattdessen nur die bestimmte **Tag restriction** auszuschliessen.
 - 4. Bei Starting with Zeichenfolge eintippen
- > Es werden nur Wörter und Zeichen, die mit der eingetippten Zeichenfolge beginnen berücksichtigt.
- > Sie können das Feld unter der Auswahl von **Starting with** ankreuzen, um stattdessen nur Wörter und Zeichen, die mit der eingetippten Zeichenfolge beginnen, auszuschliessen.
 - 5. Die Auswahl mit *Update sort* (Sortierung aktualisieren) bestätigen.



Die sortierten Suchergebnisse werden angezeigt.



Die sortierten Suchergebnisse können mit den Analysemöglichkeiten im Auswahlmenü oben rechts untersucht werden.



Die sortierten Suchergebnisse können heruntergeladen und auf CQPweb gespeichert (siehe Kap. 4.6 Suchanfrage speichern, S. 23) oder kategorisiert werden.

Kollokationen untersuchen

Bei der Analysemöglichkeit *Collocations* können die Kollokationen der Suchergebnisse berechnet werden.

Führen Sie die nachfolgenden Schritte aus, um sich die Kollokationen der Suchergebnisse berechnen zu lassen.

- Collocations... im Auswahlmenü oben rechts auswählen und daneben mit Go! bestätigen.
- → Ein Optionenmenü der Analysemöglichkeit *Collocations* wird angezeigt.
 - Durch Auswählen von *Include* oder *Exclude* bestimmen, welche Annotationsebenen als Basis für die Kollokationsberechnung verwendet werden können:
 - Named Entity Types*^G
 - Lemma*G
 - Part of Speech*^G
 - Die *Maximum window span* des Kollokationsfensters bestimmen.
 - 4. Auswahl mit Create collocation database bestätigen.





Diese Standardeinstellungen können unter bestimmten Einstellungen geändert werden (siehe Kap. 4.3 Benutzeroberfläche anpassen, S. 10).

- 5. Basis der Kollokationsberechnung bestimmen:
 - Word form: Kollokationsberechnung basierend auf Wortformen.
 - Named Entity Type*^G: Kollokationsberechnung basierend auf Eigennamen-Tags (nur möglich, wenn bei 2. bei Named Entity Type das Feld Include angekreuzt wurde).
 - Lemma*^G: Kollokationsberechnung basierend auf Lemmata (nur möglich, wenn bei 2. bei Lemma das Feld Include angekreuzt wurde).
 - Part of Speech*G: Kollokationsberechnung basierend auf Wortarten (nur möglich, wenn bei 2. bei Part of Speech das Feld Include angekreuzt wurde).
- 6. Statistik bestimmen, die der Kollokationsberechnung zu Grunde liegen soll.
- 7. Den Beginn des Kollokationsfensters bestimmen.
- 8. Das Ende des Kollokationsfensters bestimmen.
- 9. Die Mindestfrequenz von Knoten und Kollokat innerhalb des Kollokationsfensters bestimmen.
- Die Mindestfrequenz des Kollokats innerhalb des Korpus bestimmen.

- 11. Optional: Suchergebnisse mit bestimmtem Kollokat filtern.
- 12. Optional: Suchergebnisse mit bestimmter *Part of Speech* des Kollokats filtern.
- Dies ist nicht möglich, falls bei 5. Part of Speech bestimmt wurde.
 - 13. Auswahl mit *Go!* bestätigen.

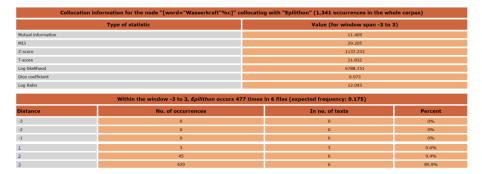




Die Suchergebnisse werden nach Frequenz sortiert in tabellarischer Form angezeigt.



Durch Auswählen eines Kollokats werden die Werte für alle verfügbaren Statistiken und die Verteilung der Positionen um den Knoten angezeigt.

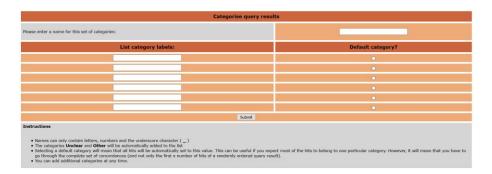


Kategorisieren

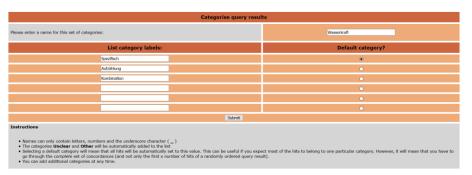
Bei der Analysemöglichkeit *Categorise* können die Suchergebnisse in verschiedene, selbstdefinierte Kategorien eingeteilt werden.

Führen Sie die nachfolgenden Schritte aus, um Suchergebnisse zu kategorisieren:

- Categorise... im Auswahlmenü oben rechts auswählen und daneben mit Go! bestätigen.
- → Ein Optionenmenü der Analysemöglichkeit *Categorise* wird angezeigt.



- 2. Die gesamte Kategorisierung benennen.
- > Die Namen dürfen nur aus Buchstaben, Ziffern und Unterstrichen bestehen. Sind andere Sonderzeichen enthalten, können die Kategorien nicht erstellt werden.
 - Die einzelnen Kategorien benennen. Das Ankreuzen von *Default* category bestimmt die ursprüngliche Kategorie, der alle Suchergebnisse zugeordnet werden.



- 4. Mit Submit bestätigen.
- → Die vorherigen Suchergebnisse werden angezeigt.
 - Rechts neben den Suchergebnissen die einzelnen Suchergebnisse einer in der bei 3. benannten Kategorie, other oder unclear, zuordnen.



- 6. Falls weitere Seiten kategorisiert werden, Save values for this page and go to next auswählen, ansonsten Save values and leave categorisation mode auswählen.
- 7. Mit Go! bestätigen.



Die Kategorisierung erscheint unter Categorised queries.



Kategorien trennen

Bei einer Kategorisierung werden die Suchergebnisse verschiedenen Kategorien zugeordnet. Diese Suchergebnisse können anschliessend in separate *Saved queries* anhand ihrer Kategorien gespeichert werden.

Führen Sie die nachfolgenden Schritte aus, um kategorisierte Suchergebnisse in ihre Kategorien zu trennen:

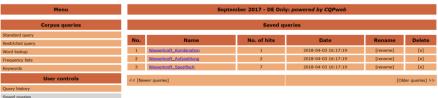


Voraussetzung: Eine Kategorisierung ist vorhanden.

- Seperate categories im Auswahlmenü rechts als Action auswählen.
- 2. Mit Go! bestätigen.



Die Suchergebnisse erscheinen in ihre Kategorien getrennt in **Saved** *queries* (siehe Kap. 4.6 Suchanfrage speichern, S. 23).



Neue Kategorien hinzufügen

Bei einer Kategorisierung werden Suchergebnisse verschiedenen Kategorien zugeordnet. Später können weitere Kategorien hinzugefügt und den Suchergebnissen zugeordnet werden.

Führen Sie die nachfolgenden Schritte aus, um einer kategorisierten Suche weitere Kategorien hinzuzufügen:



Voraussetzung: Eine Kategorisierung ist vorhanden.

- 1. Add categories im Auswahlmenü rechts als Action auswählen.
- 2. Mit Go! bestätigen.
- → Die bereits vorhandenen Kategorien werden angezeigt.
- → Ein leeres Feld, in dem eine weitere Kategorie benannt werden kann, wird angezeigt.



- 3. Die neue Kategorie benennen.
- 4. Mit Submit bestätigen.

- → Die vorherigen Suchergebnisse werden angezeigt.
 - 5. Rechts neben den Suchergebnissen die gewünschten Suchergebnisse der neuen Kategorie zuordnen.
 - 6. Falls Sie weitere Seiten kategorisieren Save values for this page and go to next auswählen, ansonsten Save values and leave categorisation mode auswählen.
 - 7. Mit Go! bestätigen.



Die Kategorisierung erscheint nun unter Categorised queries.

4.12 Word lookup: Wortformen suchen

In diesem Kapitel wird beschrieben, wie mit der Funktion word lookup ein Korpus nach einzelnen Wortformen und Tags durchsucht werden kann. Diese Funktion ist besonders für Einsteiger interessant, da man rasch einen Überblick über die Häufigkeit von Wortformen in einem Korpus erhält.





Diese Funktion kann nur auf das gesamte Korpus angewendet werden, nicht auf benutzerdefinierte Subkorpora.



Die Suche funktioniert nur mit der Simple query syntax*^G, nicht mit der CQP syntax*^G.

Korpus nach Wortformen durchsuchen

Führen Sie die nachfolgenden Schritte aus, um im Korpus nach Wortformen zu suchen:

- 1. In der Menüleiste auf Word lookup klicken.
- 2. Suchbegriff im Suchfeld eingeben.
- 3. Suchanfrage eingrenzen
- 4. Basis der Kollokation bestimmen:
 - starting with: Sucht Wörter, die mit dem Suchbegriff beginnen.
 - ending with: Sucht Wörter, die mit dem Suchbegriff enden.
 - containing: Sucht Wörter, die den Suchbegriff enthalten.
 - *matching exactly:* Sucht nur Wörter, die genau dem Suchbegriff entsprechen.
- 5. Darstellung der Suchergebnisse wählen (mit oder ohne Tags).
- 6. Gewünschte Anzahl Suchergebnisse pro Seite einstellen.
- 7. Auf *Lookup* klicken.



Die Suchergebnisse werden geladen und angezeigt.



Um die gefundenen Textstellen anzuzeigen, auf die jeweilige Wortform klicken.

Darstellung der Suchergebnisse anpassen

Im Auswahlmenü oben rechts können Sie die Darstellung der Suchergebnisse anpassen. Es gibt vier Möglichkeiten:

Frequency brekdown of words only: Zeigt die Treffer ohne Tags. Dies ist die standardmässige Ansicht, wenn eine Suche ohne Tags ausgeführt wurde.

Frequency breakdown of annotations only: Zeigt nur die Tags.

Frequency breakdown of words and annotation: Zeigt die Treffer und die Tags. Dies ist die standardmässige Ansicht, wenn eine Suche mit Tags ausgeführt wurde.

Show hits sorted by node: Ordnet die Treffer nach Text, in dem sie vorkommen.

Führen Sie die nachfolgenden Schritte aus, um die Darstellung der Suchergebnisse anzupassen:

- 1. Die gewünschte Darstellung wählen.
- 2. Mit Go! bestätigen.

Zusätzliche Funktionen des Auswahlmenüs

Sie können folgende weitere Funktionen über das Auswahlmenü ausführen:

Download frequency breakdown table: Lädt die Suchergebnisse herunter (siehe Kap. 4.15 Ergebnisse herunterladen, S. 55).

New query: Startet eine neue Suchanfrage.

Szenario: Die Erneuerbaren

Ihnen ist aufgefallen, dass in Texten zu erneuerbarer Energie öfters kurz von den "Erneuerbaren" die Rede ist. In Ihrer Bachelorarbeit wollen Sie diesem Phänomen auf den Grund gehen. Um einen ersten Anhaltspunkt über den Gebrauch des Nomens "Erneuerbare(n)" zu erhalten, können sie mit der Funktion *Word lookup* Ihr Korpus gezielt auf das Wort als Nomen (und nicht als Adjektiv) durchsuchen. Dazu führen Sie die nachfolgenden Schritte aus:

- 1. In der Menüleiste auf Word lookup klicken
- 2. Erneuerbare im Suchfeld eingeben.
- 3. Suchanfrage auf *starting with* eingrenzen, damit auch die Wortform Erneuerbaren gesucht wird.
- 4. Darstellung List by word-form and tag wählen.
- 5. 50 Suchergebnisse pro Seite einstellen.
- 6. Auf *Lookup* klicken.



Sie erhalten eine Liste mit den gefundenen Wortformen inklusive Tags. Diese könnte ungefähr so aussehen:

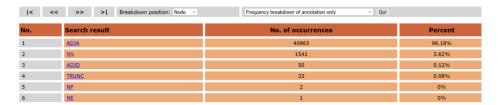
No.	Search result				
1	erneuerbaren ADJA				
2	erneuerbare ADJA				
3	erneuerbarer ADJA				
4	Erneuerbaren NN				
5	Erneuerbare NN				
6	erneuerbarem ADJA				
7	erneuerbares ADJA				

Führen Sie die nachfolgenden Schritte aus, um eine Übersicht über die Häufigkeit der Wortarten zu erhalten:

- Im Auswahlmenü oben rechts Frequency breakdown of annotation only auswählen.
- 2. Mit Go! Bestätigen.



Eine Übersicht über die Häufigkeit der Wortarten wird angezeigt.



Sie sehen unter anderem, wie oft der Suchbegriff als attributives Adjektiv (ADJA) und als normales Nomen (NN) verwendet wird. Da Sie in erster Linie an der Verwendung von Erneuerbare (n) als Nomen interessiert sind, können sie auf NN klicken. Dann werden die entsprechenden Textstellen angezeigt und Sie erhalten Informationen über die Anzahl Treffer und ihre Frequenz.

4.13 Frequency lists: Frequenzlisten erstellen

In diesem Kapitel wird beschrieben, wie mit der Funktion *Frequency lists* Frequenzlisten*^G erstellt werden. Mit dieser Funktion können Sie Frequenzlisten*^G für das ganze Korpus oder von benutzerdefinierten Subkorpora erstellen.





Um die Frequenzliste eines Subkorpus zu berechnen, muss diese zuerst kompiliert werden (siehe Kap. 4.8 Subkorpora erstellen, S. 29).

Frequenzliste anzeigen

Führen Sie die nachfolgenden Schritte aus, um die Frequenzliste für ein Korpus anzuzeigen:

- 1. In der Menüleiste auf Frequency lists klicken.
- 2. Das gewünschte Korpus auswählen.
- 3. Basis der Frequenzliste auswählen (s. unten).
- 4. Optional: Muster wählen.
- > Die Frequenzliste wird nach bestimmten Wörtern, Wortteilen, etc. gefiltert.
 - 5. Optional: Nach Frequenz filtern.
- → Nur Wörter innerhalb des gewünschten Frequenzbereichs werden angezeigt.
 - 6. Gewünschte Anzahl Suchergebnisse pro Seite einstellen.
 - 7. Gewünschte Sortierung der Suchergebnisse einstellen.
 - 8. Auf **Show frequency list** klicken.



Die Frequenzliste wird geladen und angezeigt.



Im Auswahlmenü oben rechts kann die Frequenzliste weiter analysiert werden (siehe Kap. 4.11 Korpus untersuchen, S. 40).

Basis der Frequenzliste wählen

Die Frequenzliste kann auf vier verschiedenen Kategorien basieren:

Word forms: Listet die Häufigkeit aller Wortformen auf. Unterschiedliche Formen des gleichen Lemmas werden separat erkannt (z. B. 3x ging, 4x gehst).

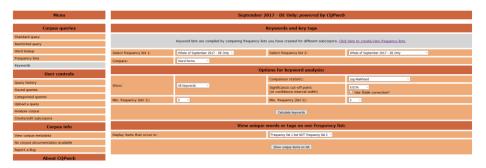
Named Entity Type*G: Listet automatisch auf, welche Ortsnamen (LOC), Personennamen (PER) und Organisationsnamen (ORG) im Text vorkommen.

Lemma*G: Listet die Häufigkeit aller Grundformen (wie im Wörterbuch) auf (z. B. 7x gehen).

Part of Speech*G: Listet die Häufigkeit grammatikalischer Kategorien auf (Anzahl Verben, Anzahl Adjektive, ...).

4.14 Keywords: Typisches Vokabular berechnen

In diesem Kapitel wird beschrieben, wie mit der Funktion *Keywords* typisches Vokabular eines Korpus berechnet wird. Dabei wird die Frequenzliste*^G eines Hauptkorpus mit der Frequenzliste eines Vergleichskorpus verglichen.





Voraussetzung: Ein Subkorpus wurde erstellt und seine Frequenzliste kompiliert (siehe Kap. 4.8 Subkorpora erstellen, S. 29).



Haupt- und Vergleichskorpus in derselben Sprache erstellen, um aussagekräftige Keywordlisten zu erhalten.

Keywords berechnen

Führen Sie die nachfolgenden Schritte aus, um die Keywords eines Korpus zu berechnen:

- 1. In der Menüleiste auf Keywords klicken.
- 2. Unter **Select frequency list 1** das Hauptkorpus wählen.
- 3. Unter **Select frequency list 2** das Vergleichskorpus wählen.
- > Standardmässig ist zweimal dieselbe Frequenzliste ausgewählt. Die Berechnung macht aber nur mit zwei unterschiedlichen Frequenzlisten Sinn.
 - 4. Vergleichskategorie wählen. In der Regel ist die Einstellung **Lemma** sinnvoll (siehe Kap. 4.13 Frequency lists: Frequenzlisten erstellen, S. 51).

- 5. Suche unter *Options for keyword analysis* anpassen. In der Regel unverändert lassen (siehe unten).
- 6. Auf calculate keywords klicken.



Die Suchergebnisse werden geladen und angezeigt.



Die Keywords des Hauptkorpus sind orange hinterlegt und mit + versehen. Die Keywords des Referenzkorpus sind grau hinterlegt und mit – versehen.

		In subcorpus "20)min_Mai_2015":	In whole "September 2017 - DE Only":			
No.	Word	Frequency (absolute)	Frequency (per mill)	Frequency (absolute)	Frequency (per mill)	+/-	Log likelihood
1		<u>z</u>	69.33	2,979,042	5,635.41	-	1065.49
2		338	3,347.50	<u>155,067</u>	293.34	+	1029.39
3)	309	3,060.28	4,373,184	8,272.69	-	440.73
4	is	<u>124</u>	1,228.08	<u>44,631</u>	84.43	+	432.86
5	:	<u>409</u>	4,050.67	<u>5,135,216</u>	9,714.21	-	431.4
6	/	<u>78</u>	772.50	2,207,628	4,176.14	-	425.21



Die Suchergebnisse werden nach **Log-Likelihood** geordnet, also nach Besonderheit/Signifikanz der Keywords. Die Frequenzspalten zeigen die absolute und relative Frequenz (pro Million Wörter) der gefundenen Keywords in den beiden Korpora.

Options for keyword analysis

In der Regel macht es Sinn, die *Options for keyword analysis* unverändert zu lassen. Hier können Sie beispielsweise folgendes einstellen:

Show All keywords: Standard. Zeigt positive und negative Keywords.

Show Positive keywords: Zeigt nur die Keywords des Hauptkorpus.

Show Negative keywords: Zeigt nur die Keywords des Referenzkorpus.

Show Lockwords: Zeigt Wörter, die in beiden Korpora eine ähnliche Frequenz haben. Funktioniert nur, wenn Sie unter **Comparison statistic** entweder **Log Ratio with Confidence Interval filter** oder **Log Ratio (unfiltered)** wählen.

Min. frequency (list 1): Frequenz, mit der ein Keyword im Hauptkorpus mindestens vorkommen soll.

Min. frequency (list 2): Frequenz, mit der ein Keyword im Referenzkorpus mindestens vorkommen soll.



Sie können die Darstellung der Suchergebnisse auch nachträglich nach positiven und negativen Keywords filtern.

Führen Sie dazu folgende Schritte aus:

- Im Auswahlmenü oben in der Mitte die gewünschte Darstellung wählen.
- 2. Mit Go! bestätigen.



Die Suchergebnisse werden nach positiven bzw. negative Keywords gefiltert.

Keywords berechnen, die nur in einem Korpus vorkommen Unter *View unique words or tags on one frequency list* können Keywords berechnet werden, die ausschliesslich im Hauptkorpus oder ausschliesslich im Vergleichskorpus vorkommen:

- 1. In der Menüleiste auf Keywords klicken.
- 2. Frequenzliste 1 oder 2 wählen. Es werden jene Keywords berechnet, die nur in der gewählten Frequenzliste vorkommen.
- 3. Auf **Show unique items on list** klicken.



Die Suchergebnisse werden geladen und angezeigt.

Szenario: energiestadt.ch

Für Ihre Bachelorarbeit über erneuerbare Energien untersuchen Sie, wie sich das Vokabular von Organisationen, die sich mit dem Energiethema befassen, vom normalen Vokabular in Zeitungen abhebt. Exemplarisch vergleichen Sie die Webseite energiestadt.ch mit derjenigen von 20 Minuten. Dazu führen Sie die nachfolgenden Schritte aus:

- 1. Subkorpus mit Texten mit der *source energiestadt* erstellen (siehe Kap. 4.8 Subkorpora erstellen, S. 29).
- 2. Subkorpus mit Texten mit der source 20min erstellen.
- 3. Für beide Subkorpora die Frequenzliste kompilieren (siehe Kap. 4.8 Subkorpora erstellen, S. 29).
- 4. In der Menüleiste auf *Keywords* klicken.
- Unter Select frequency list 1 das Subkorpus energiestadt wählen.
- 6. Unter **Select frequency list 2** das Subkorpus 20min wählen.
- 7. Auf calculate keywords klicken.



Sie erhalten eine Liste mit positiven und negativen Keywords. Diese sieht ungefähr so aus:

No.	Word	In subcorpus `	`energiestadt":	In subcorpu	us "20min":		Log likelihood	
		Frequency (absolute)	Frequency (per mill)	Frequency (absolute)	Frequency (per mill)	+/-		
1		<u>57,935</u>	36,649.23	10,371	187.82	+	359471.59	
2	Energiestadt	6,751	4,270.63	<u>10</u>	0.18	+	48236.62	
3	GEMEINDE	5,947	3,762.03	2,185	39.57	+	33278.93	
4	GEMEINDEN	4,842	3,063.01	1,744	31.58	+	27182.76	
5	energie	5,187	3,281.26	3,147	56.99	+	26299.9	
6	%	<u>4,681</u>	2,961.16	<u>5,860</u>	106.13	+	19393.19	
7	label	2,597	1,642.84	<u>416</u>	7.53	+	16211.16	
8	ENERGIESCHWEIZ	1.854	1.172.83	21	0.38	+	13053.32	



Im Auswahlmenü oben in der Mitte können Sie nun nach positiven und negativen Keywords filtern. Positive Keywords sind typisch für die Webseite energiestadt.ch, negative Keywords sind typisch für 20min und entsprechend untypisch für energiestadt.ch.

4.15 Ergebnisse herunterladen

In diesem Kapitel wird beschrieben, wie die Such- und Analyseergebnisse heruntergeladen werden können und welche spezifischen Einstellungsmöglichkeiten es für den Download gibt.



Alle Ergebnisse werden als Textdatei (.txt) heruntergeladen, die z.B. in Word oder Excel importiert und angezeigt werden kann.



Alle Suchergebnisse in CQPweb sowie darauf basierende Analyseergebnisse können entweder mit normaler oder tabellarischer Darstellung heruntergeladen werden.



Das Herunterladen kann sehr lange dauern, wenn die Anzahl der Suchoder Analyseergebnisse gross ist.

Suchergebnisse herunterladen



Führen Sie die nachfolgenden Schritte aus, um das Downloadfenster für die Suchergebnisse aufzurufen:

- Suchanfrage erstellen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15).
- 2. Im Auswahlmenü oben rechts **Download...** auswählen und daneben mit **Go!** bestätigen.



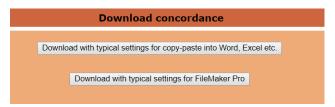
Das Fenster für den Download mit normaler Darstellung erscheint.

Normale Darstellung

Die normale Darstellung *Plain-text concordance**^G bildet die Suchergebnisse ungefähr so ab, wie sie von *CQPweb* auch im Browser angezeigt werden. Die Suchergebnisse können entweder mit standardmässigen oder mit benutzerdefinierten Einstellungen heruntergeladen werden.

Standardmässige Einstellungen

Es existieren zwei unterschiedliche Downloadmöglichkeiten mit standardmässigen Einstellungen. Die erste eignet sich für die anschliessende Nutzung in gängigen Tools wie Word oder Excel, die zweite hingegen ist speziell für den Import in das Datenbanksystem FileMaker Pro gedacht.



Download für Word, Excel etc.

Führen Sie den nachfolgenden Schritt aus, um die Suchergebnisse für Word, Excel usw. herunterzuladen:

1. Unter **Download concordance*** auf **Download with typical settings for copy-paste into Word, Excel etc.** klicken.



Die Suchergebnisse werden mit den standardmässigen Einstellungen für Word, Excel usw. heruntergeladen. Der Download wird im Browser angezeigt.

Download für FileMaker Pro

Führen Sie den nachfolgenden Schritt aus, um die Suchergebnisse für FileMaker Pro herunterzuladen:

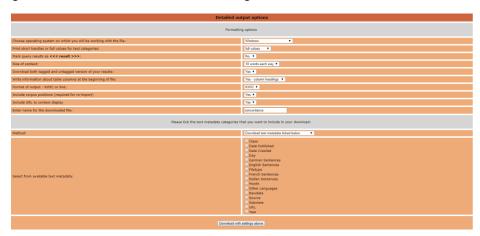
1. Unter **Download concordance*** auf **Download with typical** settings for FileMaker Pro klicken.



Die Suchergebnisse werden mit den standardmässigen Einstellungen für FileMaker Pro heruntergeladen. Der Download wird im Browser angezeigt.

Benutzerdefinierte Einstellungen

Es besteht auch die Möglichkeit, die Suchergebnisse mit benutzerdefinierten Einstellungen herunterzuladen. Dafür können diverse Einstellungen zu Format und Metadaten *G vorgenommen werden.



Führen Sie die nachfolgenden Schritte aus, um die Suchergebnisse mit benutzerdefinierten Einstellungen herunterzuladen:

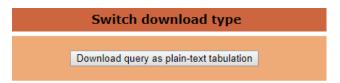
- 1. Unter *Choose operating system...* das Betriebssystem des Geräts auswählen, mit dem gearbeitet wird.
- Dieser Wert wird durch den Browser standardmässig bereits richtig gesetzt.
 - 2. Unter *Print short handles** or *full values** ... definieren, ob die Metadaten in der Textdatei abgekürzt oder ausgeschrieben werden sollen.
 - Unter *Mark query results...* definieren, ob die Suchergebnisse zusätzlich mit Klammern markiert werden sollen. Diese zusätzliche Markierung ist sinnvoll bei der *Line View*^G* (siehe Schritt 7).
 - 4. Unter **Size of context** auswählen, wie viel Kontext (Wörter vor und nach dem Suchergebnis) angezeigt werden sollen.
 - 5. Unter **Download both tagged and untagged version...** definieren, ob die Suchergebnisse mit und ohne oder nur ohne Annotation heruntergeladen werden sollen.
 - Unter Write information about table columns... auswählen, ob die Namen der Tabellenspalten zu Beginn der Datei aufgeführt werden sollen.
 - 7. Unter *Format of output* auswählen, ob die Suchergebnisse mit *KWIC View**^G oder *Line View**^G dargestellt werden sollen.
 - Unter *Include corpus positions...* definieren, ob die Position (Nummer) des Suchergebnisses im Korpus angezeigt werden soll.
 - Unter Include URL to context display definieren, ob die URL für das Abrufen des vollständigen Kontexts angezeigt werden soll
 - 10. Unter *Enter name for the downloaded file* den Namen für die zu herunterladende Textdatei angeben.
 - 11. Unter *Method for metadata**^G *download* auswählen, ob und welche Metadaten heruntergeladen werden sollen.
 - Alle Metadaten herunterladen.
 - Nur die wichtigsten Metadaten (classification-type metadata*⁶) herunterladen.
 - Nur unten angewählte Metadaten herunterladen:
 Bei dieser Variante die gewünschten Metadaten anwählen
 (für weitere Informationen zu Metadaten siehe Kap. 5.2 Metadaten, S. 64).
 - 12. Auf **Download with settings above** klicken.



Die Suchergebnisse werden mit den benutzerdefinierten Einstellungen heruntergeladen. Der Download wird im Browser angezeigt.

Tabellarische Darstellung

Die tabellarische Darstellung *Plain-text tabulation* bildet die Ergebnisse in Form einer Tabelle ab, wobei jedes Suchergebnis in einer separaten Zeile steht.



Führen Sie die nachfolgenden Schritte aus, um das Downloadfenster für die Suchergebnisse mit tabellarischer Darstellung aufzurufen:

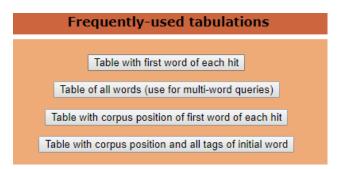
- Suchanfrage erstellen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15).
- 2. Im Auswahlmenü oben rechts auf **Download...** klicken und daneben mit **Go!** bestätigen.
- 3. Unter **Switch download type** auf **Download query as plaintext tabulation** klicken.



Das Fenster für den Download mit tabellarischer Darstellung erscheint.

Standardtabelle

Es stehen vier unterschiedliche Standardtabellen zum Download bereit. Diese Tabellen sind relativ einfach aufgebaut und werden dementsprechend oft von den Nutzern verwendet.



Führen Sie den nachfolgenden Schritt aus, um die Suchergebnisse als Standardtabelle herunterzuladen:

 Unter Frequently-used tabulations auf die gewünschte Standardtabelle klicken.



Die Suchergebnisse werden als Standardtabelle heruntergeladen. Der Download wird im Browser angezeigt.

Table with first word of each hit: Diese Tabelle enthält nur eine Spalte, in der jeweils das erste Wort von jedem Suchergebnis steht. Sofern die gestellte Suchanfrage nur ein Wort betrifft, wird das Ergebnis somit vollständig angezeigt.

Table of all words (use for multi-word queries): Diese Tabelle enthält ebenfalls nur eine Spalte, in der jeweils alle Wörter von jedem Suchergebnis stehen. Sofern die gestellte Suchanfrage nur ein Wort betrifft, ist die Tabelle identisch mit der ersten.

Table with corpus position of first word of each hit: Diese Tabelle enthält ebenfalls nur eine Spalte, in der jeweils die Position (Nummer) des ersten Wortes von jedem Suchergebnis im Korpus steht.

Table with corpus position and all tags of initial word: Diese Tabelle enthält mehrere Spalten: In der ersten steht jeweils die Position (Nummer) des ersten Wortes von jedem Suchergebnis im Korpus, in der zweiten das Wort und in der dritten das Wort mit Annotation.

Benutzerdefinierte Tabelle

Es besteht auch die Möglichkeit, eine benutzerdefinierte Tabelle zu erstellen. Dafür können Einstellungen zu Anzahl, Inhalt und Form der Spalten vorgenommen werden.



Das Erstellen einer benutzerdefinierten Tabelle ist relativ schwierig und eignet sich nur für fortgeschrittene Nutzer. Deswegen werden hier nur die wichtigsten Punkte erklärt.

				Specify	custom tabulation			
	Beg	in at	End at		Normalise?			
Col. no.	Anchor	Offset	Anchor	Offset	Attribute	Case	Diacritics	
1	match *	no offset ▼	match *	no offset ▼	(Column not in use) Y			
2	match *	no offset ▼	match •	no offset ▼	(Column not in use) •	=	=	
3	match *	no offset ₹	match *	no offset ▼	(Column not in use) *			
4	match *	no offset ▼	match *	no offset ▼	(Column not in use) ▼		=	
5	match *	no offset ▼	match *	no offset ▼	(Column not in use)			
hoose operating system on which y	ou will be working with th	e file:			₩ndres			
Enter name for the downloaded file: Labolat					tabulation			
Counted quay tabulation with sattings above								
				I need n	nore output columns!			
Do you need more tabulation columns? Use this control: I want a tabulation with (up to) 9 v columns: Create bigger form!								

Führen Sie die nachfolgenden Schritte aus, um die Suchergebnisse als benutzerdefinierte Tabelle herunterzuladen:

- Unter *I need more output columns!* die Anzahl der gewünschten Spalten auswählen und auf *Create bigger form!* klicken, um die Anzahl der Spalten zu verändern.
- 2. Unter **Begin at, Anchor** auswählen, bei welcher Art von Information das Suchergebnis beginnen soll.
 - *Match:* erstes Wort des Suchergebnisses
 - Matchend: letztes Wort des Suchergebnisses
- 3. Unter **Begin at, Offset** auswählen, ob und wieviel Kontext angezeigt werden soll.
- 4. Unter *End at, Anchor* auswählen, bei welcher Art von Information das Suchergebnis aufhören soll.

- 5. Unter *End at, Offset* auswählen, ob und wieviel Kontext angezeigt werden soll.
- 6. Unter *Attribute* auswählen, welche Information in der jeweiligen Spalte angezeigt werden soll:
 - Corpus position number
 - Word
 - Entitytype*^G
 - Lemma*G
 - Part-of-speech*^G
- 7. Unter **Normalise?** anwählen, ob die Suchergebnisse klein **(Case)** und/oder ohne diakritische Zeichen*^G **(Diacritics***^G) angezeigt werden sollen.
- 8. Schritte 2 bis 7 für jede Spalte wiederholen.
- Unter Choose operating system... Betriebssystem des Geräts auswählen, mit dem gearbeitet wird.
 Dieser Wert wird durch den Browser standardmässig bereits richtig gesetzt.
- 10. Unter *Enter name for the downloaded file...* den Namen für die zu herunterladende Textdatei angeben.
- 11. Auf *Download query tabulation with settings above* klicken.



Die Suchergebnisse werden als benutzerdefinierte Tabelle heruntergeladen. Der Download wird im Browser angezeigt.

Analyseergebnisse herunterladen

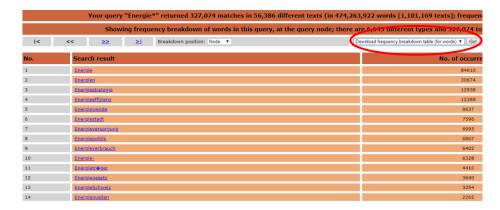
Nebst der Möglichkeit, Ergebnisse basierend auf einer Suchanfrage zu analysieren, gibt es noch weitere Analysemöglichkeiten, die auf der Startseite unter *Corpus queries* zu finden sind.

Corpus queries					
Standard query					
Restricted query					
Word lookup					
Frequency lists					
Keywords					



Alle untenstehenden Analyseergebnisse werden automatisch mit tabellarischer Darstellung heruntergeladen.

Word lookup: Wortformen suchen



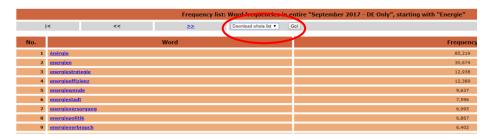
Führen Sie die nachfolgenden Schritte aus, um die Ergebnisse der Funktion *Word lookup* herunterzuladen:

- 1. Unter *Word lookup* eine Suchanfrage erstellen (siehe Kap. 4.12 Word lookup: Wortformen suchen, S. 48).
- 2. Im Auswahlmenü oben rechts auf *Download frequency break-down table (for words)* und daneben mit *Go!* bestätigen.



Die Ergebnisse der Funktion *Word lookup* werden heruntergeladen. Der Download wird im Browser angezeigt.

Frequency lists: Frequenzlisten erstellen



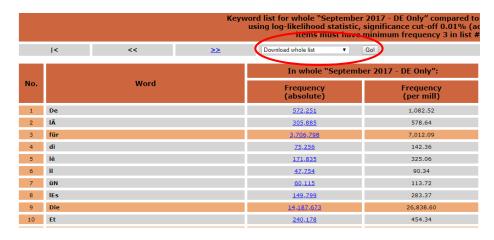
Führen Sie die nachfolgenden Schritte aus, um die Ergebnisse der Funktion *Frequency lists**⁶ herunterzuladen:

- 1. Unter *Frequency lists**^G eine Suchanfrage erstellen (siehe Kap. 4.13 Frequency lists: Frequenzlisten , S. 51).
- 2. Im Auswahlmenü oben in der Mitte auf **Download whole list** und daneben mit **Go!** bestätigen.



Die Ergebnisse der Funktion *Frequency lists**^G werden heruntergeladen. Der Download wird im Browser angezeigt.

Keywords: Typisches Vokabular berechnen



Führen Sie die nachfolgenden Schritte aus, um die Ergebnisse der Funktion *Keywords**^G herunterzuladen:

- 1. Unter Keywords eine Suchanfrage erstellen (siehe Kap. 4.14 Keywords: Typisches Vokabular berechnen, S. 52).
- 2. Im Auswahlmenü oben in der Mitte auf **Download whole list** und daneben mit **Go!** bestätigen.



Die Ergebnisse der Funktion *Keywords**⁶ werden heruntergeladen. Der Download wird im Browser angezeigt.

5 Glossar

Das vorliegende Kapitel umfasst drei verschiedene Glossare, die das Verständnis von CQPweb und der Benutzerdokumentation unterstützen sollen.

5.1 **Fachbegriffe**

In diesem Kapitel werden die Bedeutungen aller im Text mit *G markier-

ten Fachbegriffe erklärt.

Concordance (dt. Konkordanz) Elektronisch erstellte Trefferliste, welche das Ergebnis

einer Suche nach Wörtern/Phrasen ist.

CQP Syntax (dt. CQP Syntax) Standardisierte Art, einen Suchbegriff einzugeben.

Funktioniert mit regulären Ausdrücken*G.

Diacritics (dt. diakritische Zeichen) Zusätze an oder in Schriftzeichen, mit denen

bestimmte Unterscheidungen getroffen werden sollen (Akzent- und Be-

tonungszeichen).

Eigenname in einem Korpus. **Entitytype**

(dt. Frequenzliste) Liste von Suchbegriffen mit Angabe der Häufigkeit ih-Frequency list

res Vorkommens im Korpus.

Full value Vollständiger Name der Metadaten.

Keyword Wort, das in einem Korpus häufiger als erwartet vorkommt. Das heisst,

es tritt im Korpus häufiger auf als in den anderen Korpora.

KWIC View -Keyword in Context

(dt. Schlüsselwort im Kontext) Kontexte links und rechts vom Sucher-

gebnis werden in separaten Spalten dargestellt.

Lemma (dt. Lemma) Grundform eines Wortes, also die Wortform, welche in ei-

nem Nachschlagwerk vorkommt.

Line View Suchergebnis und Kontext werden in einer gemeinsamen Spalte darge-

stellt.

Metadata (dt. Metadaten) Daten, welche Informationen über die Texte im Korpus

geben.

Named Entity **Type**

Siehe Entitytype.

POS - part of Die Wortart, bzw. grammatische Kategorie eines Wortes. Werden mit speech

Tags beschrieben (siehe Kap. 5.3 Tagset, S. 65).

Short handle Abkürzung der Metadaten. Simple query syntax

(dt. Simple Query Syntax) Relativ einfache, aber nicht standardisierte

Art, einen Suchbegriff einzugeben.

Subcorpus (dt. Subkorpus) Korpus, das einen grösseren Korpus weiter eingrenzt.

Regulärer Ausdruck

Bekannte und weit verbreitete Symbole für das Erweitern von Suchbe-

griffen.

5.2 Metadaten

In diesem Kapitel werden die Bedeutungen aller Metadaten erklärt, mit denen Texte im Korpus Swiss-AL ausgezeichnet werden. Je nach Funktion sind unterschiedliche Metadaten zur Auswahl verfügbar resp. einsehbar.

class Akteursklasse, zu der der Text gehört.

Die Akteursklassen teilen sich auf in:

• Politikadressaten und -betroffene (PAB).

• Politikformulierung und -umsetzung (PFU).

Politikbeobachtung und -vermittlung (PBV).

• Politikentwicklung und -beratung (PEB).

day Tag, an dem der Text publiziert wurde (falls nicht vorhanden: 11).

filetype Datenyp der Quelle: html otder pdf.

main language (Überwiegende) Sprache des Textes.

month Monat, in dem der Text publiziert wurde (falls nicht vorhanden: 11).

source Quelle des Textes als Kurzname.

subclass Spezifikation der Akteursklassen.

year Jahr, in dem der Text publiziert wurde (falls nicht vorhanden: 9999).

5.3 Tagset

In diesem Kapitel werden alle möglichen Tags für deutsche Texte aufgelistet. Die Tags bezeichnen z.B. die Wortarten in *CQPweb*. Diese Tags

wurden im Stuttgart-Tübingen-Tagset (STTS) definiert.

ADJA attributives Adjektiv

ADJD adverbiales oder prädikatives Adjektiv

ADV Adverb

APPR Präposition; Zirkumposition links

APPRART Präposition mit Artikel

APPO Postposition

APZR Präposition; Zirkumposition rechts

ART bestimmter oder unbestimmter Artikel

CARD Kardinalzahl

FM Fremdsprachliches Material

ITJ Interjektion

KOUI unterordnende Konjunktion mit "zu" und Infinitiv

KOUS unterordnende Konjunktion mit Satz

KON nebenordnende Konjunktion

KOKOM Vergleichspartikel, ohne Satz

NN normales Nomen

NE Eigennamen

PDS substituierendes Demonstrativpronomen

PDAT attribuierendes Demonstrativpronomen

PIS substituierendes Indefinitpronomen

PIAT attribuierendes Indefinitpronomen ohne Determinativ

PIDAT attribuierendes Indefinitpronomen mit Determinativ

PPER irreflexives Personalpronomen

PPOSAT attribuierendes Possessivpronomen

PPOSS substituierendes Possessivpronomen

PRELS substituierendes Relativpronomen

PRELAT attribuierendes Relativpronomen

PRF reexives Personalpronomen

PWS substituierendes Interrogativpronomen

PWAT attribuierendes Interrogativpronomen

PWAV adverbiales Interrogativ oder Relativpronomen

PAV Pronominaladverb

PTKZU "zu" vor Infinitiv

PTKNEG Negationspartikel

PTKVZ abgetrennter Verbzusatz

PTKANT Antwortpartikel

PTKA Partikel bei Adjektiv oder Adverb

TRUNC Kompositions-Erstglied

VVFIN Finites Verb, voll

VVIMP Imperativ, voll

VVINF Infinitiv, voll

VVIZU Infinitiv mit "zu", voll

VVPP Partizip Perfekt, aux

VAFIN Finites Verb, aux

VAIMP Imperativ, aux

VAINF Infinitiv, aux

VAPP Partizip Perfekt, aux

VMFIN Finites Verb, modal

VMINF Infinitiv, modal

VMPP Partizip Perfekt, modal

XY Nichtwort, Sonderzeichen enthaltend

\$, Komma

\$. Satzbeendende Interpunktion

\$(sonstige Satzzeichen; satzintern

6 FAQ

Dieses Kapitel bietet Antworten auf oft gestellte Fragen. Die Fragen sind unterschiedlichen Kategorien zugeordnet.

Kontakt

Wem kann ich schreiben, wenn etwas nicht funktioniert?

Sie können an folgende E-Mail-Adresse schreiben: korpuslinguistik@zhaw.ch.

Korpus Swiss-AL

Welche Quellen sind im Korpus enthalten?

Die vorinstallierten Korpora enthalten Texte von Schweizer Webseiten in den Sprachen Deutsch, Französisch, Italienisch und Englisch. Die Korpora sind nach ihrem Release (Jahr und Monat) und nach Sprache der Quelltexte geordnet.

Kann ich den gesamten Text einsehen, in dem ein Beleg gefunden wird?

Ja, bei den meisten Suchergebnissen können Sie den gesamten Quelltext aufrufen. Führen Sie dafür die nachfolgenden Schritte aus:

- Bei jeder gestellten Suchanfrage erscheint in der Ergebnisliste auf der linken Seite zu jedem Ergebnis ein verlinkter Filename. Klicken Sie auf diesen Filename.
 - → Es erscheint ein neues Fenster mit den Metadaten zur Suchanfrage bzw. zum Quelltext.
- Wenn bei der URL-Angabe ein Link angegeben ist, klicken Sie darauf, um sich den Quelltext anzeigen zu lassen. Steht dort N/A kann die entsprechende URL nicht aufgerufen werden.

Suchanfragen

Wie finde ich Texte, in denen zwei Wörter vorkommen, allerdings nicht unbedingt direkt nacheinander?

Führen Sie die nachfolgenden Schritte aus, um Texte zu suchen, in denen zwei Wörter vorkommen:

- Normale oder eingeschränkte Suchanfrage mit dem ersten gesuchten Begriff erstellen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15) und Suchanfrage speichern (siehe Kap. 4.6 Suchanfrage speichern, S. 23).
- 2. Subkorpus mit *Texts found in a saved query* erstellen (siehe Kap.4.8 Subkorpora erstellen, S. 29).
- 3. Im erstellten Subkorpus normale Suchanfrage mit dem zweiten gesuchten Begriff erstellen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15).

Wie finde ich Abfolgen von Wörtern, die direkt nacheinander kommen oder mit einer bestimmten Anzahl Wörtern dazwischen?

Führen Sie den nachfolgenden Schritt aus, um direkt nebeneinander stehende Wörter zu finden:

- Normale oder eingeschränkte Suchanfrage erstellen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15) und die gesuchten Wörter direkt nebeneinander schreiben. z.B. [word="Energie"] [word="Bewegung"].
- → Es werden nur Suchergebnisse angezeigt, in denen die gesuchten Wörter direkt nebeneinander stehen.

Führen Sie den nachfolgenden Schritt aus, um Wörter zu finden, zwischen denen fakultativ noch andere Wörter stehen können:

- Normale oder eingeschränkte Suchanfrage erstellen (siehe Kap. 4.5 Suchanfrage erstellen, S. 15) und zwischen den gesuchten Wörtern noch den regulären Ausdruck für ein unbestimmtes Wort [] und den regulären Ausdruck für die Wiederholung von Elementen {x,y}. z.B. [word="Energie"] [] {0,3} [word="Bewegung"].
- → Es werden Suchergebnisse angezeigt, in denen 0 bis 3 beliebige Wörter zwischen den gesuchten Wörtern vorkommen.

Woher weiss ich, wie die Wortarten abgekürzt werden?

Eine Übersicht für die Abkürzungen der Wortarten ist im Glossar im Kap. 5.3 Tagset, S. 65 enthalten.

Meine Suchanfrage funktioniert nicht. Was mache ich falsch?

Überprüfen Sie die folgenden Punkte, welche die häufigsten Fehlerquellen sind:

- Anführungs- und Schlusszeichen richtig setzen.
- · Klammern richtig öffnen und schliessen.
- Bei der CQP Syntax die Wortform richtig deklarieren (z.B. word und nicht wort).
- Richtigen Query mode setzen (der Query mode wird standardmässig auf Simple query (ignore case) gesetzt).

Kann ich Suchanfragen speichern und später wieder nutzen?

Ja, die Suchanfragen können gespeichert werden. Sie können eine Suchanfrage entweder im System speichern (siehe Kap. 4.6 Suchanfrage speichern, S. 23) und diese dort zu einem späteren Zeitpunkt aufrufen, oder Sie können die Ergebnisse herunterladen (siehe Kap. 4.15 Ergebnisse herunterladen, S. 55).

Texte herunter-/hochladen

Können Texte in Form von Textdateien (.txt) von CQPweb heruntergeladen bzw. für eine Analyse auf CQPweb hochgeladen werden?

Nein, es ist nicht möglich, Texte herunter- oder hochzuladen.

7 Weiterführende Informationen

In diesem Kapitel sind Literatur und Links aufgeführt, die weitere, nützliche Informationen enthalten und als Nachschlagewerke dienen können.

7.1 Literatur

Hardie, A. (2012). CQPweb. Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, *17*(3), 380-409.

Lemnitzer, L., & Zinsmeister, H. (2015). *Korpuslinguistik. Eine Einführung.* (3., überarbeitete und erweiterte Auflage). Tübingen: Narr Studienbücher.

7.2 Links

Wiki der Korpusgruppe (nur im internen ZHAW-Netz oder über VPN erreichbar): http://160.85.180.39/dokuwiki/doku.php/start

Youtube-Channel mit Tutorials zu CQPweb:

https://www.youtube.com/playlist?list=PL2XtJIhhrHNQqf4Dp6sckGZRU4NiUVw1e

IMS Open Corpus Workbench – Informationen und offizielle Dokumentation: http://cwb.sourceforge.net/cqpweb.php

Informationen zum Korpus Swiss-AL:

https://www.zhaw.ch/de/linguistik/forschung/swiss-al/