Das Portal eAQUA

Zugangsbereiche

Das Portal eAQUA ist gegenwärtig in zwei voneinander getrennte Bereiche unterteilt. In einen frei zugänglichen Bereich, der zu Demonstrationszwecken eine eingeschränkte Nutzung der möglichen Funktionsvielfalt anbietet, und einen geschlossenen Bereich, welcher nur durch einen Login über Benutzerkennung möglich ist. Eine derartige Trennung ist in erster Linie lizenzrechtlichen Problemen geschuldet. Diverse Korpora sind an Benutzerlizenzen der herausgebenden Verlage gebunden. Wir sind deswegen angehalten, vor Herausgabe von Benutzerkennungen zu überprüfen, ob eine entsprechende Lizenz vorliegt.

Darüber hinaus existieren zwei Frontend-Versionen, die unterschiedliche (Programmier-)Techniken einsetzen. Um den Entwicklungen im Bereich der dynamischen Internetseitengenierung auch durch die sogenannte HTML5-Technik gerecht zu werden, ist ein zweite Variante erstellt worden, die bei der Visualisierung vollständig auf JavaScript basiert. Die erste benutzt zur visuellen Darstellung eine proprietäre Software der Firma Adobe, genannt Flash, für die ein zusätzliches Browser-Plugin notwendig ist.

Deshalb werden bei Erklärungen der Tools zwei Versionen gegeben, die vor allem den unterschiedlichen visuellen Darstellungen, nicht aber den Analyseergebnissen, geschuldet sind.

Demonstrationen

Demonstration Kookkurrenz-Analyse

Die Demonstration der Kookkurrenz-Analyse ist über den Menü-Punkt Tools erreichbar. Sie beinhaltet einige grundsätzliche Analyseergebnisse aus dem Projekt-Portal und wurde im Laufe der Dissemination um weiter Korpora bereichert. Für den Login in den geschützten Bereich, in dem Ergebnisse von Korpora abgerufen werden können, die an Benutzungslizenzen gebunden sind, ist am rechten Bildschirmrand ein Link **Login** vorgesehen.

Nach Auswahl des Korpus im oberen Bereich gibt es mehrere Möglichkeiten für eine Suche in der Datenbank. Im Feld **Wort-Suche** kann nach einem Wort oder Wort-Teil in der Datenbank geschaut werden. Das Feld ist mit einer **Autovervollständigung** versehen, die ab **3 Buchstaben** wirksam wird. Diese Listet alle gefunden Wörter (mit ihren Häufigkeiten in Klammern) auf. Bei Klick auf eines der Listenergebnisse wird dieses Wort übernommen. Es ist daran zu erkennen, dass rechts im Feld **Wort-ID** eine Zahl steht. Bei Klick auf den Button Start würde dieses Wort analysiert.

Falls keine Vorschläge unterbreitet werden, kann in der Datenbank nach Treffern gefahndet werden, indem nur die Anfangsbuchstaben in das Feld gegeben werden und dann auf Start geklickt wird. Es erscheint eine Liste aller Terme, die mit diesen Buchstaben beginnen. Die Begriffe sind mit einem Link versehen, der die Kookkurrenz-Analyse für denjenigen startet.

Falls durch eine vorhegende Suche die ID des Wortes bekannt ist, kann diese auch in das Feld **Wort-ID** eingetragen und die Analyse gestartet werden.

Es gibt in der Demo-Version, im Gegensatz zum Portal selbst, eine kleine Einschränkung hinsichtlich der maximal anzuzeigenden Treffer, die über das Dropdown **Trefferanzahl** definierbar ist. Dies ist wegen der dynamisch geladenen Visualisierung notwendig, damit die Seite nicht allzu lang zum Laden braucht. Um auch unsignifikante Kookkurrenzen finden zu können, ist ein zusätzlicher **Sortierung**sfilter eingebaut. So lassen sich zu einem Wort zum Beispiel die 100 unsignifikantesten Kookkurrenzen anzeigen, indem die Trefferanzahl auf 100 und die Sortierung auf selten gesetzt wird.

Das Ergebnis präsentiert sich ähnlich wie im Portal als Liste unterteilt in Wörter mit ähnlichem Zusammenhang, den hundert signifikantesten (bzw. hier auch unsignifikantesten) Satzkookkurrenzen (gesamt / links / rechts) sowie den (un-) signifikantesten rechten und linken Nachbarn. Außerdem werden die (un-) signifikantesten Kookkurrenzen in einer Visualisierung dargestellt. Zum Aufruf dieser werden in der linken Spalte unterhalb von „Signifikante Kookkurrenzen“ zwei Buttons eingeblendet.

Ganz am Ende finden sich Belegstellen, in denen das Wort Verwendung findet.

Die Listen der Kookkurrenzen sind jeweils mit Querverweisen hinterlegt. Bei Klick auf ein Wort, wird die Kookkurrenzsuche genau für diesen Term gestartet, bei Klick auf die Zahl in Klammern, die die Häufigkeit der Vorkommen angibt, werden Belegstellen mit beiden Ausdrücken angezeigt.

Wie aus der Kookkurrenz-Analyse Daten exportiert werden können, ist hier notiert.

Visualisierung der Kookkurrenzen

Das Fenster mit der in JavaScript geschriebenen Netzwerk-Visualisierung vis.js (http://visjs.org/; Apache 2.0 / MIT-Lizenz), wird über die zwei Buttons in der linken Spalte unterhalb von „Signifikante Kookkurrenzen“ aufgerufen, gruppiert einmal nach der **Frequenz** (Häufigkeit) der Kookkurrenzen, zum anderen nach dem Signifikanzmaß **Log-Likelihood**. Zu den Signifikanzmaßen und deren Bedeutung finden sich ein paar Erklärungen hier.

Gruppierungen werden optisch durch unterschiedliche Farben dargestellt. Ähnliche Werte erhalten die gleiche Farbe. Für die Stabilisierung des sich darstellenden Netzwerkes können verschiedene Algorithmen benutzt werden, die in der rechten Konfigurationsspalte auswählbar sind. Voreingestellt ist hierbei der **Barnes-Hut-Algorithmus** (siehe J. Barnes, P. Hut (1986): A hierarchical O(N log N) force-calculation algorithm, in *Nature* 324, 446-449; doi:10.1038/324446a0), der die Anzahl der zu berechnenden Wechselwirkungen durch geschicktes Zusammenfassen reduziert.

Um die Darstellungszeiten zu beschleunigen, werden die ersten (genaugenommen 8 mal Knotenanzahl) Berechnungsrunden unsichtbar durchgeführt, danach werden lediglich die Knoten und erst ganz zum Schluss die Kanten (Verbindungslinien) dargestellt.

Auch noch während der Stabilisierung können die Kanten an- oder abgestellt werden und zwar rechts in den Einstellungs-Optionen **edges→hidden**.

Belegstellen und Wortbaum anzeigen

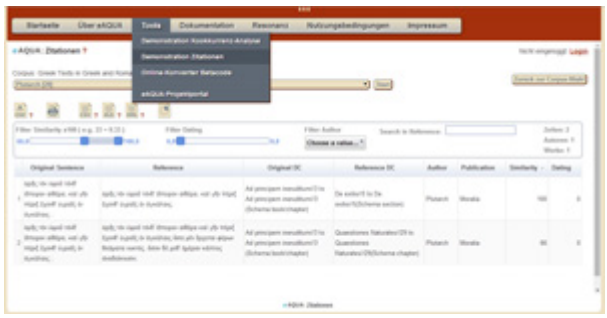
Der Wortbaum stellt mehrere parallele Wortfolgen dar, um zu zeigen, welche Wörter einem Zielwort am häufigsten folgen oder vorangehen. Als Basis dienen die kompletten Tokens (Sätze) aller Satz-Kookkurrenzen zu einem Wort, die hier als Belegstellen bzw. Beispielsätze bezeichnet sind.

Wurde zuvor ein bestimmtes Wort als Kookkurrent zum Zielwort ausgewählt, reduziert sich die angezeigte Belegstellenliste auf die Token, die beide Begriffe enthalten und der Wortbaum rückt an die Stelle vor. Der gleiche Effekt lässt sich erreichen, indem innerhalb des Wortbaumes auf eine Begrifflichkeit geklickt wird.

Es gibt mehrere Möglichkeiten, das Fenster mit den Fundstellen und dem Wortbaum aufzurufen:

- Nach der Suche in den Listen zu den signifikanten Kookkurrenzen, indem auf die Zahl in den runden Klammern geklickt wird.
- Unterhalb der signifikanten Kookkurrenzen, indem auf den Button „Zeige alle“ geklickt wird.
- In der Netzwerk-Visualisierung, indem auf einen Knoten geklickt wird.

Demonstration Zitationen



Die Demonstration Zitationserkennung ist über den Menü-Punkt Tools erreichbar. Sie beinhaltet neben den Analyse-Ergebnissen des Projekt-Portal einige zusätzliche Korpora, die während oder nach der zweiten Förderungsphase des eAQUA-Projektes auch zu Lehrzwecken angelegt wurden.

Für den Login in den geschützten Bereich, in dem Ergebnisse von Korpora abgerufen werden können, die an Benutzungslizenzen gebunden sind, ist am rechten Bildschirmrand ein Link **Login** vorgesehen.

Der Auswahlprozess erfolgt in drei Schritten: Korpus auswählen - Autor(en) und Werk(e) anhaken - auf den Button Start klicken. Bei den als Subkorpora bezeichneten, sind alle Werke eines Autors zusammenfasst. Insbesondere bei diesen ist zu beachten, dass die Ladezeit der Tabellenansicht sehr lange braucht. Das Subkorpus ARISTOTELES et CORPUS ARISTOTELICUM hat eine Ergebnismenge in der Größe von ca. 180 MB. Diese müssen sowohl vom Server an den Browser übertragen werden, was wiederum abhängig von der Anbindung des Nutzers ist, und darüber hinaus vom Browser weiterverarbeitet und dargestellt werden, was gelegentlich auch zum Absturz des Browser führen kann, wie es besonders bei älteren Systemen oder exotischen Konstellationen beobachtet wurde. Aber auch schon bei geringeren Ergebnismengen können Wartezeiten von mehreren Minuten entstehen. Da ist etwas Geduld gefragt. Nachdem in der Tabellenansicht die Ergebnisse der Parallelstellensuche geladen sind, ergeben sich verschieden Filter- und Sortiermöglichkeiten.

Wie die gefilterten Daten exportiert werden können, ist hier notiert. Bei größeren Ergebnismengen ist nur der Export in das Format CSV zu empfehlen, weil hier die Daten direkt vom Browser an den Client geschickt werden können. Sowohl bei XML als auch XLS werden die Listen vom Browser nochmals an den Server zur Umwandlung transportiert und von dort wieder geladen, was die Ausführungszeit erheblich verlängert.

Online-Konverter Betacode

Texte in Beta Code zu transkripieren ermöglicht die unkomplizierte Eingabe von Altgriechisch mit handelsüblicher Tastatur.

Griechischer Beta Code ist die 7-Bit-sichere Kodierung mittels des US-ASCII-Zeichensatzes. Jedes diakritische Zeichen wird durch ein eigenes Zeichen dargestellt, welches dem Buchstaben folgt (Ausnahme: bei Großbuchstaben vor dem Buchstaben). Beta Code unterscheidet nicht zwischen Klein-/Großschreibung, Großbuchstaben werden durch Voranstellung von * Asteriskos (griech. ἀστερίσκος) gekennzeichnet. Einige Projekte benutzen nur Großbuchstaben (z.B. TLG), andere nur Kleinbuchstaben (z.B. das Perseus Project). Beide Varianten werden vom Konverter berücksichtigt. Darüber hinaus gibt es traditionell unterschiedliche Schreibweisen des Endsigmas, welche ebenfalls einstellbar ist.

Der Online-Konverter wandelt den Beta Code in Unicode UTF-8 bzw. Unicode in Beta Code. Dafür sind aus der UTF-8-Codetabelle aus den Bereichen *Greek and Coptic* (U+0370 - U+03FF) und *Greek Extended* (U+1F00 - U+1FFF) die Zeichenkombinationen mit ihren Beta Code Varianten versehen. Ein fallback-Mechanismus versucht dabei, die Codepositions aus dem Bereich *Combining Diacritical Marks* (U+0300 - U+036F) in ihr Equivalent aus den anderen Bereichen umzuwandeln. Alle zusätzlichen Formatierungen, wie beispielsweise Zeilenumbrüche oder HTML-Formatierungen werden vom Konverter entfernt. Über den De-bug-mode, einzustellen am rechten unteren Rand, können sich die Umwandlungspaare schrittweise angezeigt werden.

Projekt-Portal

Sobald man sich über die Hauptseite des eAQUA-Portals eingeloggt hat, sieht man eine Startseite, auf der die verschiedenen digitalen Werkzeuge (Citationsgraph, Differenzanalyse, Klassifikation, Mental Maps, Plautinische Metrik, Suchmaske und Textvervollständigung) zur Auswahl stehen. Die meisten Tools wenden sich an dezidiert altertumswissenschaftliche Fragestellungen und können nicht ohne Weiteres in andere Fachrichtungen übertragen werden. Dagegen sind die Suchmaske und die dahinter liegende Kookkurrenzanalyse sowie die Zitationsanalyse grundsätzlich für alle textbasierten Disziplinen offen und nicht an eine spezielle Fachrichtung gebunden.

Links oben ist der Zugang zur Zitations-Analyse, bezeichnet als Citationsgraph, links unten der zur Kokkurrenz-Suche, bezeichnet als Suchmaske.

Kookkurrenz-Analyse

Nach Klick auf auf das linke, untere Icon erscheint die Suchmaske.

Dort ist zuerst das Suchkorpus mittels anwählen eines Radio-Buttons festzulegen. Danach kann ein Wort in das Suchfeld eingegeben oder einkopiert werden. Es ist auf die korrekte Schreibweise zu achten. Groß- und Kleinschreibung werden genauso beachtet, wie die Verwendung der Diakritika. **Altgriechisch** kann sowohl im Original, beispielsweise durch Benutzung eines entsprechenden Tastaturlayouts (Wie das für einzelne Betriebssystem funktioniert, läßt sich im Internet leicht nachschlagen. Für Windows haben z.B. die Kollegen der Katholisch-Theologischen Fakultät an der Ludwig-Maximilians-Universität München ein Dokument vorbereitet.) oder durch Verwendung einer virtuellen Tastatur. Ebenso besteht hier die Möglichkeit, mittels transkripiertem Betacode zu arbeiten.

Falls die Suche erfolgreich war, wird unterhalb das Ergebnis präsentiert unterteilt in Wörter mit ähnlichem Zusammenhang, den hundert signifikantesten Satzkookkurrenzen (gesamt / links / rechts) sowie den signifikantesten rechten und linken Nachbarn.

Außerdem werden die signifikantesten Kookkurrenzen in einer Visualisierung dargestellt. Das Laden des Netzwerkes kann unter Umständen sehr lange dauern und wird am unteren Rand der Visualisierung angezeigt. Hier haben Sie auch die Möglichkeit, unterschiedliche Signifikanzmaße einzustellen und deren Wertebereiche zu filtern (voreingestellt ist Log-Likelihood). Mehr Informationen zu den Signifikanzmaßen finden Sie hier.

Ganz am Ende finden sich Belegstellen, in denen das Wort Verwendung findet.

Die Listen der Kookkurrenzen sind jeweils mit Querverweisen hinterlegt. Bei Klick auf ein Wort, wird die Kookkurrenzsuche genau für diesen Term gestartet, bei Klick auf die Zahl in Klammern, die die Häufigkeit der Vorkommen angibt, werden Belegstellen mit beiden Ausdrücken angezeigt.

Zitationsgraph

Nach Klick auf auf das linke, obere Icon erscheint eine Liste mit allen Autoren im Korpus.

Nach Auswahl des Autors wird eine zweites Auswahlfeld geladen, wo alle Werke verzeichnet sind, die zum Autoren gefunden wurden. Nachdem das entsprechende Werk angewählt wurde, wird die Datenbankabfrage gestartet und eine Ergebnisliste angezeigt.

Innerhalb der Liste ist für jeden angegebenen Satz hinter dem Stichwort References ein Link auf die Seite mit den Belegstellen dafür verzeichnet. Ganz am Anfang neben Results ist ein Link, der mit „view in new flash layout“ bezeichnet ist und auf die Visualisierung der Ergebnisliste zu dem Werk verweist. Das Laden kann abhängig von der Menge der Belegstellen einige Zeit in Anspruch nehmen.

Nachdem die Visualisierung geladen wurde, ist eine Chart-View mit Informationen zu Autoren, Zeiträumen oder der Anzahl der Zitationen nach unterschiedlichen Kriterien zu sehen. Im linken oberen Bereich findet sich die Möglichkeit, auf eine Tabellenansicht zu wechseln, die neben einer Sortierfunktion es auch vorsieht, innerhalb einzelner Spalten Suchfilter zu definieren. Zur Darstellung der Visualisierung im Browser ist ein Add-on zum Abspielen von SWF-Dateien notwendig, wie zum Beispiel der Adobe Flash Player.

From:
http://www.eaqua.net/doku/ - **Wissensdatenbank**

Permanent link:
http://www.eaqua.net/doku/doku.php/portal?rev=1479395628

Last update: **2021/03/31 12:49**