Table of Contents

The eAQUA portal 2
Access areas 2
Demonstrations 2
Demonstration co-occurrence analysis 2
Visualization of the co-occurrences 3
Show References and Word Tree 4
Demonstration citations 4
Online converter beta code 5
Project portal 6
Co-occurrence analysis 6
Citation graph 7

Fix Me! **This page was translated from German into English using DeepL.**
(remove this paragraph once the translation is finished)

The eAQUA portal

Access areas

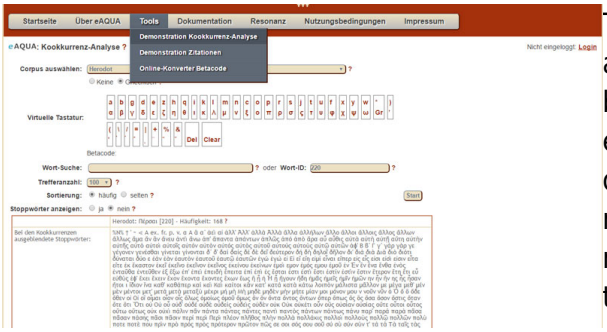
The eAQUA portal is currently divided into two separate areas. In a freely accessible area, which offers a limited use of the possible functionality for demonstration purposes, and a closed area, which is only possible by a login via user ID. Such a separation is primarily due to licensing issues. Various corpora are bound to user licenses of the publishing houses. We are therefore required to check whether a corresponding license is available before issuing user IDs.

Furthermore, there are two frontend versions which use different (programming) techniques. In order to keep up with the developments in the field of dynamic web page generation, including the so-called HTML5 technology, a second version has been created, which is completely based on JavaScript for visualization. The first one uses a proprietary software of the company Adobe, called Flash, for the visual representation, for which an additional browser plugin is necessary.

Therefore, two versions are given in explanations of the tools, which are mainly due to the different visual representations, but not the analysis results.

Demonstrations

Demonstration co-occurrence analysis



The demonstration of the co-occurrence analysis is accessible via the menu item Tools. It contains some basic analysis results from the project portal and was enriched by further corpora in the course of dissemination. For login to the protected area, where results from corpora bound to user licenses can be retrieved, a link **Login** is provided on the right side of the screen.

After selecting the corpus in the upper area, there are several possibilities for a search in the database. In the field **Word Search** it is possible to look for a word or part of a word in the database. The field is provided with an **Autocomplete** that takes effect from **3 letters**. This lists all found words (with their frequencies in brackets). If you click on one of the list results, this word will be taken over. It can be recognized by the fact that there is a number in the **Word ID** field on the right. Clicking the Start button would analyze this word.



If no suggestions are made, it is possible to search the database for hits by entering only the initial letters in the field and then clicking Start. A list of all terms beginning with these letters will appear. The terms are provided with a link that starts the co-occurrence analysis for that one.

If the ID of the word is known from a previous search, it can also be entered in the **Word ID** field and the analysis started.

There is a small restriction in the demo version, in contrast to the portal itself, regarding the maximum number of hits to be displayed, which can be defined via the **Number of hits** dropdown. This is necessary because of the dynamically loaded visualization, so that the page does not take too long to load. To be able to find also unsignificant co-occurrences, an additional **sort** filter is built in. For example, the 100 most unsignificant co-occurrences can be displayed for a word by setting the number of hits to 100 and the sorting to rare.



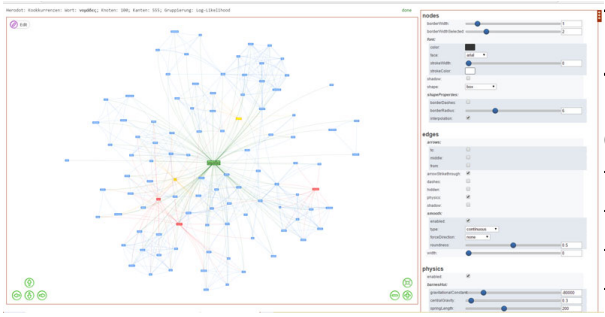
Similar to the portal, the result is presented as a list divided into words with a similar context, the hundred most significant (or here also the most insignificant) sentence co-occurrences (total / left / right) as well as the most (un-)significant right and left neighbors. In addition, the most (un-)significant co-occurrences are shown in a visualization. To call this up, two buttons are displayed in the left column below “Significant co-occurrences”.

At the very end, there are passages in which the word is used.

The lists of co-occurrences are cross-referenced. By clicking on a word, the co-occurrence search is started exactly for this term, by clicking on the number in brackets, which indicates the frequency of occurrence, citations with both expressions are displayed.

How to export data from the co-occurrence analysis is noted here.

Visualization of the co-occurrences



The window with the network visualization written in JavaScript vis.js (http://visjs.org/; Apache 2.0 / MIT-Lizenz), is called up via the two buttons in the left column below “Significant co-occurrences”, grouped on the one hand according to the **frequency** (frequency) of the co-occurrences, and on the other hand according to the significance measure **log-likelihood**. There are a few explanations about the significance measures and their meaning here.

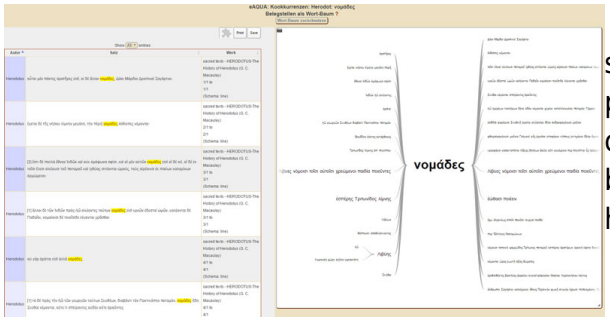
Groupings are visually represented by different colors. Similar values receive the same color. For the stabilization of the displayed network different algorithms can be used, which are selectable in the right configuration column. The default setting is the **Barnes-Hut Algorithm** (siehe J. Barnes, P. Hut

(1986): A hierarchical O(N log N) force-calculation algorithm, in *Nature* 324, 446-449; doi:10.1038/324446a0), which reduces the number of interactions to be calculated by clever grouping.

To speed up the display times, the first (8 times the number of nodes, to be exact) calculation rounds are performed invisibly, then only the nodes are displayed and only at the very end the edges (connecting lines).

Even during stabilization, the edges can be switched on or off in the settings options **edges→hidden** on the right.

Show References and Word Tree



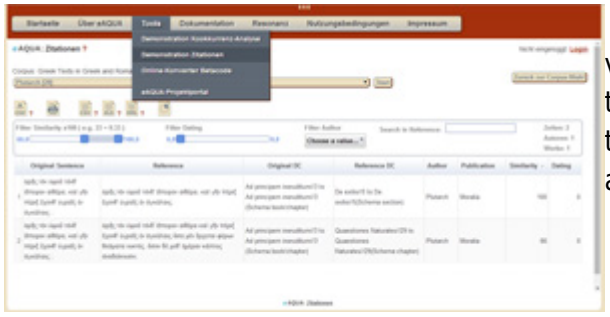
The word tree represents several parallel word sequences to show which words most often follow or precede a target word. The complete tokens (sentences) of all sentence co-occurrences to a word serve as a basis, which are called vouchers or example sentences here.

If a certain word was previously selected as a co-occurrence to the target word, the displayed token list is reduced to the tokens that contain both terms and the word tree advances to the position. The same effect can be achieved by clicking on a term within the word tree.

There are several ways to access the Findings and Word Tree window:

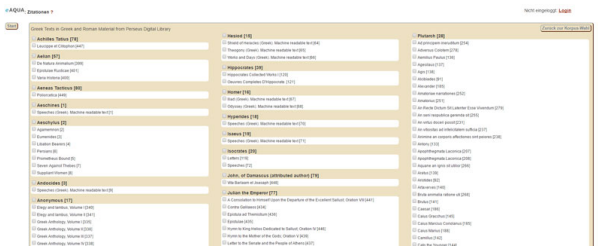
- After searching the lists to the significant co-occurrences by clicking on the number in the round brackets.
- Below the significant co-occurrences, by clicking on the “Show all” button.
- In the network visualization, by clicking on a node.

Demonstration citations



The citation detection demonstration can be accessed via the Tools menu item. Besides the analysis results of the project portal, it contains some additional corpora that were also created for teaching purposes during or after the second funding phase of the eAQUA project.

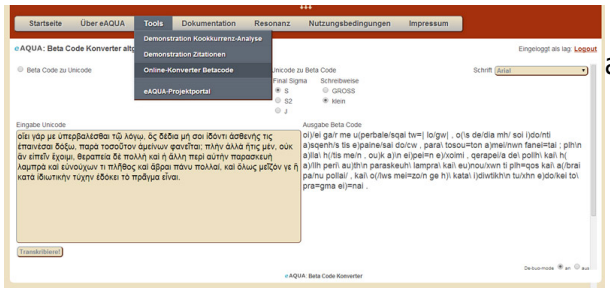
For login to the protected area, where results of corpora bound to user licenses can be retrieved, a link **Login** is provided on the right side of the screen.



The selection process takes place in three steps: select corpus – check author(s) and work(s) – click on the Start button on the far left of the window. In the case of the selection options known as subcorpora, all the works of an author are grouped together, so that here the last two steps are omitted and the loading of the data is started immediately. Especially with these subcorpora, it should be noted that the loading times of the table view can be very long. For example, the TLG subcorpus ARISTOTELES et CORPUS ARISTOTELICUM has a result set in the size of about 180 MB. These have to be transferred from the server to the browser, which depends on the connection of the user, and furthermore have to be processed and displayed by the browser, which can occasionally lead to a crash of the browser, as it has been observed especially with older systems or exotic constellations. However, waiting times of several minutes can occur even with smaller result sets. A little patience is required. After the results of the parallel search have been loaded in the table view, various filtering and sorting options are available.

How to export the filtered data is noted here. For larger result sets, only the export to CSV format is recommended, because here the data can be sent directly from the browser to the client. With both XML and XLS, the lists are transported again from the browser to the server for conversion and loaded again from there, which considerably increases the execution time.

Online converter beta code

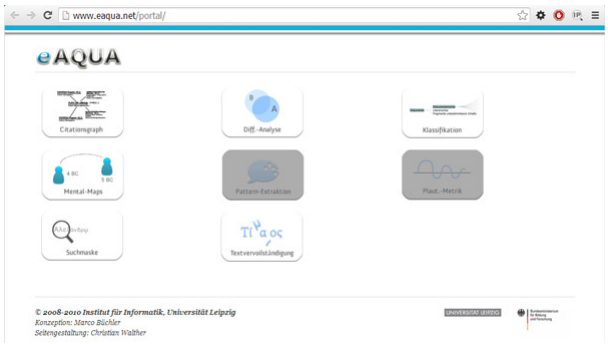


Transcribing texts in Beta Code allows for easy input of ancient Greek using a standard keyboard.

Greek Beta Code is the 7-bit secure encoding using the US ASCII character set. Each diacritical character is represented by a separate character following the letter (exception: for uppercase letters preceding the letter). Beta Code does not distinguish between lowercase and uppercase letters; uppercase letters are preceded by * Asteriskos (Greek ἀστερίσκος). Some projects use only uppercase letters (e.g. TLG), others only lowercase letters (e.g. the Perseus Project). Both variants are taken into account by the converter. In addition, there are traditionally different spellings of the final sigma, which can also be set.

The online converter converts the beta code into Unicode UTF-8 or Unicode into beta code. For this purpose, the character combinations from the UTF-8 code table from the areas *Greek and Coptic* (U+0370 - U+03FF) and *Greek Extended* (U+1F00 - U+1FFF) are provided with their beta code variants. A fallback mechanism tries to convert the code positions from the *Combining Diacritical Marks* (U+0300 - U+036F) area into their equivalent from the other areas. All additional formatting, such as line breaks or HTML formatting, is removed by the converter. Via the de-bug-mode, set at the bottom right edge, the conversion pairs can be displayed step by step.

Project portal



As soon as one has logged in via the main page of the eAQUA portal, one sees a start page on which the various digital tools (citation graph, difference analysis, classification, mental maps, plautinic metrics, search mask and text completion) are available for selection. Most of the tools address decidedly ancient studies questions and cannot be easily transferred to other disciplines. In contrast, the search mask and the underlying co-occurrence analysis as well as the citation analysis are basically open to all text-based disciplines and not bound to a specific discipline.

On the top left is the access to the citation analysis, called citation graph, on the bottom left the access to the co-occurrence search, called search mask.

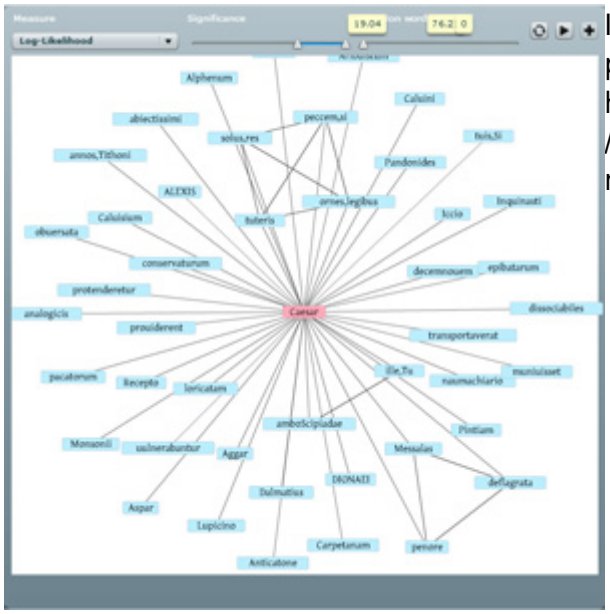
Co-occurrence analysis



After clicking on the lower left icon, the search mask appears.



There, the search corpus must first be defined by selecting a radio button. Then a word can be entered or copied into the search field. Attention must be paid to the correct spelling. Upper and lower case are observed as well as the use of diacritics. **Ancient Greek** can be used both in the original, for example by using an appropriate keyboard layout (How this works for individual operating systems can be easily looked up on the Internet. For Windows, for example, the colleagues of the Faculty of Catholic Theology at the Ludwig Maximilian University of Munich have prepared a document) or by using a virtual keyboard. There is also the possibility to work by means of transcribed beta code.



If the search was successful, below the result is presented divided into words with similar context, the hundred most significant sentence co-occurrences (total / left / right) and the most significant right and left neighbors.

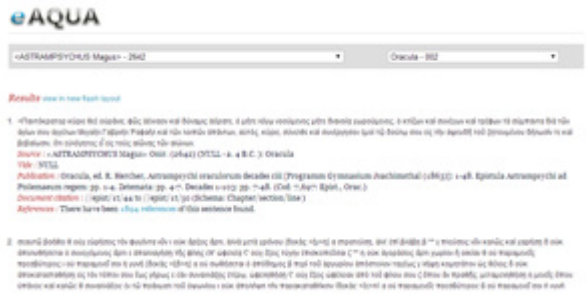
In addition, the most significant co-occurrences are displayed in a visualization. Loading the network may take a long time and is displayed at the bottom of the visualization. Here you also have the possibility to set different significance measures and to filter their value ranges (default is log-likelihood). More information about the significance measures can be found here. At the very end, there are passages in which the word is used.

The lists of co-occurrences are cross-referenced. By clicking on a word, the co-occurrence search is started exactly for this term, by clicking on the number in brackets, which indicates the frequency of occurrence, citations with both expressions are displayed.

Citation graph



After clicking on the upper left icon, a list of all authors in the corpus appears.



After selecting the author, a second selection field is loaded, where all works are listed that were found for the author. After selecting the corresponding work, the database query is started and a list of results is displayed.

Within the list, for each given sentence, after the keyword References, there is a link to the page with the references for it. At the very beginning, next to Results, there is a link labeled “view in new flash layout” that points to the visualization of the results list for the work. Loading may take some time depending on the amount of references.



After the visualization has been loaded, a chart view with information on authors, time periods or the number of citations according to different criteria can be seen. In the upper left area, there is the option to switch to a table view, which, in addition to a sorting function, also provides for defining search filters within individual columns. To display the visualization in the browser, an add-on for playing SWF files is required, such as Adobe Flash Player.

From:
http://replicatio.science/dokuwiki/ - **documentatio replicationis**

Permanent link:
http://replicatio.science/dokuwiki/doku.php/en/eaqua/portal?rev=1680192586

Last update: **2023-03-30**