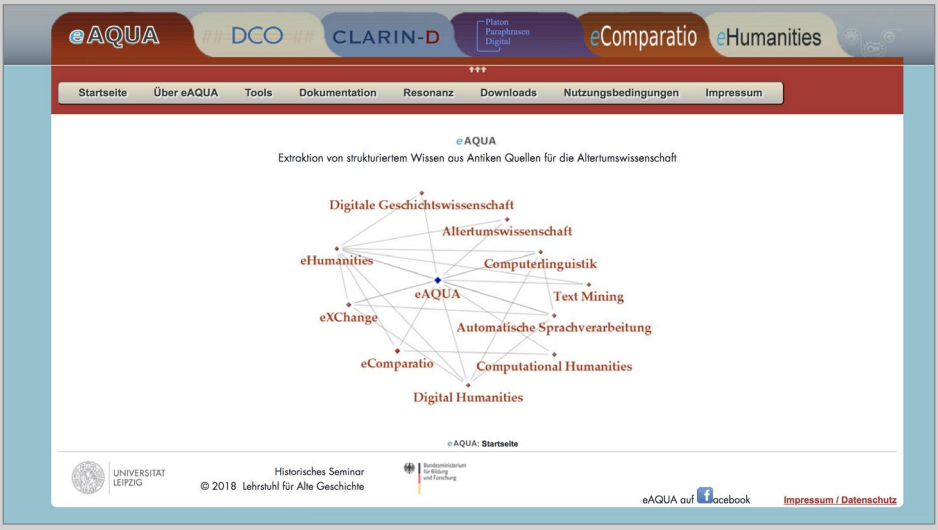

Abbildung 1. Startseite www.eaqua.net

eAQUA – Was ist das?

Zur Geschichte des Projektes

Das heutige Portal eAQUA (Extraktion von strukturiertem Wissen aus Antiken Quellen für die Altertumswissenschaft) ist aus einer Projektförderung des Bundesministeriums für Bildung und Forschung (BMBF) hervorgegangen (■ **Abbildung 1**). Die Förderung umfasste eine erste Phase von 2008 bis 2011, in der Altertumswissenschaftler[1], Frühneuzeitler und Informatiker zusammenarbeiteten, um die Anwendung fortgeschrittener Werkzeuge aus dem Bereich des Text Mining für die beteiligten Fachdisziplinen erstmals experimentell zu erproben. Das Projekt wurde damals im Rahmen des vom BMBF im Förderschwerpunkt „Geistes- und Sozialwissenschaften" aufgelegten Programms „Wechselwirkungen zwischen Geistes- und Naturwissenschaften" (2008–2011) gefördert und umfasste in diesem Zeitrahmen acht Teilprojekte, die Verfahren aus dem Bereich des Text Mining für Anwendungsszenarien unterschiedlicher Genres und Quellengattungen entwickelten:

- Projekt Atthidographen (Leitung: Ch. Schubert, Alte Geschichte, Universität Leipzig)
- Projekt Platon (Leitung: K. Sier, Gräzistik, Universität Leipzig)
- Projekt Metrik (Leitung: M. Deufert, Latinistik, Universität Leipzig)
- Projekt Camena (Leitung: W. Kühlmann, Germanistik, Universität Heidelberg)
- Projekt Inschriften (Leitung: B. Meißner, Alte Geschichte, Universität der Bundeswehr Hamburg)
- Projekt Papyri (Leitung: R. Scholl, Alte Geschichte, Universität Leipzig)
- Projekt Fehlererkennung (Leitung: G. Heyer, Informatik, Universität Leipzig)
- Projekt Mental Maps (Leitung: Ch. Schubert, Alte Geschichte, Universität Leipzig).

1 Das Autorenteam hat sich beim Formulieren der Texte um eine genderneutrale Sprache bemüht. In den wenigen Fällen, in denen dies nicht möglich war, haben wir uns zugunsten der Lesbarkeit und des flüssigen Textlaufs für das generische Maskulinum entschieden.

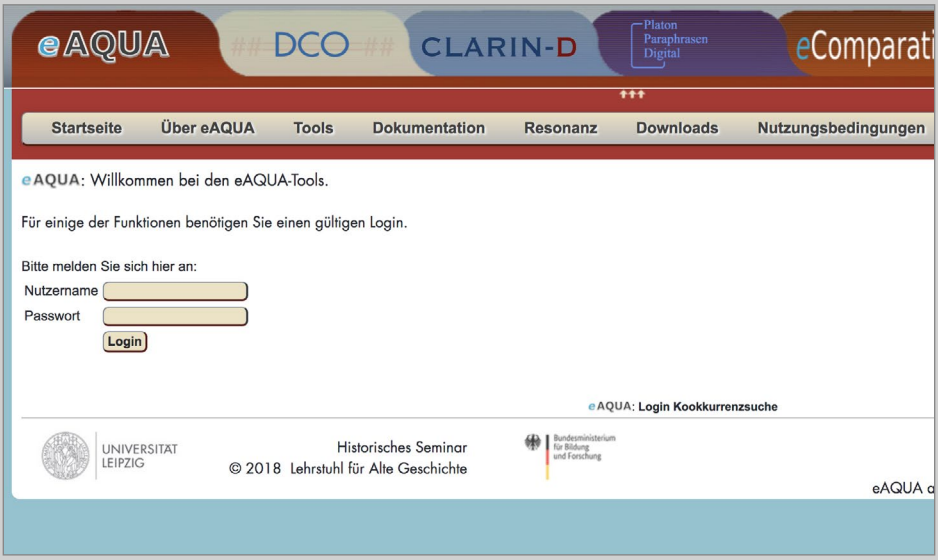

Abbildung 2. Login geschützter Zugang

Ziel der interdisziplinären Zusammenarbeit zwischen geisteswissenschaftlichen Fächern und der Informatik war es, neues und strukturiertes Wissen aus antiken oder frühneuzeitlichen Quellen zu gewinnen und dabei Werkzeuge und Verfahren aus dem Segment des Text Mining weiterzuentwickeln.

In der zweiten Projektphase (2011–2013) wurde das Projekt in der alleinigen Verantwortung des Lehrstuhls für Alte Geschichte vom BMBF weitergefördert, die über den reinen Projektstatus hinausführen sollte, um eine nachhaltige und breite Nutzung zu ermöglichen.

Die in diesem Buch vorrangig behandelten Suchvarianten Kookkurrenzsuche und Parallelstellensuche,[2] die in dieser zweiten Projektphase maßgeblich auf Anwendungsstabilität hin weiterentwickelt wurden, gehen über die üblichen Möglichkeiten digitaler Bibliotheken hinaus und ermöglichen die Erschließung von Abhängigkeiten, Einflüssen und Transferwegen des Wissens in größerem Maßstab.

Die Kookkurrenzsuche, vorrangig im eAQUA-Teilprojekt Atthidographen 2008–2011 entwickelt, führt heute zur Erschließung von semantischen Zusammenhängen, die Zitationssuche, vorrangig im eAQUA-Teilprojekt Platon 2008–2011 entwickelt, ermöglicht heute die Auflistung von Textpassagen, die Parallelen zwischen einem Werk und dem gesamten Referenzkorpus darstellen. Die hier vorgestellten Tools sind in der zweiten Förderphase des Projektes 2011–2013, sowie auch seither umfangreich weiterentwickelt und verbessert worden. Sie werden auf der Internetseite von eAQUA[3] zur Anwendung sowohl für frei zugängliche Textkorpora als auch für Korpora, die einer Lizenzpflicht unterliegen, für alle Interessenten angeboten. Bei Benutzung lizenzpflichtiger Korpora ist nach gegenwärtigem Stand des Urheberrechts[4] ein individuell zu vergebender Zugang in den geschützten Bereich der Webseite notwendig (■ **Abbildung 2**). Der geschützte Zugang trägt dabei dem Umstand Rechnung, dass es laut Urheberrechtsgesetz zwar erlaubt ist, 75 Prozent eines Werkes für die eigene Forschung weiterzuverarbeiten, aber nur bis zu 15 Prozent eines veröffentlichten Werkes vervielfältigt, verbreitet, öffentlich zugänglich gemacht oder wiedergegeben werden dürfen.[5]

---

2 Parallelstellen- und Zitationssuche werden synonym verwendet, auf eine genaue Unterscheidung wird hier verzichtet.

3 URL: http://www.eaqua.net.

4 Urheberrechtsreform 2018 mit dem Namen „Urheberrechts-Wissensgesellschafts-Gesetz (UrhWissG)".

5 Gesetz über Urheberrecht und verwandte Schutzrechte – UrhG: Teil 1, Abschnitt 6, Unterabschnitt 4: Gesetzlich erlaubte Nutzungen für Unterricht, Wissenschaft und Institutionen; insbesondere §60c und §60d. URL: https://www.gesetze-im-internet.de/urhg/.

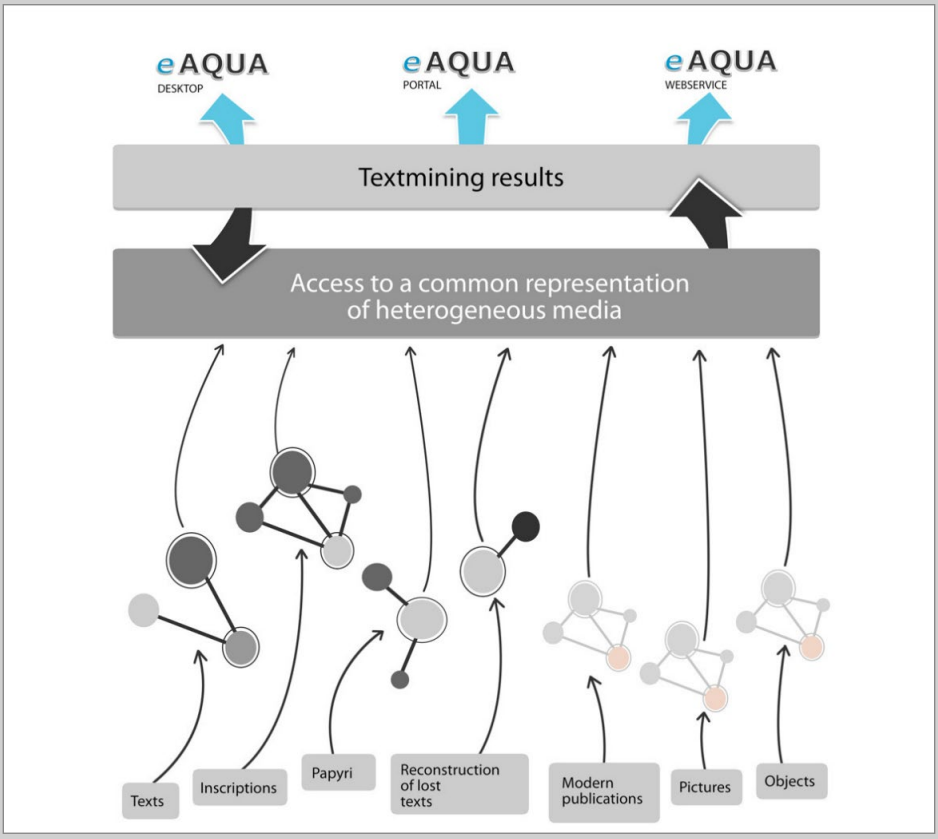

Abbildung 3. Das Portal eAQUA[6]

6 Charlotte Schubert, Gerhard Heyer: Working Papers Contested Order: Das Portal eAQUA – Neue Methoden in der geisteswissenschaftlichen Forschung I – Eine Einführung in das Portal eAQUA; Abb. 1, Seite 7; DOI: https://doi.org/10.11588/ea.2010.0.

Technische Komponenten von eAQUA

In technischer Sicht bestehen die unter dem Begriff eAQUA subsumierten Software-Entwicklungen aus zwei Kernbereichen (■ **Abbildung 3**). Auf der einen Seite aus einer Reihe von Tools, die sogenanntes Text Mining betreiben, die also digitalisierte Korpora verarbeiten, berechnen und Ergebnisse ermitteln. Diese Tools laufen in separierten Serverumgebungen mit entsprechender Rechenkapazität, sind jedoch zumeist so konzipiert, dass sie auch auf Desktop-Rechnern zum Einsatz kommen könnten, wenn die benötigten Laufzeitumgebungen installiert sind. Bei der Konzeption wurde auf eine aufwendig zu programmierende GUI verzichtet, die Programme werden mittels Konsolenbefehlen in der Shell gestartet.

Die so bezeichnete eAQUA-Toolchain besteht aus einer Ansammlung von JAVA-Programmen und Shell-Skripten, die mittels Apache Ant erstellt wurden und demzufolge eine Java-Laufzeitumgebung (JRE) benötigen. Darüber hinaus sollten die für die Erzeugung der Datenbanken nötigen Datenbankmanagementsysteme (DBMS), MySQL oder MariaDB vorhanden und installiert sein, da die Berechnungsergebnisse in diese eingespielt werden.

Auf der anderen Seite gibt es webbasierte Anwendungen zur Präsentation der Ergebnisse und zur Interaktion von Nutzern in dem aufbereiteten Datenmaterial. Im Wesentlichen setzen die webbasierten Programme auf einen LAMP Stack (Linux, Apache, MySQL, PHP) auf, sind also so konzipiert, dass sie auf einem üblichen Web-Server einsatzfähig sind. Konkret werden alle dynamischen Inhalte mittels PHP aus MySQL- (bzw. MariaDB-) Datenbanken ausgelesen und zu statisch auslieferbaren Inhalten verarbeitet. Dabei werden nicht nur HTML-Seiten erzeugt, sondern die Daten auch in JSON- oder CSV-Dateien umgeschrieben, um sie später über JavaScript-Bibliotheken darzustellen bzw. interaktiv herunterladen zu können.

Zur Visualisierung innerhalb der Internetseiten kommen freie JavaScript-Bibliotheken zum Einsatz, die entweder unter Creative Commons, Apache 2.0 oder MIT lizensiert sind. Dabei handelt es sich um Google Visualization API, jQuery, jquery-svg-pan-zoom und vis.js.

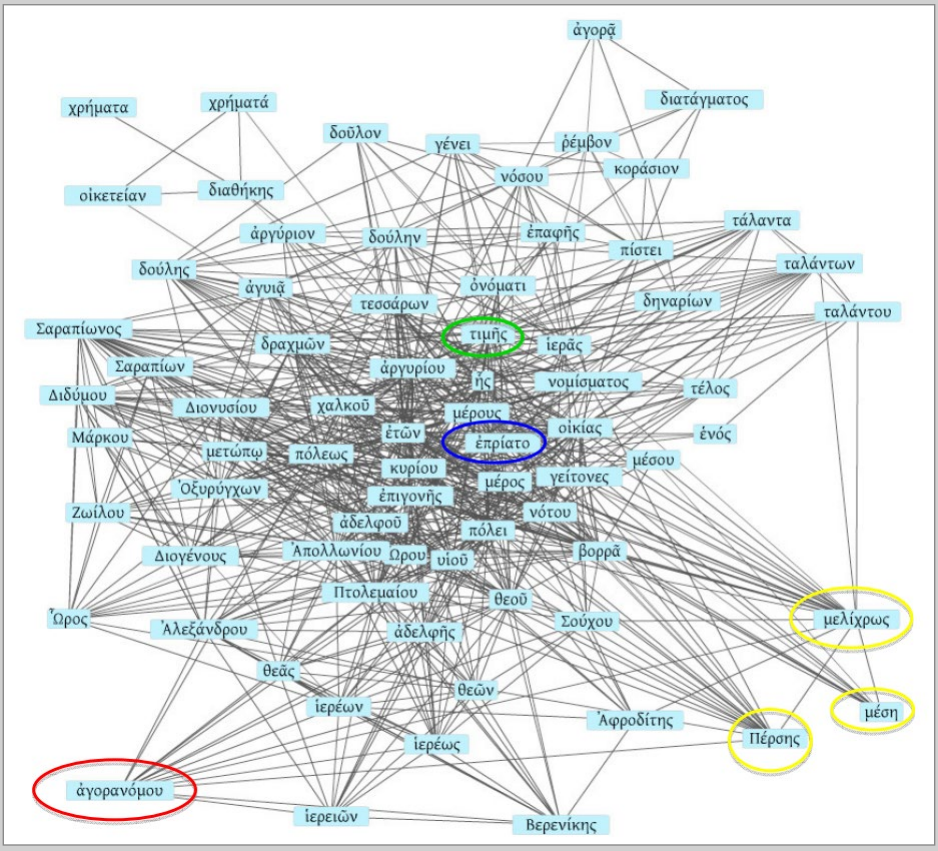

Abbildung 4. Der ursprüngliche Kookkurrenzgraph in Flash[7]

7 Charlotte Schubert, Gerhard Heyer: Working Papers Contested Order: Das Portal eAQUA – Neue Methoden in der geisteswissenschaftlichen Forschung I – Eine Einführung in das Portal eAQUA; Abb. 7, Seite 7; DOI: https://doi.org/10.11588/ea.2010.0.

Im älteren Begleitmaterial zum Portal eAQUA[8] ist von einer Cocoon-basierten Architektur die Rede, welche bei der internen und externen Darstellung vollständig auf XML setzte. Aus Performance- und Lesbarkeitsgründen lassen sich die Ergebnisse, hier insbesondere die der Kookkurrenz- und Parallelstellenanalyse nur unzureichend in XML darstellen, so dass im Zuge der Neugestaltung der webbasierten Anwendungen nicht nur auf die zuvor eingesetzte Rich-Client-Technologie Flash,[9] sondern auch auf die Java-Servlet-Technologie zur Darstellung von Webinhalten verzichtet wurde (■ **Abbildung 4**). Im Bereich des Preprocessing wird weiterhin auf die ANT-basierte Toolchain zurückgegriffen, die für die Aufbereitung und Extraktion der Texte, die Segmentierung der Sätze, Tokenisierung von Wörtern bis hin zur Erstellung der Datenbank eingesetzt wird.

8 Beispielsweise in den eAQUA Working Papers, einer Reihe der Working Papers Contested Order des Profilbildenden Bereichs Contested Order der Universität Leipzig; URL: https://journals.ub.uni-heidelberg.de/index.php/eaqua-wp.

9 Insbesondere als Visualisierungskomponente.