

Bearbeitung und Annotation historischer Texte mittels Graph-Datenbanken am Beispiel der Chronik des Matthias von Edessa

Kaufmann, Sascha

sascha.kaufmann@dh.unibe.ch
Universität Bern, Schweiz

Andrews, Tara Lee

tara.andrews@kps.unibe.ch
Universität Bern, Schweiz

Die Chronik des Matthias von Edessa ist den meisten Wissenschaftlern des mittelalterlichen Nahen Ostens und des ersten Kreuzzuges bekannt für ihren Reichtum an Informationen, sowie der (mutmaßlichen) Ignoranz und Naivität ihres Autors. Matthias von Edessa war ein Armenischer Priester, der in der Kreuzfahrer-Grafschaft Edessa lebte und die Chronik zwischen den Jahren 1110 und 1132 verfasste. Gleichwohl der Text oft von Historikern verwendet wird, liegt er bis heute in keiner kritischen Edition vor und wurde zuletzt 1898 (Matthias von Edessa 1898) veröffentlicht. Die Chronik umfasst 35 (kopierte) Manuskripte, deren ältestes auf mindestens 450 Jahre nach dem Tod des Autors datiert werden kann. Diese werden zur Zeit für eine digitale Gesamtedition vorbereitet.

Die Herausforderungen, die sich dabei stellen, beschränken sich nicht nur auf die Bearbeitung aus philologischer Sicht, sondern auch auf die Annotation und Präsentation als historisches Werk, mit dem Ziel, sie auch als Plattform für Historiker zur Verfügung zu stellen (z. B. mit Zeit- und Ortsangaben, etc.). Hierzu greifen wir zum Einen auf eine Reihe aktueller Methoden und Werkzeuge der digitalen Philologie zurück, wie zum Beispiel die Transkription aller Manuskripte, palaeographisches Markup unter der Benutzung des TEI-Vokabulars (TEI Consortium 2015), automatische Manuskript-Kollation mit CollateX (Dekker et al. 2014), stemmatische Analyse mit Hilfe der Werkzeuge von Stemmaweb (Andrews / Macé 2013) und der Publikation aller Transkriptionen, sowie einer editorischen Rekonstruktion des Textes. Zum Anderen werden auch Textkommentare von der digitalen Plattform profitieren, da sie mit weiteren Informationen angereichert werden können. So können unter Anderem Ortsnamen nicht nur ausgezeichnet („getagged“) werden, sondern ihre mögliche Lokalisierung auch angegeben und soweit möglich geographisch angezeigt werden. Personen und ethnographische Bezeichnungen werden nicht nur in einem Index erfasst, sie werden, soweit möglich, mit

prosopographischen Datenbanken oder relevanten Seiten auf Wikipedia verlinkt.

Das derzeitige Projekt *The Chronicle of Matthew of Edessa Online* wird bis 2018 von dem Schweizer National Fond (SNF) finanziert und baut auf verschiedenen Projekten und gesammelten Erfahrungen der vergangenen fünf Jahre auf.

Ziel ist es, dem Forscher oder der Forscherin ein Werkzeug in die Hand zu geben, das es ihm / ihr erlaubt Bewegungen von Individuen und Gruppen über Raum und Zeit zu verfolgen, indem wir die Textedition selbst zu einer Plattform für mittelalterliche Geschichte machen.

Im Folgenden werden wir zwei, aus technischer Sicht wesentliche Aspekte des Projektes vorstellen (Speichern und Wiedergabe von Informationen) und erste Ergebnisse präsentieren.

Effizientes Speichern und Bearbeiten von Manuskripten und Annotationen

Aufbauend auf den Erkenntnissen, die wir aus dem Stemmaweb-Projekt gewonnen haben, werden die Manuskripte in Form von Graphen gespeichert. Daher ist für uns der naheliegendste Schritt, die gesamte Arbeit des Benutzers im Graph zu repräsentieren. Als mögliche Darstellungsformen für den Benutzer kommen hierfür Online-Präsentationen oder ein Export, zum Beispiel nach TEI, in Betracht.

Während die Speicherung in Stemmaweb als Perl-Objekte in einer Relationalen-Datenbank (MySQL (Oracle Cooperation 2015)) erfolgt, haben wir uns für dieses Projekt dafür entschieden, Neo4J (Neo Technology Inc 2015), eine Graph-Datenbank zu verwenden. Zum einen kommt dies unserer intern verwendeten Datenstruktur entgegen, die auf einer Modellierung auf Graphen basiert. Damit verbunden hat dies auch den Vorteil, dass viele, für uns wichtige Funktionen (z. B. Depth First Search, Breadth First Search, etc.), nativ vom Datenbanksystem zur Verfügung gestellt werden und speziell benötigte Operationen, wie z. B. Plausibilitätsprüfungen, einfacher implementiert und effizienter ausgeführt werden können.

Des Weiteren migrieren wir derzeit große Teile, des aus Stemmaweb vorliegenden Quellcodes, von Perl nach Java. Gleichzeitig überarbeiten wir die API, mit dem Ziel, den Service auch über eine Web-API nutzen zu können. Die Hauptgründe hierfür sind, die bestehende Stemmaweb Backend-Engine zu modernisieren und effizienter zu gestalten. Dieser Schritt soll bis Ende 2015 abgeschlossen sein.

Parallel zu der Migration, haben wir bereits einen ersten Prototypen des Editions-Interfaces erstellt. Dieser Prototyp basiert bereits auf Neo4J und implementiert ein Textmodell das interoperabel mit Stemmaweb ist und wir

zu Testzwecken bereits mit zusätzlichen Annotationen versehen haben.

In Abbildung 1 ist ein typischer Ausschnitt unserer Neo4J Datenbankstruktur zu sehen. Man erkennt, dass sich eine „TRADITION“ aus ein oder mehreren „SECTION“-Knoten zusammensetzt, die wiederum mittels gerichteter „NEXT“-Kanten miteinander verbunden sind und somit ihre Reihenfolge gewährleistet bleibt. Jeder „SECTION“-Knoten verweist wiederum auf eine Folge von untereinander mit „LEMMA_TEXT“ verbundenen „READING“-Knoten. Darüber hinaus existieren noch sogenannte „TRANSLATION“-Knoten, die die jeweilige Übersetzung in einer Sequenz aus READING-Knoten speichern. Hierbei bleibt es dem Bearbeiter freigestellt, die Granularität der Übersetzung festzulegen. So sind Einheiten, wie zum Beispiel Sigle-Wörter, Sätze oder ganze Paragraphen, etc. vorstellbar.

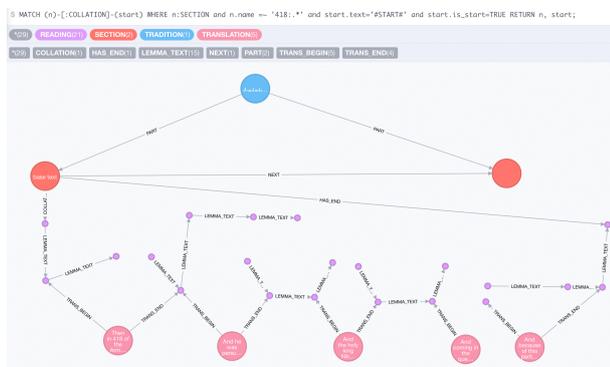


Abb. 1 : Beispiel für die derzeitige Datenbankstruktur in Neo4J. Gut zu Erkennen ist die Gliederung in Tradition (blau), Sections (orange), Readings (pink) und Translations (rosa).

Eine adäquate Wiedergabe der gespeicherten Informationen

Wie bereits erwähnt, liegt ein weiterer Schwerpunkt des Projektes auf einer angemessenen Wiedergabe der im System vorhandenen Informationen. Aus diesem Grund haben wir, parallel zu den Arbeiten am Backend, bereits damit begonnen, den Prototypen einer Webseite zur Darstellung der Informationen zu implementieren.

Dieser Prototyp stellt zur Zeit vier Abschnitte (Sections) aus der Chronik bereit, zwischen denen der Benutzer auswählen kann (Abbildung 2). Zu einem ausgewählten Abschnitt werden dessen Transkription, sowie (englische) Übersetzung angezeigt. Eine Karte, in Form einer eingebetteten Google-Map (Google Inc. 2012), zeigt zusätzlich alle im Abschnitt vorkommenden Örtlichkeiten an, sofern sie lokalisierbar sind. In der Transkription werden Textteile, zu denen Zusatzinformationen vorliegen, dem Benutzer farblich kodiert angezeigt (Ortsangaben (blau), Personenangaben (rot) und Zeitangaben (gelb)).

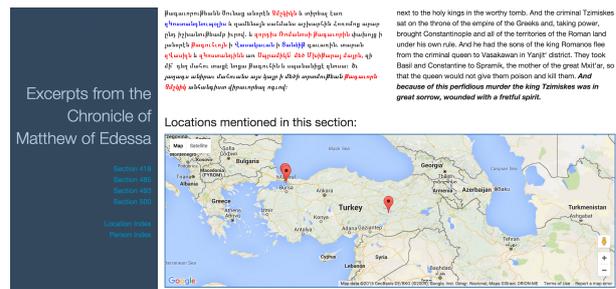


Abb. 2 : Screenshot der Webseite. Neben der Transkription (linke Spalte), befinden sich die englische Übersetzung (rechte Spalte), sowie ein Karte mit Positionsmarkierungen der im Text vorkommenden Ortsbeschreibungen.

Führt nun der Benutzer mit dem Mauszeiger über einen solchen farblich hervorgehobenen Textteil, werden zu diesem in einem Popup-Fenster weitere Informationen oder Verlinkungen angezeigt. So ist zum Beispiel für Ortsangaben, wie in Abbildung 3 zu sehen, dies üblicherweise der Ortsname, sowie ein vergrößerter Ausschnitt der Karte mit einer Markierung der genauen Position, soweit diese bekannt ist.

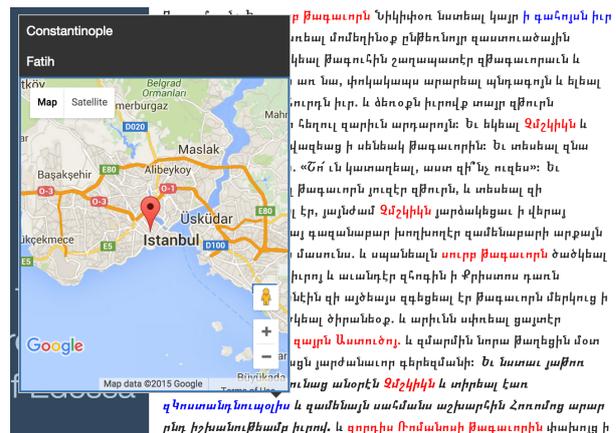


Abb. 3: Beispiel für Zusatzinformationen bezüglich Constantinople (Faith).

Bei Personenangaben, siehe Abbildung 4, wird die Übersetzung des Namens, der volle Name und weitere Informationen, sowie Links, die zu Einträgen bezüglich der Person auf externen Seiten verweisen, hier Links zu *Wikipedia* (Wikimedia Foundation 2015) und *Prosopography of the Byzantine World* (Jeffreys et al. 2011), in einem Popup-Fenster angezeigt.

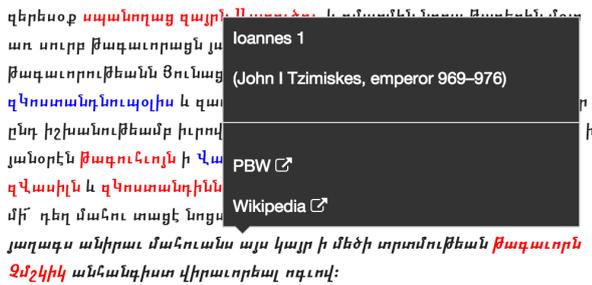


Abb. 4 : Beispiel für Zusatzinformationen bezüglich *Ioannes I (John I Tzimiskes)*.

Darüber hinaus arbeiten wir daran, sowohl individuelle Textzeugen als auch verschiedene Variationen eines Textes anzeigen zu können.

Hervorzuheben ist, wie oben bereits angedeutet, das die Plattform auch die Möglichkeit bietet, die Edition zu exportieren. Hierzu werden relevante Standards wie zum Beispiel TEI oder CIDOC-CRM (International Council of Museums 2014) angeboten.

Zusammenfassung und Ausblick

Wir haben einen kurzen Einblick in das Projekt The Chronicle of Matthew of Edessa Online gegeben. Dieses hat als Ziel eine digitale Plattform zur Verfügung zu stellen, welche das Untersuchen und Bearbeiten von Manuskripten aus philologischer, als auch historischer Sichtweise unterstützt. Hierzu haben wir einen Einblick in technische Aspekte und dem derzeitigen Stand der Implementierung gegeben. Zusammenfassend kann man sagen, dass wir schon eine gute Strecke zurückgelegt haben und dass das Projekt noch einige interessante Aufgaben für uns bereit hält .

Fußnoten

1. Stemmaweb bietet seinen Benutzern die Möglichkeit mit Text-Kollationen und Varianten zur stemmatischen Analyse zu arbeiten und diese zu modifizieren. Des Weiteren können Texte basierend auf Kollationen und Stemmata rekonstruiert und erstellt werden. Hierfür hat es sich bewährt, Texte in einem Graph abzubilden.
2. Dieser Ansatz wurde gewählt, da eine automatische Kollation für längere Texte am besten mit einer Einteilung in diskrete Abschnitte (Sections) gelingt. Desweiteren können wir somit Textzeugen, deren Abschnitte abweichend angeordnet sind, ohne Schwierigkeiten darstellen
3. Interessante Aufgaben wären zum Beispiel die Gestaltung der Web-Oberfläche, die Interaktion mit dem Benutzer, Optimierung der Graph-Datenbank, Web-API oder die Bereitstellung verschiedener Exportformate, um nur einige zu nennen.

Bibliographie

- Andrews, Tara L. / Macé, Caroline** (2013): „Beyond the tree of texts: Building an empirical model of scribal variation through graph analysis of texts and stemmata“, in: *Literary and Linguistic Computing* 28, 4: 504-21 <http://dx.doi.org/10.1093/lc/fqu072> [letzter Zugriff 09. Februar 2016].
- Dekker, Ronald Haentjens / Hulle, Dirk van / Middell, Gregor / Neyt, Vincent / Zundert, Joris van** (2014): „Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project“, in *Literary and Linguistic Computing* 25: 1-19 <http://dx.doi.org/10.1093/lc/fqu007> [letzter Zugriff 09. Februar 2016].
- Google Inc** (2012): *Google Maps* <http://www.google.com/maps> [letzter Zugriff 15. Oktober 2015].
- International Council of Museums** (2014): *The CIDOC Conceptual Reference Model* <http://www.cidoc-crm.org> [letzter Zugriff 15. Oktober 2015].
- Jeffreys, Michael** (2011): *Prosopography of the Byzantine World* www.pbw.kcl.ac.uk [letzter Zugriff 15. Oktober 2015].
- Matthias von Edessa (Matt#e#os Ur#hayec#i)** (1898): *Die Croniken des Matthias von Edessa* [Žamanakagrut#iwn.Va#aršapat](http://www.zamanakagrut.iwn.Va#aršapat).
- Neo Technology Inc.** (2015): *Neo4j* <http://www.neo4j.com> [letzter Zugriff 15. Oktober 2015].
- Oracle Cooperation** (2015): *MySQL* <https://www.mysql.com> [letzter Zugriff 15. Oktober 2015].
- TEI Consortium** (2015): *Guidelines for Electronic Text Encoding and Interchange. Version 2.8.0* <http://www.tei-c.org/Vault/P5/2.8.0/doc/tei-p5-doc/en/html/> [letzter Zugriff 15. Oktober 2015].
- Wikimedia Foundation** (2015): *Wikipedia* <http://www.wikipedia.org> [letzter Zugriff 15. Oktober 2015].