

"Bleeding Edge" – Datenmodellierung, Softwareentwicklung und die Freuden und Leiden forschungsgetriebener Entwicklung am Beispiel der Datenbank der Islamic Scientific Manuscript Initiative (ISMI)

,
casties@mpiwg-berlin.mpg.de
MPI für Wissenschaftsgeschichte, Deutschland

Wissenschaftliche Manuskripte des Islamischen Raums

Das ISMI Projekt (Islamic Scientific Manuscripts Initiative) ist ein gemeinsames Projekt des Max-Planck-Instituts für Wissenschaftsgeschichte und des Institute of Islamic Studies der McGill University in Montreal mit dem Ziel Informationen über alle islamischen Manuskripte in den "exakten Wissenschaften" (Astronomie, Mathematik, Optik, Mathematische Geographie, Musik, Mechanik und verwandte Disziplinen) in Arabisch, Persisch, Türkisch und anderen Sprachen aus der Zeit zwischen dem 8. und dem 19. Jahrhundert (CE) zu sammeln.

Im Projekt werden vor allem bibliographische und kodikologische Informationen zu den Manuskripten gesammelt, aber auch Informationen zu Nutzung und Besitz der Manuskripte und damit verbundenen Orten und Personen, die sich aus Kommentaren, Verkaufsbemerkungen und anderen Anmerkungen auf dem Manuskript ergeben können.

Arabische Manuskripte, aber auch andere alte Handschriften stellen besondere Anforderungen an die Datenmodellierung einer bibliographischen Datenbank. Informationen über Autor, Titel und Erstellungsdatum, die Standardfelder eines modernen bibliographischen Eintrags sind oft genug unbekannt. Dazu kommen viele unterschiedliche Schreibweisen des Namens einer Person und viele Fälle gleicher Namen, die zu unterschiedlichen Personen gehören. Es gibt viele Kopien des gleichen Textes von unterschiedlichen Kopisten und eine elaborierte Kultur von Kommentaren und Kommentaren höherer Ordnung.

Zusätzlich zu den bibliographischen Informationen werden auch Digitalisate präsentiert wenn sie vorhanden sind und die Lizenzbedingungen eine öffentliche Präsentation zulassen.

Von Tabellen zu Daten-Graphen

Die ursprüngliche ISMI Datenbank, erstellt von Prof. Jamil Ragep bestand aus einem Satz von relationalen Tabellen in MS-Access. In dieser Struktur gab es bereits die für die Arbeit mit Manuskripten wichtige Unterscheidung von abstraktem Text und der konkreten Manifestation in Form eines Manuskripts durch getrennte Tabellen für Texte und Manuskripte. Diese Unterscheidung erleichtert den Umgang mit vielen verschiedenen Kopien des gleichen Texts und erlaubt es den Blick auf die Beziehungen der Texte untereinander zu richten und beispielsweise durch zusätzliche Spalten Kommentarbeziehungen bis zu dritten Ordnung abzubilden.

Es schloss sich 2006 eine neue konzeptionelle Phase an, in der in Zusammenarbeit mit der IT-Gruppe des MPIWG ein neues Datenmodell entworfen wurde. Das neue Datenmodell orientiert sich an konzeptionellen Objekten, die es teilweise bereits im alten Datenbankschema gab: Texten, Manuskripten und Personen mit ihren jeweiligen Attributen und erweitert es durch weitere Objekttypen und individuelle Relations-Objekte mit frei definierbaren Typen und Attributen.

Das neue Modell sollte in der Software als möglichst flexible Attribut-Graph-Datenbank umgesetzt werden, so dass es jederzeit möglich sein sollte Objekten zusätzliche Attribute zu geben oder zusätzliche Relationen einzuführen, sobald dies während des Prozesses der Dateneingabe sinnvoll erscheint.

Im neuen Modell wurde beispielsweise die feste Relation des Author-ID Feldes der Text-Tabelle zur Autor-Tabelle durch ein Relationsobjekt des Typs "was_created_by" zwischen einem Text-Objekt und einem Personen-Objekt ersetzt.

Neue Möglichkeiten durch ein neues Datenmodell

Die mit dem neuen Datenmodell einhergehende Umstellung des Umgangs mit Daten von der vergleichsweise vertrauten Welt der Tabellen und tabellarischen Daten in eine Welt individueller Objekte und Relationen erweist sich bis heute einerseits als Chance, andererseits aber auch als Herausforderung in technischer und konzeptioneller Hinsicht.

Die Chancen zeigten sich in den Diskussionen über das Datenmodell und während der Dateneingabe. So entstanden neue Ideen wie das Konzept auch falsche Zuschreibungen (Misattribution) abzubilden: Wenn in der Literatur für lange Zeit eine Zuschreibung für die

Autorschaft eines bestimmten Textes existiert, die sich im Lauf der aktuellen Forschung als falsch herausstellt, dann ist es wichtig nicht nur die korrigierte Information aufzunehmen und anzuzeigen, sondern auch die als inkorrekt markierte alte Zuschreibung, um Forscher darauf hinzuweisen, dass die alte Zuschreibung bekannt ist und dass sie durch neue Informationen überholt ist.

Herausforderungen durch neue Technologie

Der Preis für das neue Datenmodell war neben der konzeptionellen Arbeit des Umdenkens und der Umstellung der bestehenden Daten zunächst vor allem die Abwesenheit von existierender Software zur effektiven Umsetzung der geplanten Datenstrukturen. Zum Zeitpunkt der Umstellung existierte keine verbreitete Graphendatenbanksoftware, so dass in mehreren Anläufen eine spezifische Datenbank „OpenMind“ und ein webbasiertes Frontend entwickelt wurde. Der Aufwand für die Wartung der Software und die Implementierung neuer Anforderungen steigt jedoch ständig.

Eine weitere Herausforderung, die sich nach der Eingabe grosser Datenmengen (derzeit 4000 Texte in 14000 Manuskripten in 7500 Codices und 2200 Personen) stellt, ist die Suche und Analyse der Daten. Vorhandene Werkzeuge, wie tabellenorientierte Abfragen sind auf den Graphen von Daten-Objekten nicht ohne weiteres anwendbar und es müssen Abfrage-Oberflächen entwickelt werden, die es ermöglichen das Potential der vernetzten Daten auch für die Forschenden nutzbar zu machen.

Neben der eher klassischen Webpräsentation und Browsing-Umgebung werden derzeit verschiedene Graphen-Visualisierungs- und Abfragetools getestet. In einem geplanten Workshop soll eine grösserer Kreis von Fachwissenschaftlern Zugang zur Datenbank und den experimentellen Werkzeugen erhalten um eine weitere Diskussion in Gang zu setzen und Erfahrungen und mögliche Fragestellungen für die weitere Entwicklung zu sammeln.

Bibliographie

Ragep, Jamil F. / Ragep, Sally P. (2008): „The Islamic Scientific Manuscript Initiative (ISMI). Towards a Sociology of the Exact Sciences in Islam“, in: Calvo, Emilia / Comes, Mercè / Puig, Roser / Rius, Monica (eds.): *A Shared Legacy: Islamic Science East and West*. Homage to Professor J. M. Millàs Vallicrosa. Barcelona: University of Barcelona 15–21 https://www.rasi.mcgill.ca/ISMI_SharedLegacy.pdf [letzter Zugriff 17. Februar 2016].