

WissKI – Wissenschaftliche Kommunikations- Infrastruktur

Merz, Dorian

dorian.merz@fau.de

Friedrich-Alexander Universität Erlangen-Nürnberg,
Deutschland

Fichtner, Mark

m.fichtner@wiss-ki.eu

Germanisches Nationalmuseum Nürnberg, Deutschland

Das DFG-finanzierte Projekt "WissKI" hat in den Jahren 2009 bis 2011 zur Entwicklung einer digitalen Forschungsumgebung für die Anwendung im Bereich der Digital Humanities geführt. Hauptaspekt der Datenerfassung und -haltung in WissKI sind die semantischen Zusammenhänge zwischen einzelnen Fakten und Datensätzen. Dies wird durch umfassende Unterstützung aktueller Semantic Web Technologien erreicht. Die Einordnung und Speicherung der erhobenen Daten erfolgt auf Grundlage einer Domänenontologie, deren Konzepte und Relationen - zu sogenannten Pfaden verbunden - als Vorlage für die Masken und Felder im System dienen. Auf Basis dieser Technologie werden solitär erscheinende Daten zu einem gemeinsamen, semantischen Netzwerk verbunden und damit die unmittelbare Sichtbarkeit weiterer, tiefergehender Zusammenhänge ermöglicht. Hierdurch bietet sich ein Mehrwert, der in der Vergangenheit in flachen Hierarchien wie Datenbanktabellen gar nicht oder nur mit sehr hohem Aufwand erkennbar gemacht werden konnte. Das Web-basierte Systemdesign und der dadurch ermöglichte Zugriff über das Internet, die Anbindung von externen kuratierten Datenquellen (sog. Authority Files) und die Möglichkeit zur Bereitstellung ausgewählter Daten über gängige Online-Schnittstellen (Web-Frontend, SPARQL-Endpoint, ...) betonen den Semantic-Web-Gedanken hinter der Infrastruktur.

Die Speicherung der Daten erfolgt in einem Triple-Store, der die eingegebenen Fakten in einer Subjekt-Prädikat-Objekt-Satzform ablegt. Die Aneinanderreihung der hier verwendeten Prädikate zu Pfaden erfolgt im Kern des Systems, dem sogenannten Pathbuilder, mit dem die semantische Bedeutung der einzelnen Masken-Einträge in Bezug auf das beschriebene Objekt (auch Person, Ort o. Ä.) anhand der Ontologie festgelegt wird. Die Eingabe der Daten erfolgt über eine, mit den gängigen Datenbankoberflächen vergleichbare, Editier-Oberfläche. Sie ist aus Feldern aufgebaut, die wiederum je einem bestimmten Feldtyp zugeordnet sind. Feldtypen bestimmen die Ein- und Ausgabemodalitäten der Daten.

So stehen zum Beispiel ein- oder mehrzeilige Textfelder, verschiedenartige Auswahldialoge und Möglichkeiten zur Bildanzeige zur Verfügung. In einigen davon kann WissKI den Anwender durch Auto-Vervollständigung unterstützen, indem zu angefangenen Eingaben ähnlich lautende Einträge aus der lokalen Datenhaltung oder den Authority Files vorgeschlagen werden.

Weitere umfassende Möglichkeiten zur Datenaufbereitung sind das halbautomatische Auszeichnen und Erkennen bzw. Einordnen von Personen, Orten, Zeiten und anderen Entitäten aus Texten (sog. Named Entity Recognition), die Verwaltung der Revisionsgeschichte von Texten, das Werkzeug zur Bildannotation und die Verknüpfung mit der Literaturverwaltungssoftware Zotero. Dabei verzichtet die Software nicht auf die aus dem Bereich der Content Management Systeme bekannten Funktionalitäten wie z. B. die Generierung von Websites, Foren, Wikis oder auch die detaillierte Verwaltung der Nutzer und ihrer Zugriffsrechte.

Inzwischen ist die Software in verschiedenen Forschungsprojekten an unterschiedlichen, namhaften Institutionen im kunst- und kulturhistorischen, sowie biologischen und technischen Bereich erfolgreich im Einsatz. Als Domänenontologie im Museums- und Sammlungsbetrieb kommen individuelle Erweiterungen des "Conceptual Reference Model" des Comité international pour la documentation zum Einsatz (CIDOC-CRM: ISO 21127), dessen Umsetzung in der Web-Ontology-Language OWL ebenfalls vom Projekt besorgt wurde und über die Website <http://erlangen-crm.org> frei zur Verfügung steht.

Das Poster stellt nun die konsequente Weiterführung des Projektes (DFG Proj.-Nr. GR1471/9) und das damit einhergehende Update der Software vor. Neben der grundlegenden Aktualisierung der zugrundeliegenden Frameworks und Technologien (Drupal 7, php 5.5, SPARQL 1.1) sind dabei einige Erweiterungen der Oberflächenfunktionalität sowie deutliche Erleichterungen in der Bedienbarkeit vorgesehen. So können nun beliebige Feldtypen aus der "Field API" des verwendeten Content Management Systems Drupal in die Ein- und Ausgabeansichten integriert werden. Dies umfasst neben den altbewährten Textfeldern und -bereichen und Bildern (incl. Zoomviewer für sehr hochauflösende Bilder) auch interaktive Landkarten, 3D-Animationen, Zeitstrahlen und alle denkbaren Medientypen, sowohl zur direkten Ansicht als auch zum Download. Zusätzlich zu diesen bereits eingebundenen Formaten ermöglicht die offene Architektur von WissKI² auch die Einbindung anderer, gängiger Feldtypmodule, die für Drupal zur Verfügung stehen.

Zu den erwähnten Erleichterungen zählt ebenso ein Update des System-Kerns, dem Pathbuilder, mit dem die Pfadschablonen durch die Domänenontologie auf einer graphischen Oberfläche ausgewählt bzw. erzeugt werden können. Daneben soll eine umfassende Bibliothek mit Musterontologien, -masken und -pfaden bereitgestellt werden, die die Einstiegshürde für Erstbenutzer minimal halten wird.

Ein weiteres neu integriertes Werkzeug zur Verbesserung der Datenqualität werden automatische Inferenzmaschinen, sog. Reasoner, sein, die in der Lage sind, widersprüchliche Daten zu erkennen und somit etwaige Eingabe- oder Modellierungsfehler aufzudecken.