

Projektdokumentation im Modul Semantic Web

Entwicklung eines Onlinelexikon der französischen Landschaftsmaler des 19. Jhd.

Philipp Knorr

31. Juli 2016

Aufgabenstellung: Das Bildermuseum in Leipzig hat von einem Autor einen Entwurf eines Buches über französischen Landschaftsmaler des 19. Jahrhundert erhalten. Da diese Daten in dieser Form nicht leicht zu veröffentlichen bzw. zu nutzen sind, hat es sich an Professor Riechert und Dr. Bergmeyer mit der Anfrage gewandt, ob diese Daten nicht im Rahmen Open-Linked-Data veröffentlicht werden könnten.

1 Inhaltliche Interpretation der Fragestellung

1.1 Aufgabenstellung und Gegebenheiten

Die Daten des Bildermuseum sollen als Open-Linked-Data veröffentlicht werden. Außerdem soll ein Interface zur Verfügung stehen, das die Eingabe der vorhanden Daten ermöglicht. Um die Daten abzufragen, müssen diese in einer Wissensbasis oder Datenbank gespeichert werden. Die Daten, die in Museen vorkommen, sind nicht homogen und daher eignet sich kein bekanntes Schema für das Ablegen der Daten, des Weiteren sind hier oft Objekte Unikate. Erschwerend ist, dass eine strukturierte Bestandsaufnahme lange Zeit vernachlässigt worden ist. Deshalb wird ein Schema benötigt, welches das Ablegen von diesen spezifischen Datensätzen ermöglicht. Hierfür wird ein geeignetes Formular benötigt, das auf das zu entwickelnde Schema angepasst ist und den Anforderungen des Museums entspricht.

1.2 Interpretation der Aufgabe

Die Aufgaben mussten zunächst in eine sinnvolle Reihenfolge gebracht und gegebenenfalls in kleinere Teilaufgaben unterteilt werden. Die Unterteilung sowie die Einordnung und Neuformulierung von Aufgaben musste auch in Absprache mit den Museologen stattfinden, um auch immer die Nutzung im späteren Betrieb zu gewährleisten.

Als Erstes musste ein System gefunden werden, welches die Repräsentierung und Speicherung von Wissen ermöglicht. Infolgedessen wurde das Schema entwickelt, das die Speicherung des Wissens in der Wissensbasis ermöglicht. Zum Abschluss wurde ein Formular entwickelt, welches das Einlesen und Bearbeiten der Daten in der Wissensbasis gewährleistet und erleichtert.

2 Ontowiki

Nach einer Recherche von Wissensrepräsentationssystemen hat sich Ontowiki als vorteilhaft herausgestellt. Dieses System wurde von der AKSW research group der Universität Leipzig entwickelt. Es handelt sich hierbei um eine OpenSource Software mit GPL Lizenz. Das System ist eine Webanwendung, welche in PHP entwickelt wurden und mit Hilfe von einer MYSQL Datenbank oder eines Virtuoso triple store die Daten abspeichert.

Das Ontowiki besteht im Wesentlichen aus zwei Kernkomponenten, dem Ontologien Editor (ontology editor) und dem Wissensmanagement System (knowledge acquisition system). Der Ontologien Editor ermöglicht es auch angepasste, selbst entwickelte Ontologien zu verwalten und einzusetzen. Das Wissensmanagement System erleichtert die Verwaltung des vorhandenen Wissens und den Zugriff sowie die Verlinkung. Der SPARQL Endpunkt kann zur Abfrage von Informationen sowie zur Entwicklung eines Webinterface genutzt werden.

Anders als bei den meisten Systemen ist Ontowiki formularbasierend statt syntaxbasierend, was das Einfügen von einem angepassten Formular zum Einlesen von Museumsdaten ermöglicht. Außerdem ist diesem System der Fokus auf vereinfachter Wissensrepräsentierung gelegt worden, was bei dem Einsatz im Museum von Vorteil ist, denn so kann das System leicht erlernt werden, da die meisten Benutzer Laien sind. Auch wird bei Ontowiki dem Nutzer die komplexe Wissensdarstellung mit deren Formalismen erspart und im Hintergrund verborgen. Ontowiki wurde für die erste Phase als Testsystem eingerichtet, dabei war ein weiterer Vorteil, dass es auch auf einem Docker leicht eingerichtet werden kann.

3 Schemata

Zur Speicherung der Daten wurde ein angepasstes Schema benötigt. Zunächst wurde hierfür die vorhandene Datenstruktur der vorliegenden Daten analysiert. Es handelt sich hier um 3.200 Einträge in unterschiedlicher Länge mit 4.000.000 Zeichen inklusive Leerzeichen und zudem werden diese aktuell noch lektoriert. Dabei hat sich leider vor allem herausgestellt, dass diese Daten sehr heterogen und auch teils unvollständig sind. Dazu besteht eine weitere Schwierigkeit, da diese Daten auch in Normdaten umgewandelt werden sollten, sodass diese besser verlinkt werden können. Das Format soll auch die Publikation der Daten als LOD (Linked Open Data) ermöglichen.

Um einen besseren Überblick über Formate und Strukturen von Daten in Museen zu bekommen und Methoden kennenzulernen, habe ich eine Recherche im Oberseminar zu diesem Fach durchgeführt und so Erkenntnisse einfließen lassen. Die dadurch kennengelernten Methoden, die solche Daten in ein RDF-Format überführen, haben das Entwickeln eines Schemas beschleunigt und erleichtert. Betrachtet wurde die Überführung von Museumsdaten in Schweden (Verweis Artikel1) und Amerika (Verweis Artikel2). Dabei hat sich von Vorteil erwiesen, dass das eingesetzte System Ontowiki formularbasierend ist, denn so kann ein Formular auf dem entwickelten und angepassten Schema die Daten in die Wissensbasis überführen. Allerdings muss hier noch eine Vorverarbeitung der Daten stattfinden, denn die gegebenen Datensätzen enthalten teils große Blöcke. Diese beinhalten viele Informationen, die extrahiert werden sollen, um diese auch als Normdaten sowie als Open-Linked-Data verfügbar zu machen.

3.1 Klassen

Das Schema baut sich um die im Zentrum stehenden Person auf. Die Klassen bilden alle vorhandenen Informationen der gegebenen Daten ab.

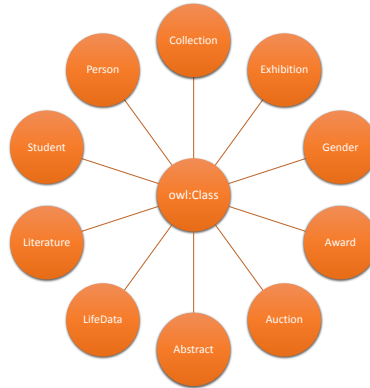


Abbildung 1: Klassen des Schemas

3.2 Strukturbeispiel

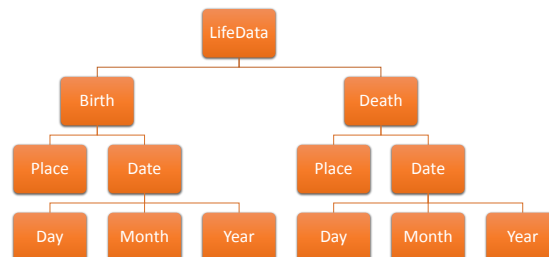


Abbildung 2: Klassen des Schemas

Das erste Beispiel (Abbildung 2) bildet alle Informationen über die Lebensdaten der Person ab. Außerdem ist diese Klasse mit einer mehrstufigen Hierarchie ausgestattet, sodass die einzelnen Daten individuell gespeichert werden können. Das ermöglicht auch unvollständige Datensätze zu speichern. Im Beispiel ist so Birth ein Object Property von LifeData. Die Instanz Birth kann dann auch Place als Datatype Property und Date als Object Property besitzen. Das Date Objekt kann aus Day, Month und/oder Year bestehen. Diese spezielle Unterscheidung ist zum Beispiel nötig, damit auch ungenaue

Jahresdaten gespeichert werden können, wenn in diesem Beispiel das genaue Geburtsdatum unbekannt oder nur geschätzt werden kann.

Im zweiten Beispiel (Abbildung 3) sind die Daten als Datatype Property der Instanz der Klasse Name, zugeordnet.

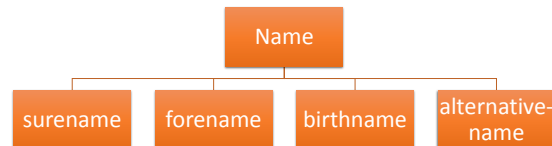


Abbildung 3: Klassen des Schemas

3.3 Hierarchiebeispiel

Die jeweilige Instanz wird der Person als Object Property oder Datatype Property zugeordnet. Alle Klassen, welche die Person direkt betreffen, sind der Person untergeordnet. Das sind Klassen, wie die in den zuvor aufgeführten Beispielen. Die übrigen Klassen wie Award oder Collection sind nicht der Klasse Person unterstellt, aber werden mit der jeweiligen Instanz von Person verlinkt. Durch diese Verlinkung wird sichergestellt, dass es sich hierbei um Normdaten handelt.

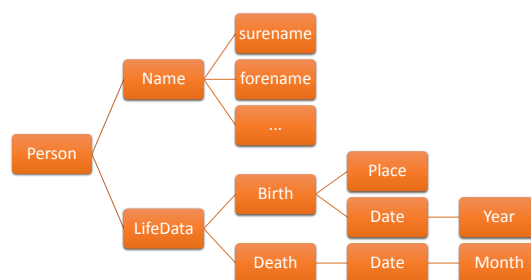


Abbildung 4: Klassen des Schemas

3.4 Zugriff des Schemas

Der Zugriff auf die Daten kann über den in Ontowiki integrierten SPARQL Endpunkt gewährleistet werden. Außerdem kann so ein Webinterface entwickelt werden, welches den Zugriff und die Darstellung der Daten vereinfacht. Das Beispiel beschreibt eine Abfrage aller Personen und deren zugehöriger Lebensdaten mit Hilfe von SPARQL.

Beispiel Query:

```
PREFIX flms: <http://cours.imn.htwk-leipzig.de/ont/flms/>
SELECT ?person ?object
WHERE {
    ?person flms:LifeData ?object .
}
```

Zugriff des Schemas - Ergebnis der Query

```
person:
http://cours.imn.htwk-leipzig.de/ont/flmd/Person/Karl
object:
http://cours.imn.htwk-leipzig.de/ont/flmd/Lebensdaten/Lebensdaten_von_Karl_dem_Grossen
```

4 Zusammenfassung und Ergebnisse

Der Aufbau eines Testsystems wurde erfolgreich abgeschlossen. Dafür wird Ontowiki als System zur Wissensrepräsentierung und Abspeicherung der Daten genutzt. Die Überführung der bestehenden Daten in ein RDF-Format wird mit Hilfe des angepassten Schemas gewährleistet. Das entwickelte Schema erfüllt auch die Vorgabe, die Daten als Normdaten zur Verlinkung zu nutzen.

Weiterhin soll ein Formular angepasst und implementiert werden, welches die Wisenseingaben ermöglicht. Als Vorlage für das Formular dient das RForm (<https://github.com/simeonackermann/RForm>).

Eine weitere offene Aufgabe ist die Entwicklung eines Webinterface zur Wissensrepräsentierung. Das System Ontowiki bietet zwar bereits ein Webinterface, aber eine Vorgabe ist die Entwicklung einer Webseite, welche die Daten übersichtlicher präsentiert.