Aim of the Project

The aim of the project is to create an input system to feed the Literature Georeferencing Database, so the software is able to generate Literature Map out of the data.

A Literature Map Software must be able to:
1. create a geographical map given an author showing georeferences in his life and in his works;
2. create a geographical map given a work showing georeferences in this work;
3. create a geographical map given a geo-location showing other georeferences linked to this location through literary works [3]

Workflow

Rational behind the data modelling:
1. The researchers should be able to enter large amount of data quickly.
2. We apply DRY-Principle (Don’t Repeat Yourself) as good as possible.
3. Maps are generated from the database, and not drawn by the scholars using GIS tools.

Thus, we come up with this solution: For the metadatas, a web frontend for database operations is created, for the TEI data we found a solution based upon DOCX to TEI Conversion. Furthermore, queries on the authorities given (GND, Geonames.org, and WikiData) is made during the data ingest, thus the database entries are linked to other authorities. However, the same data are done with open government data made available by the local government of Tyrol and South Tyrol: the reason these data can only be downloaded but not queried through a web service. Therefore, the data is stored locally and queried during the data ingest.

Benefits of this modelling:
1. Captures many aspects of use cases (e.g. queries on objects are just queries on databases.)
2. Database operation on metadata of objects can be done quickly.
3. Inserting and Updated tasks can be done easily.
4. Data models reflect their usage in humanities research.

Drawbacks of this modelling:
1. Complicated database scheme is difficult to maintain.
2. Redundant data storage, no back-propagation of changes in RDBMS to TEI Document.
3. Dependencies on data of other relations.

Map Generation:
Places have coordinates, they can be plotted in a HTML map. Using LeafJS, one can embed interactive maps within a web page. Different colors and different icons mark different meaning of a place.

Having the relationships between authors, works and geographical places in focus, the Project The Tyrol / The South Tyrol - A literary Topography (FWF P26039) created a database on these entities.

At the beginning, data modelling only captures basic metadata of the entities and their relationships. However, it was impossible to annotate texts, thus scholars could not tag the actual section of the texts with additional references.

After careful consideration, it was decided to redo the data modeling: While the metadata could still be kept and maintained by a RDBMS (in our case: MySQL), the actual texts (primary sources, biographies and description of places) should be modelled as XML/TEI data and kept in a repository (in our case: Apache Jackrabbit). Furthermore, we want to reuse authority controlled data and open government data.

Starting Point

From DOCX via TEI to Literature Map
A presentation at the TEI Conference and Members’ Meeting 2016 in Vienna, September 2016, by Joseph Wang, University of Innsbruck

Convert comments marked using milestones to <rs> in 3 steps
A complete example without markup.

Step 1: convert all elements to milestones and texts.
Step 2: Convert the relevant milestones to <rs> and copy the whole text. Furthermore, we want to reuse authority controlled data and open government data.

Step 3: convert the remaining milestones back to the original milestones.

Data Modelling

References