BauDenkMalNetz – Creating a Semantically Annotated Web Resource for Historical Buildings

Anca Dumitrache

Supervised by Prof. Dr. Michael Kohlhase
Computer Science
Jacobs University Bremen

May 13, 2011

Introduction

- **BauDenkMalNetz** $\rightarrow$ "listed buildings network"
- a web portal for discovering Bremen's architectural landscape that is suited for the tech-savvy tourist
- based on a series of guide books by Dr. Nils Aschenbeck
- 4 books, more than 100 buildings

Why Semantic Web

Definition

An **ontology** is a data conceptualization, formalized in some logic by defining classes of concepts, together with the relations between them.

Why Semantic Web

Definition

An **ontology** is a data conceptualization, formalized in some logic by defining classes of concepts, together with the relations between them.

Definition

An **RDF triple** is a structure comprised of three parts: subject, predicate and object, each of them represented either by a unique URI, or a standard XML data type, defined as a literal.

Why Semantic Web

Definition

An **ontology** is a data conceptualization, formalized in some logic by defining classes of concepts, together with the relations between them.

Definition

An **RDF triple** is a structure comprised of three parts: subject, predicate and object, each of them represented either by a unique URI, or a standard XML data type, defined as a literal.

Definition

An **n-gram model** refers to a probabilistic model that, given the first $n-1$ words in a sentence, will predict the n^{th} word.

n-gram Analysis

Groups of words most likely to appear together:

- physical buildings

n-gram Analysis

Groups of words most likely to appear together:

- physical buildings
- personal names

n-gram Analysis

Groups of words most likely to appear together:

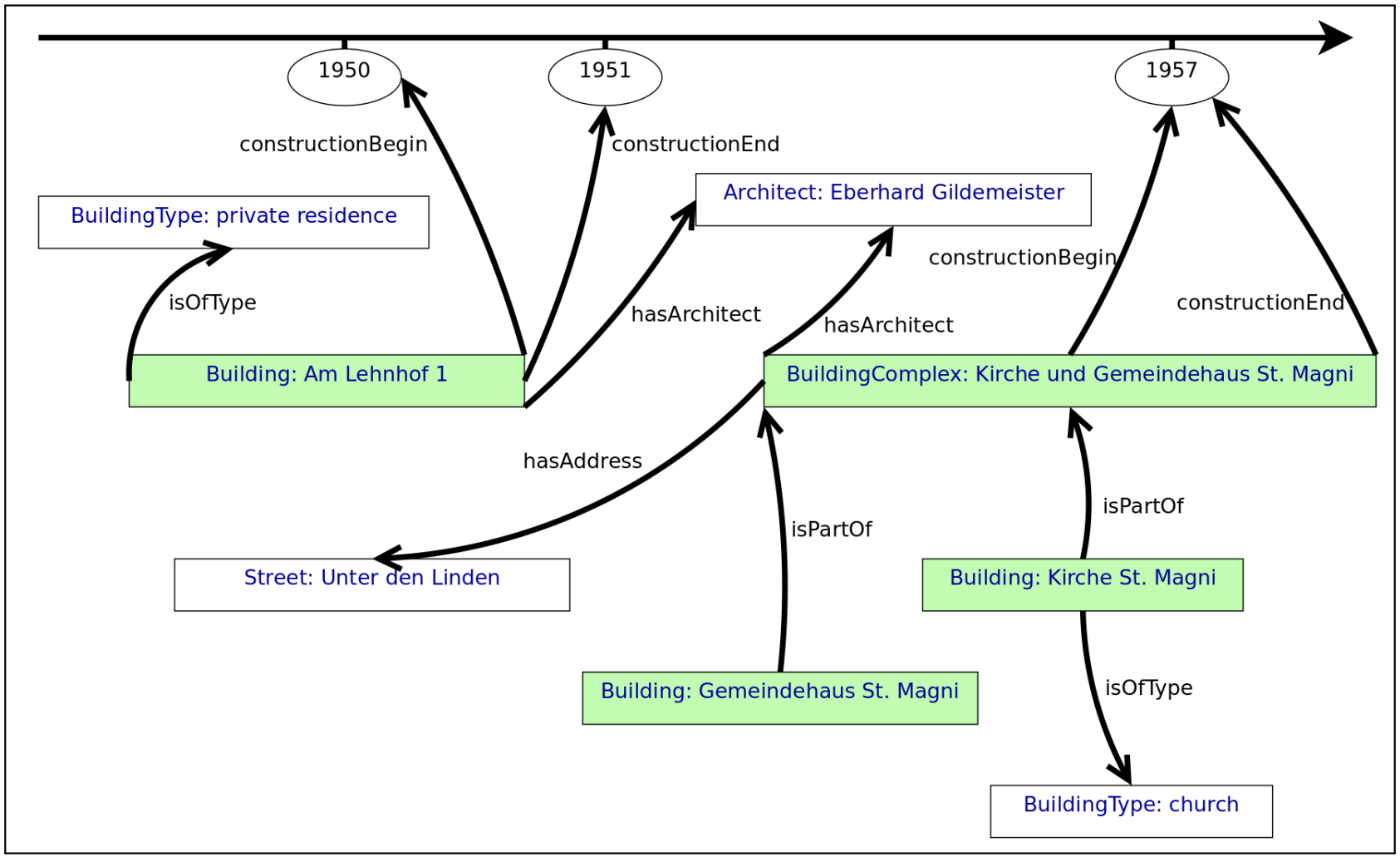
- physical buildings
- personal names
- human-readable addresses

n-gram Analysis

Groups of words most likely to appear together:

- physical buildings
- personal names
- human-readable addresses
- building features

The BDMN Ontology



Choosing a Deployment Environment

- **Semantic content management systems:** Semantic MediaWiki, Drupal 7.

Choosing a Deployment Environment

- **Semantic content management systems:** Semantic MediaWiki, Drupal 7.
- Requirements:
  1. possibility of integrating ontologies;
  2. browsing based on the semantic metadata;
  3. support for querying the RDF;
  4. possibility of importing large amounts of text.

Drupal 7 Modules

- **BDMN module** – custom module for creating content types;

Drupal 7 Modules

- **BDMN module** – custom module for creating content types;
- **Evoc** – for importing ontologies;

Drupal 7 Modules

- **BDMN module** – custom module for creating content types;
- **Evoc** – for importing ontologies;
- **SPARQL Views** – for queries over RDF;

Definition

SPARQL is the standard language for querying data stored in RDF triple format.

Drupal 7 Modules

- **BDMN module** – custom module for creating content types;
- **Evoc** – for importing ontologies;
- **SPARQL Views** – for queries over RDF;

Definition

SPARQL is the standard language for querying data stored in RDF triple format.

- **GMap** – for Google Maps visualization.

Evaluation

- Evaluation criteria: **usability** and **usefulness**.
- Task-based evaluation – users were presented with an assignment that they had to solve.
- Results:
  - 100% success in solving the task
  - user suggestions: integration with social networking platforms, enhanced map visualization...

Future Work

Definition

XSPARQL: XML results to queries over RDF, RDF data from XML documents;
KML: XML notation for representing geolocation, used by Google Earth.

Example

Use XSPARQL queries over RDF to return KML data that can be visualized in Google Earth.

Conclusion

- **Aim:** to develop a web portal that publishes printed text about Bremen's architecture.
- **How it was accomplished:** creating an ontology, creating a Drupal 7 website with pages mapped to the ontology, enhancing its functionality with semantic queries.
- **Future work:** semantic geolocation, reusing the ontology / reapplying the methodology for similar web resources...

Thank you for your attention!