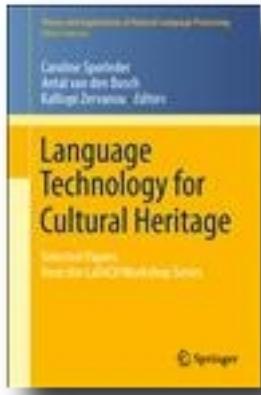# New Book Chapter on Semantic Wikis and Natural Language Processing for Cultural Heritage Data

- [Cultural Heritage Data](#)
- [Durm](#)
- [Semantic Wiki](#)
- [NLP](#)

Springer just published a new book, *Language Technology for Cultural Heritage*, where we also contributed a chapter: "Integrating Wiki Systems, Natural Language Processing, and Semantic Technologies for Cultural Heritage Data Management". The book collects selected, extended papers from several years of the LaTeCH workshop series, where we presented our work on the Durm Project back in 2008.

In this project, which ran from 2004–2006, we analysed the historic *Encyclopedia of Architecture*, which was written in German between 1880-1943. It was one of the largest projects aiming at conserving all architectural knowledge available at that time. Today, its vast amount of content is mostly lost: few complete sets are available and its complex structure does not lend itself easily to contemporary application. We were able to track down one of the rare complete sets in the Karlsruhe University's library, where it fills *several meters* of shelves in the archives. The goal, then, was to apply "modern" (as of 2005) semantic technologies to make these heritage documents accessible again by transforming them into a semantic knowledge base (due to funding limitations, we only worked with one book in this project, but the system was developed to be able to eventually cover the complete set). Using techniques from Natural Language Processing and Semantic Computing, we automatically populate an ontology that can be used for various application scenarios: Building historians can use it to navigate and query the encyclopedia, while architects can directly integrate it into contemporary construction tools. Additionally, we made all content accessible through a user-friendly Wiki interface, which combines original text with NLP-derived metadata and adds annotation capabilities for collaborative use (note that not all features are enabled in the public demo version).

All data created in the project (scanned book images, generated corpora, etc.) is publicly available under open content licenses. We also still maintain a number of open source tools that were originally developed for this project, such as the Durm German Lemmatizer. A new version of our Wiki/NLP integration, which will allow everyone to easily set up a similar system, is currently under development and will be available early 2012.

Semantics for the Masses

Source URL (retrieved on *2026-01-16 00:58*):
https://www.semanticsoftware.info/content/new-book-chapter-semantic-wikis-and-natural-language-processing-cultural-heritage-data