

A Storage Solution for the WOM with Revision Control

Summary

Most wikis store their pages (articles) as text blobs in an SQL database. We have developed a formal parser for MediaWiki's markup format, a format that drives Wikipedia and many other wiki instances. This parser produces a structured document (WOM) which unlocks the information stored in each article and enables us to bring refactoring and other technologies to the wiki world. The goal of this thesis is to implement a storage solution for WOM documents that supports quick, arbitrary searches over all articles and facilitates revision control.

Work Results

- Selection of base technology
 - It is not expected to invent an entirely new database technology (although this is an option).
 - The goal should be achieved by combining existing technologies (e.g. XML, SQL, No-SQL databases, text search engine) into a working storage solution.
 - The query interface should be based on existing query languages, as needed (e.g. XPath, XQuery, SPARQL, SQL).
- Design and implementation of an appropriate storage layer and interface
 - Design and implementation of a **storage layer** that facilitates fast retrieval of content, quick searches over the whole wiki, enables revision control and compressed storage.
 - Design and implementation of a well-defined **storage layer interface** for storing and querying articles and article meta data (the interface should be specific to the task it fulfils for wikis and not be of a generic nature).
- Discussion and evaluation of the performance and limitations of the
 - resulting software architecture,
 - implemented solution.

Supervisor

Dipl.-Inf. Hannes Dohrn, hannes.dohrn@fau.de

Prof. Dr. Dirk Riehle, dirk.riehle@fau.de

Open Source Research Group, Computer Science Department, Friedrich-Alexander University

Link to resources: <http://osr.cs.fau.de/theses/resources/>