Chapter IV
SWiM:
A Semantic Wiki for Mathematical Knowledge Management

Christoph Lange
Jacobs University, Germany

Michael Kohlhase
Jacobs University, Germany

ABSTRACT

In this chapter, we present the SWiM system, a prototype semantic wiki for collaboratively building, editing, and browsing mathematical knowledge. SWiM is based on the semantic wiki IkeWiki, but replaces the wiki text with OMDoc, a markup format and ontology language for mathematical documents as the underlying knowledge representation format. Our long-term objective is to evolve SWiM into an integrated platform for ontology-based added-value services. As a social semantic work environment, it will facilitate the creation of a shared, public collection of mathematical knowledge (e.g., for education) and serve scientists as a tool for collaborative development of new theories. We discuss the architecture of the SWiM system focusing on its conceptual base, the OMDoc system ontology. In contrast to other semantic wikis, SWiM uses the system ontology to operationalize the fragments and relations of the underlying representation format, not only the domain ontology, that is, the relations between the represented objects themselves. We will present the prototype implementation of the SWiM system and propose its further evolution into a service platform for science and technology.

INTRODUCTION

The Internet plays an ever-increasing role in our everyday life, and science is no exception. The way we do (conceive, develop, communicate about, and publish) scientific or mathematical knowledge will change considerably in the next 10 years. In particular, most of the activities will be supported by *scientific services*, that is, software systems connected by a commonly accepted distribution architecture. It is a crucial but obvious insight that true cooperation of such services is only feasible if they have access to a joint corpus of joint knowledge. A central prerequisite for this is a technology that is capable to create, maintain, and deploy *content-oriented* (a.k.a. "semantic") libraries of science on the Web that make the structure of scientific knowledge explicit, so that it can serve as a context for the objects of science.

As mathematics is a discipline, which has always been especially aware of the syntax/semantics distinction and has treated its theories as first-class objects, we will pursue this ambitious goal for mathematical knowledge first and extend the methods and technologies to the (natural) sciences and engineering later.

If we extend Tim Berners-Lee's vision of a "Semantic Web" as a *Web* of *data* for *applications* to mathematics, many services come to mind: cutting and pasting mathematical text from a search engine into a computer algebra system, automated checking or explanation of proofs published on the Web, if they are sufficiently detailed and structured, semantic search or data mining for mathematical concepts ("Are there any objects with the group property out there?"), classification (given a concrete mathematical structure, is there a general theory for it?), and many more. All of these services can currently only be performed by humans, limiting the accessibility and thus the potential value of the information.

On the other hand, the content-oriented mathematical libraries can only be generated by humans, as has been proved by the successful *PlanetMath* project (Krowne, 2003; "*PlanetMath*," 2007), which features free, collaboratively created entries on more than 10,000 mathematical concepts. *PlanetMath*, however, is not completely machine-understandable. There is a fixed set of metadata associated with each article, including its type (definition, theorem, etc.), parent topic, Mathematics Subject Classification, synonyms, and keywords, but the content itself is written in LaTeX and can only be searched in full-text mode.

SEMANTIC MATHEMATICAL KNOWLEDGE MARKUP WITH OMDoc

We make use of the structural/semantic markup approaches using formats such as OpenMath (Buswell, Caprotti, Carlisle, Dewar, Gaetano, & Kohlhase, 2004), Content MathML (Carlisle, Ion, Miner, & Poppelier, 2003), and OMDoc (Open Mathematical Documents (Kohlhase, 2006)), the latter of which embeds and extends the former ones. These formats, constituting the state of the art for representing mathematical knowledge, are now used in a large set of projects in automated theorem proving (Müller, 2006a), e-learning (Melis et al., 2006), e-publishing, semantic search (Kohlhase & Şucan, 2006), and in formal digital libraries. OMDoc builds on OpenMath or Content MathML for mathematical formulae and extends this by an infrastructure for context and domain models from "formal methods." In contrast to those, a structural/semantic approach does not require the full formalization of mathematical knowledge, but only the explicit markup of important structural properties. For instance, a statement will already be considered as "true" if there is a proof object that has certain structural properties, not only if there is a formally verifiable proof for it. This allows OMDoc to be used, for example, as a development/migration format,