

# Towards an Ontology-Driven Management of Change

Normen Müller and Michael Kohlhase

Computer Science, International University Bremen  
n.mueller@iu-bremen.de m.kohlhase@iu-bremen.de

Current document management systems (DMS) are designed to coordinate the collaborative creation and maintenance process of documents through the provision of a centralized repository. The focus is primarily on the management documents themselves. Relations between and within documents and effects of changes are largely neglected. To avoid inefficiencies, conflicts, and delays the support of modification management is indispensable. Up to now humans are needed for such a *management of change* (MOC), i.e. to maintain consistency after any modifications.

Some initial research has been conducted on methods and tools managing the consistency and change of documents: For formal documents like programs or specifications we refer to the HETS [Het] and MAYA [May] systems. For informal documents like mathematical textbooks we refer to the MMiSS [MMi] project. However, all these systems ([MS05,Mos05]) base their MOC on the inherent underlying (formal) mathematical structure of the documents. Thus they are restricted to specific problem domains, where we have mathematical formalizations.

Here we present the design of the LOCUTOR system that aims to provide MOC for arbitrary XML documents ranging from *informal*, e.g. instruction or construction manuals, to *formal* documents. In order to be also able to handle (purely) formal documents we will use the insights of the above mentioned systems.

*A Structured View of Documents* We use a structured view of documents to facilitate management of change, information reuse, and consistency. In contrast to file- and line-based systems like the SUBVERSION system [SVN], we consider documents as structured collections of information units. For the purpose of this article one can think of information units as “*tangible/visual text fragments potentially adequate for reuse*”. To identify information units as well as to define non-structural relations between information units we base our approach on knowledge representation methods, in particular on the notion of a *System’s Ontology* [KBM06]. This is an ontology describing the data models behind the representation formats independently of their respective syntactical realization. Thus we are not bound to any specific application area or specific document format but yet able to capture semantic interrelations, e.g. *illustrates*, *refines* or *depends-on*, even between (fragments of) informal documents.

To distinguish the ontological concept from colloquial usage of the “information unit”, we will use the word *infom* for the former. In this regard we consider any (user-defined) concept “is\_a”-related to the concept *infom* to be an informa-

tion unit. We call non-structural relations *ontological* relations as distinct from structural relations.

Following [VD04] and [Koh06] we separate documents into two layers: A *narrative* and a *content* layer both of which consist of *infoms* and are structure by *ontological* relations. The order of information units in documents is represented on the narrative layer. The information units themselves are placed in the content layer and referenced by *narrative links* (analog to symbolic links in UNIX). Ontological relations are represented by *ontological links* in the content layer. The information units and the ontological links build up the “content commons” [CNX05].

Following the MMiSS [MMi] project, our approach is also aware of the concept of *variants*. This expands the application area not only “in-the-breadth” but also “in-the-depth”. Thus, by extending the well-known concept of *versions* and *revisions* by the concept of variants, the life-cycle of documents will no longer be only along a horizontal time line but also along a vertical line of variants. On the document level we call the concept of versions, revisions, and variants *document states*.

Thus the two layered, two-dimensional view of documents represented by directed acyclic (narrative layer) combined with directed cyclic (content layer) graphs (NARCON) already facilitate *information harvesting*. In the next paragraph we will give a survey, how NARCONS support management of change for (informal) documents. We develop a MOC on NARCONS to maintain consistency during the development of various document states to additionally provide reliable *information reuse*.

*Management of Change* In our approach documents have to explicitly identify the underlying language  $\mathcal{L} := \langle \mathcal{G}, \mathcal{O} \rangle$  they are an instance of. So we regard a language definition to be a tuple consisting of a grammar  $\mathcal{G}$  and an ontology  $\mathcal{O}$ . A grammar  $\mathcal{G}$  constitutes the syntactical rules to build up a document, e.g. an XML Schema, whereas an ontology  $\mathcal{O}$  enriches a document by further semantic interrelations (the system’s ontology described above).

We base our computation of structural differences on the insights of XML-diff tools [Har,XML] and adapt the algorithms to operate on NARCONS. Based on a model  $\mathcal{M}$  for NARCONS, i.e. to define when two NARCONS are considered to be equal, we develop a MDiff algorithm:  $\text{MDiff} : \mathcal{D} \rightarrow \mathcal{D} \rightarrow \mathcal{G} \rightarrow \Delta$ , where  $\mathcal{D}$  denotes a NARCON,  $\mathcal{G}$  a grammar, and  $\Delta$  a set of structural differences between NARCONS (we achieved an XML-based SUBVERSION so far).

We can extend  $\Delta$  to also capture the *structural semantic closure*<sup>1</sup> (SSC), denoted by  $\delta^*$ , of each structural difference  $\delta$ . Analog to ontologies enriching documents, we develop a set  $\mathcal{CR}$  of *change-relations* to enable authors to annotate computed structural difference  $\delta \in \Delta$ . We denote the set of annotated

---

<sup>1</sup> We use the term “structural semantics” in sense of marking up the meaning by structure. We do not need any entailment relation to model but rather concentrate just on structural and ontological relations.

structural differences by  $\ddot{\Delta}$ . So we extend the *two-valued states* of changes, i.e. modified and non-modified, to *annotated two-valued states* of changes.

To systematically reason on annotated changes, say to reason on  $\ddot{\Delta}$ , we on the one hand type-cast sorts of changes by elaborating a taxonomy of  $\mathcal{CR}$ . On the other hand we define *equivalence classes* on  $\mathcal{CR}$  to operate on representatives rather than on singletons. We will develop inference rules consolidated in a  $\mathcal{CR}$ -Calculus operating on NARCONS to deduce the effect of changes. With these “rules of re-action to changes” we are able to define an algorithm to compute the SSC of each  $\delta \in \ddot{\Delta}$ :  $\mathcal{ODiff} : \ddot{\Delta} \rightarrow \ddot{\Delta}^*$ .

We want to lead authors to annotate informal documents step by step, i.e. to provide informal documents more and more with structural semantics. As a consequence of each single step the “informal wave” shifts a bit more towards the formal world and thus can be kept better under control by formal systems:

*“The flatter a document the less the assistance!”*

## References

- [CNX05] CONNEXIONS. Project home page at <http://cnx.rice.edu/>, seen January 2005.
- [Har] The Harmony Project. Web site at <http://www.seas.upenn.edu/~harmony/>.
- [Het] Hets: The Heterogeneous Tool Set. Web site at [http://www.informatik.uni-bremen.de/agbkb/forschung/formal\\_methods/CoFI/hets/](http://www.informatik.uni-bremen.de/agbkb/forschung/formal_methods/CoFI/hets/).
- [KBM06] Bern Krieg-Brückner and Achim Mahnke. Semantic Interrelation and Change Management. In *OMDoc an open markup format for mathematical documents (version 1.2)*, chapter 26.6, pages 274–277. Springer LNAI, 2006.
- [Koh06] Michael Kohlhase. OMDoc an open markup format for mathematical documents (version 1.2). Manuscript, <http://www.mathweb.org/omdoc/omdoc1.2.ps> to appear in Springer LNAI, 2006.
- [May] MAYA: The Development Graph Manager. Web site at <http://www.dfki.de/~inka/maya.html>.
- [MMi] MMiSS: Multimedia in Safe and Secure Systems. Web site at [www.mmiss.de](http://www.mmiss.de).
- [Mos05] Till Mossakowski. *Heterogeneous Specification and the Heterogeneous Tool Set*. Habilitation, Universität Bremen, 2005.
- [MS05] Achim Mahnke and Jan Scheffczyk. Engineering Mathematical Knowledge. In Michael Kohlhase, editor, *Mathematical Knowledge Management*, number 3863 in LNAI. Springer Verlag, 2005.
- [SVN] The Subversion Project. Web site at <http://subversion.tigris.org/>.
- [VD04] Katrien Verbert and Erik Duval. Towards a Global Component Architecture for Learning Objects: A Comparative Analysis of Learning Object Content Models. In *Proceedings of the EDMEDIA 2004 World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 202–208, 2004.
- [XML] The XMLDiff Project. Web site at <https://xmldiff.dev.java.net/>.