WorldML vs. YaML – On the scope and purpose of mark-up languages

Jens Klump (1) and Robert Huber (2)

(1) German Research Centre for Geosciences, CeGIT, Potsdam, Germany (jens.klump@gfz-potsdam.de, +49 (331) 2881703),
(2) MARUM, Univ. Bremen, Bremen, Germany (rhuber@pangaea.de, +49 (421) 218-65505

With the rise of web services and XML came the idea that it should be possible to assemble components of a service oriented architecture in a modular fashion, similar to "plug and play" hardware. Web services were thought to integrate into web based architectures by automatic inclusion in a registry of available services with standardised, machine interpretable descriptions of their interfaces. Information exchange was thought to be between machines only with no need for human intervention. Mark-up languages were seen as well suited media for machine-machine communication since they can be validated by machines and leave no room for interpretation or ambiguity. But does this assumption hold true? Is it possible to pre-define an unambiguous structure for this communication – a perfect language? Or is the solution "yet another mark-up language"?

Since the advent of XML mark-up languages there has been an antagonism between top-down standardisation efforts and bottom-up definition by communities of practice. The same is about to happen in the field of semantic web technologies. In practice we can observe a widespread use of simple mark-up languages (e.g. Dublin Core), while more complex mark-up schemes (e.g. ABCDEFG) were never adopted by their designated user communities. Further complexities arise, if the XML schema allows application profiles deviating from the standard schema. Simple, purpose driven ontologies and "folksonomies" thrive, while top-down high level ontologies are still awaiting practical application.

In practice, it turns out that in many instances communication between the parties operating the respective web services is required to allow machine-machine communication. This is not a mere practical difficulty but is caused by an underlying property of communication that has been the subject of research and debate for centuries.

There are principal considerations that can help us to decide where to direct our efforts in the choice or design of mark-up languages and ontologies. For centuries humans have dreamt of the perfect language that allows unambiguous communication. Philosophers have postulated that such a language we could even allow us to chart our knowledge and define its boundaries. Philosophical work in the early 20th Century, such as Ludwig Wittgenstein's theory on language, investigated whether it is possible to define a "precision language" that would allow communication without ambiguity. However, further work on theory and experiments showed that language is a social construct where terms and meaning are defined through negotiation between communicating parties.

These fundamental considerations can be applied to the design of mark-up languages and ontologies. They show that there are limits to generalisation and, in turn, give an important role to negotiation in the definition of language elements and their meaning. This is particularly important in the distinction between top-down and bottom-up definition of mark-up languages and ontologies. This presentation characterises different use cases and outlines the scope of mark-up languages and ontologies.