



Executive Summary of D 11.8

Working with living documents

Authors: Johannes Busse
Ontoprise
busse@ontoprise.de

Bernt Bremdal
Cognit
Bernt.Bremdal@cognit.no

Work-package: WP11
Type: Deliverable
Distribution: Public
Status: Version 1.0 (final)
Release Date: 2009-02-27
Deliverable Coordinator: Johannes Busse
Reviewer: Bernt Bremdal
Area Coordinator: Luca Gilardoni - Quinary Spa, Milan, Italy

Project Coordinator: Fabio Ciravegna, University of Sheffield
EU Project Officer: Dirk Van der Eecken

Summary

A living document is a document "@work". Why should one annotate a document instead of simply perform another authoring step? Because annotating text (i.e. structuring / segmenting text, communicating disambiguation or anaphor resolution information or simply citing and commenting 3rd party text) is a genuine act of authoring!

Problems however occur when looking in detail at the semantics and pragmatics of annotations. The deliverable analyses in detail what a user might intend when she annotates a text. We point to some inherent complexities and fundamental issues of tagging i.e. of living documents. When looking for more generic answers to these one would have to dive deeply into semiotics. However, this is a complex area. We cannot do here more than bringing about the necessary respect for the matter which we otherwise find rather absent in discussions on semantic tagging. For the difficulties are not so much triggered by the pure semantics, but the subjectivity related to author's intention compared to the reader's assumed intention.

Dependent on the granularity of the selected chunk, the meaning of the annotation target in an ontology and -- of course -- the context of the annotation itself there are different possibilities of what an annotation is intended to mean. Annotating a single phrase simply supplies additional data to the phrase - data which either helps the reader to disambiguate the phrase or which adds information to an already identified object. If the user annotates a more medium grained text chunk (e.g. a whole paragraph, a table) she might add data, metadata or meta meta data to it - very similar to the possibilities if she'd annotate a document as a whole. In fact annotating a medium grained text chunk might be interpreted as the act of declaring that the respective text chunk could be interpreted as a stand alone information object (describing a distinct subject) which is worth being annotated.

The most relevant variable however is the modelling of the objects we use for annotating within the ontology itself. In more simple cases the annotation tag is a class or an instance. In a more complex case the annotation tag refers to a property in the ontology. The ultimate and most expressive case is that the user want's to annotate a chunk of text with a more or less RDF graph itself. For we cannot decide in advance how to represent a user given ontology adequately -- i.e. we cannot please this ultimate approach in general -- we have to look for a much more simple solution.

In order to allow for interactive annotation we suggest to aggregate annotations to an intermediate level of text granularity like a section, a paragraph, a table, a figure and so on. Annotation aggregation per definition is a lossy transformation. However, the losses are not necessarily bad. Throwing away information can be understood as an act of information consolidation, which is i.e. true if it is the user which amends the aggregation.