Attempto Controlled English as Ontology Language

Tobias Kuhn
Department of Informatics
University of Zurich
Switzerland
Email: t.kuhn@gmx.ch

***Abstract*— Using the domain of protein interactions we introduce Attempto Controlled English (ACE) as an ontology language. We then use this ontology in two example applications. In the first example, we propose to summarize the results of scientific papers in ACE. These summaries would make scientific results accessible by computers and thus directly usable for any reasoning task. In the second example we present a novel kind of wiki that allows users to extend and to maintain ACE ontologies.**

This work is a cooperation of Zurich (working group I2) and Dresden (working group A2).

I. Introduction

A major drawback of most ontology languages is their formal appearance. Domain specialists are often not familiar with formal languages and they do not want to learn a complicated formal knowledge representation language. Attempto Controlled English (ACE) could close this gap. ACE is a formal language that looks like natural English. Every ACE text can be translated unambiguously into first-order logic.

Table I shows how the fact 'everything that is a protein has a terminus' is expressed in different knowledge representation languages: first-order logic (FOL), OWL, UML, and ACE. The OWL representation (using the RDF/XML syntax) is the most verbose and – from the human perspective – the least readable one. The representations in first-order logic and UML are more concise, but they are still not understandable for people who are not familiar with formal notations. The ACE representation, in contrast, should be immediately understandable for any English speaking person. It looks perfectly like natural English and thus the reader might not even recognize that it is a formal language.

We present now how ACE can be used as ontology language, and afterwards we will show two possible applications. In both cases we focus on the domain of protein interactions, and all the examples in this text are chosen from this domain.

II. ACE as an Ontology Language

In order to provide basic structures for ontologies in ACE, we adopt the elements from Description Logics: individuals, concepts, and roles. We call them *ontology elements*, and use them to represent the foundation of ontologies in ACE.

Individuals stand for single objects of the domain. They are represented in ACE as *proper names* like 'IRAK2' or 'Alzheimer'.

Concepts stand for classes of objects and there are two possibilities how to express them in ACE. *Common nouns* are the most straight-forward way. The noun 'protein', for example, can stand for the concept of all proteins. As a second possibility we can use the *positive form of adjectives*. The adjective 'organic', for example, might be used for the concept of all organic substances.

Roles stand for binary relations between objects, and they can be expressed in four different ways. First of all, we can use *transitive verbs* for expressing roles. For example, we can use 'interacts-with' to express a relationship between proteins. Next, we can combine transitive verbs with *adverbs*. For example, we can use the adverb 'directly' together with the transitive verb 'interacts-with' to express the role 'directly interacts-with'. As a third possibility we can use *of-constructs* like 'is a part of'. Due to the syntax of ACE, 'of' is the only allowed preposition for nouns. Finally, we can use *constructs with the comparative form of adjectives* like 'is larger than'.

The examination of statements about protein interactions showed that normal roles are often not sufficient to express the needed information. We can express simple statements like 'P1 interacts-with P2', but we cannot express statements with contextual information like 'P1 interacts-with P2 in Yeast' or 'P1 interacts-with P2 in Microfilament for Motor-Activity'. In order to be able to express such statements, we want to allow roles to have such additional information. In natural English we usually express such information with prepositional phrases, and this is exactly the way we do it in ACE.

Using these ontology elements, we can express for example the sentence

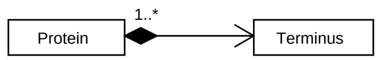
P1 is a protein and directly interacts-with P2 in Yeast.

where 'P1', 'P2', and 'Yeast' are individuals, 'protein' stands for a concept, and 'directly interacts-with' stands for a role. The phrase 'is a' is used to assign the individual 'P1' to the concept 'protein'. The conjunction 'and' connects the statements flanking left and right, and the preposition 'in', finally, connects to the context 'Yeast'.

In the following we present two example applications of an ACE ontology. First, we show how ACE can be used to express the results of scientific papers in formal abstracts, in oder to make text mining more powerful. Second, we present a novel kind of wiki that contains articles written in ACE. Extending and maintaining an ontology becomes (almost) as easy as editing a common wiki.

TABLE I

EXAMPLE IN DIFFERENT KNOWLEDGE REPRESENTATION LANGUAGES

FOL	$\forall X(protein(X) \rightarrow \exists Y(terminus(Y) \wedge has(X,Y)))$
OWL	<owl:Class rdf:ID="Protein"> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty rdf:resource="#has"/> <owl:someValuesFrom rdf:resource="#Terminus"/> </owl:Restriction> </rdfs:subClassOf> </owl:Class>
UML	
ACE	Every protein has a terminus.

III. FORMAL SUMMARIES FOR SCIENTIFIC PAPERS

Biomedical scientists are challenged by an ever-increasing amount of scientific papers. The indexing service *PubMed*[1] shows the huge quantity of literature that the scientists have to face. It contains at the moment 16 million articles and grows every year by over 600'000 articles. All these biomedical articles are written in natural language. That means that we cannot easily process them with computers. But, facing the quantity of literature, it is clear that we need computational support in order to manage the contained knowledge.

Instead of using natural language processing, we suggest an alternative approach: The authors of scientific articles formally summarize their own results in ACE. Such formal summaries are added to the articles which makes them processable by computers. ACE is on the one hand easy to learn and understand, and on the other hand it is expressive enough to represent even complicated scientific results.

It is clear that this approach is not applicable for papers that have been written without the formal summaries, and that means that we still need natural language processing or manual extraction for such papers. Thus it is rather a concept for the future than a solution for today's problems.

Figure 1 shows how the frontpage of an article with an ACE summary could look like. In order to allow the researchers to write such formal summaries, we support them with an authoring tool that hides the technical details. It helps the researchers to choose the right words from the lexicon, to use these words as intended by the ontology, and to write texts that are compliant with the ACE syntax and with the ontology.

IV. ACE FOR SEMANTIC WIKI

If the extension and maintenance of an ontology is as easy as editing a wiki, then it would be much easier for a scientific community to provide a consistent and complete knowledge base. Each member would be able to contribute, without learning a complicated language. We built the prototype ACEWIKI that shows how a semantic wiki using ACE could look like.

Figure 2 shows a screenshot. Like common wikis, ACEWIKI consists of articles. Every article is attached to exactly one ontology element. By clicking the *add*-button, the user can create ACE sentences and the words of the sentence are automatically linked to the existing articles.

For the creation of ACE sentences, the user is supported by a special editor, that shows step by step the possible continuations of the sentence. Thus the user does not need to know the details of the ACE syntax. The words have to be chosen by navigating through the concept- and role-hierarchies. If the user wants to use a word that does not yet exist, he can create it on-the-fly. The editor is aware of the underlying ontology and thus it allows only to create sentences that comply with that ontology. If the role 'localizes-to', for example, has the range 'cellular-component', then we would not be able to create a sentence like 'Act1 localizes-to a protein', since this would violate the range-condition.

There exist many other attempts for semantic wikis (e.g. IkeWiki[2]). ACEWIKI is different in that the formal semantics are not just attached to the articles, but the contents of the articles themselves are completely formal. Nevertheless ACEWIKI looks almost as natural as a normal wiki.

V. CONCLUSIONS

Formal summaries for biomedical papers could be a first step towards better communication and persistence of biomedical knowledge. In the case of wikis, ACE could move them to a higher level, providing a logical representation of the contained knowledge.

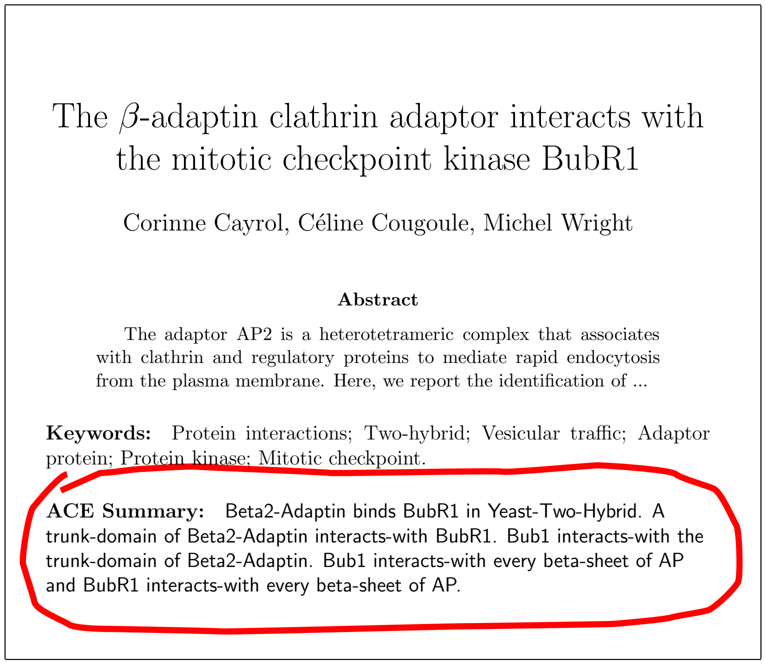
Altogether, we believe that controlled natural languages like ACE can combine the power of ontologies with the convenience of natural language. ACE can make scientific results readable by humans (since it looks like natural English) *and* processable by machines (since it is automatically translatable into first-order logic). Humans and machines could work hand in hand.

REFERENCES

[1] Corinne Cayrol, Céline Cougoule, Michel Wright. *The β-adaptin clathrin adaptor interacts with the mitotic checkpoint kinase BubR1*. In "Biochemical and Biophysical Research Communications", Volume 298, Issue 5, Pages 720-730. 2002.

[1] http://www.pubmed.gov

[2] http://ikewiki.salzburgresearch.at



The β-adaptin clathrin adaptor interacts with the mitotic checkpoint kinase BubR1

Corinne Cayrol, Céline Cougoule, Michel Wright

Abstract

The adaptor AP2 is a heterotetrameric complex that associates with clathrin and regulatory proteins to mediate rapid endocytosis from the plasma membrane. Here, we report the identification of ...

Keywords: Protein interactions; Two-hybrid; Vesicular traffic; Adaptor protein; Protein kinase; Mitotic checkpoint.

ACE Summary: Beta2-Adaptin binds BubR1 in Yeast-Two-Hybrid. A trunk-domain of Beta2-Adaptin interacts-with BubR1. Bub1 interacts-with the trunk-domain of Beta2-Adaptin. Bub1 interacts-with every beta-sheet of AP and BubR1 interacts-with every beta-sheet of AP.

Fig. 1. Article with ACE summary (see [1])

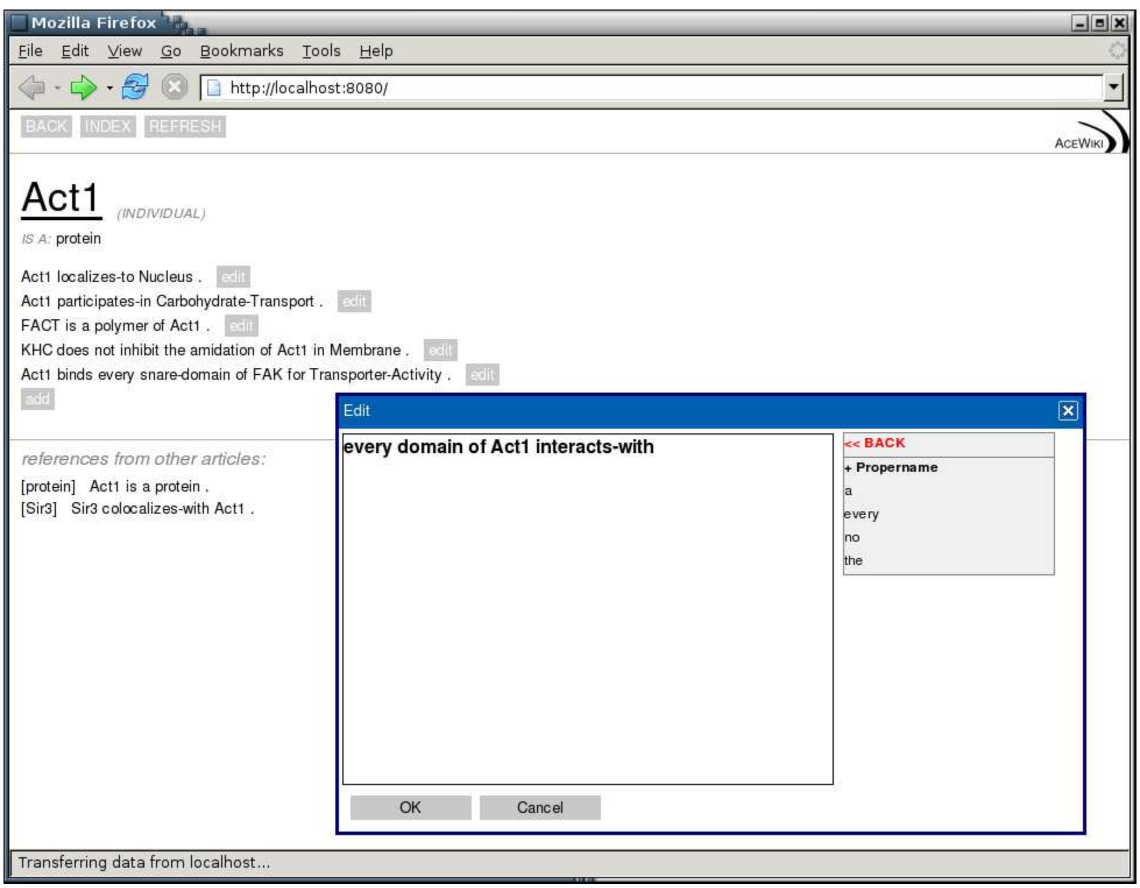


Fig. 2. AceWiki