

Adding Semantics to Scientific Collaboration Documents

Muthukkaruppan Annamalai¹, Lito Cruz², Leon Sterling¹, Glenn Moloney¹

¹ The University of Melbourne, Victoria 3010, Australia
{mkppan,leon}@cs.mu.oz.au, glenn@physics.unimelb.edu.au

² RMIT, Victoria 3000, Australia
lcruz@cs.rmit.edu.au

Introduction

The WWW has become the *defacto* collaborating medium for distributed scientific community to interchange information among them. There is still much human mediation involved to utilise this information. The human effort can be largely reduced, when the information is exchanged with meanings attached. The key enabler for this meaningful collaboration on the Semantic Web is ontology.

An ontology does not have to be a universally standardised language. However, its usability depends chiefly on its adoption as a collaborating language by a user community. Following this, our research aims to demonstrate that suitable ontologies can be constructed to support the exchange of information at the *knowledge level* within a distributed community, such as the Experimental High Energy Physics (EHEP) collaboration.

Tacit Knowledge in Collaboration Documents

The EHEP collaborative work revolves around experimental analyses. The research groups within a collaboration analyse the huge sets of data produced in an experiment, using various analysis techniques. The results of the analyses are communicated to fellow researchers in the form of pre-prints and research notes.

In the absence of a prescribed set of analysis description guidelines, authors generally state aspects of the analysis procedure, which they think is essential to be conveyed to the readers. As in the case of experimental science publications, there is a tendency among authors to presume readers already have knowledge about the analysis procedure. The Experimental analyses described in this fashion, with publication bogged down with implicit knowledge are prone to be misunderstood, particularly by researchers who are not familiar with the kind of analyses mentioned in the document. Often times, a researcher trying to replicate published experimental analyses, ends up with relatively different result. Precious time is expended trying to correctly interpret the experimental analyses, which often results in tedious debugging of the analysis procedure. Debugging an experimental analysis described by authors

who profess somewhat different ontological commitment about the domain is indeed a daunting task.

Explication of the Experimental Analyses

The misinterpretation problem can be traced to lack of structure and semantics in the published scientific documents. We believe it can be safely resolved if an analysis process is described explicitly in definite terms to peer researchers.

To begin, we propose the creation of a formal scientific document, called *analyses report*, which describes the completed experimental analyses according to EHEP ontologies in an orderly manner. An organised narrative of the analyses would allow a meaningful description of the content. Publishing the analyses with annotations linked to ontologies published on the Semantic Web can ensure optimal exchange of information between researchers within a collaboration. This opportunity to embark upon an innovative way of handling scientific information generated within an EHEP collaboration is illustrated in **Fig. 1**.

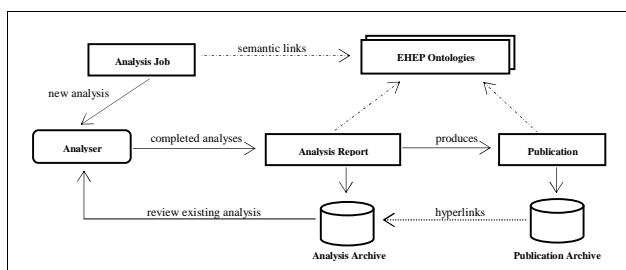


Fig. 1. Handling the EHEP scientific collaboration documents. Researchers prepare and deliver the semantically marked up analysis reports and publications, which can be archived and referred during subsequent experimental analysis. The content of the archives can also be searched more productively using precisely defined queries. Jobs described using machine-readable ontological terms can be processed directly by agent analysers

It affirms the belief that the Semantic Web can indeed change the way scientific knowledge is produced and shared, as envisaged by Berners-Lee and Hendler [2001].