COD: Context-preserving Ontology Decomposition Development and Application

1. ABSTRACT

In Computer Science, Artificial Intelligence is "the science and engineering of making intelligent machines" [1]. In the last few decades, machines and software have been developed to simulate "intelligent behaviour": for example, the field of Knowledge Representation and Reasoning (KR) is concerned with developing suitable languages and computer systems to represent knowledge and automate the explicitation of implicit knowledge. **Ontologies** are systems of this kind: in practice, they are computer processable sets of special logical formulae known as **axioms**. These axioms are modelled by knowledge engineers who encode the interrelations between the objects of the domain of interest that experts indicate, e.g., "a femur is a kind of bone". The Web Ontology Language **OWL** is a standard language that has strong computational properties and can be processed by special inferences engines known as **reasoners**. Ontologies enable experts to model the knowledge of a domain of interest by providing a syntax for describing domain concepts in an unambiguous and non contradictory way, facilitating shared understanding and knowledge exchange.

The last 3 decades have seen a rapid growth in the development of ontologies especially over the bio-medical domain: one notable example is the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) adopted as the standard for the treatment of medical knowledge in more than 50 countries, including the UK; SNOMED CT covers more than 300K terms, from diseases to diagnosis, from drugs to body structures, and describes their interrelations. Users, though, do not generally need all the knowledge derivable from such large, comprehensive ontologies: instead, they could be interested in a selection of topically-coherent fragments of the ontology, and in their interrelations.

Unfortunately, there is still no logically sound notion of **topicality** for ontologies; i.e., there is no method to automatically determine whether, and if so how pieces of an ontology link together; in particular, it is not possible to identify a **topic** (e.g., "bones") with relevant axioms/knowledge, and then suitably combine it with other topics (e.g., "bones and fractures") and their corresponding axioms.

In this direction, it would be interesting to investigate the notion of a **context** [2], defined to be a subontology that a user can view. Ideally, a context is a fragment that domain experts and ontology developers have identified as topically coherent. However, contexts are not "safe" w.r.t. semantics, i.e., the meaning of a term *t* in a context and the meaning of the same term *t* in the whole ontology could significantly differ.

The COD project presented here aims at strenghtening the logical properties of contexts to preserve the semantics of terms, and at formalising a notion of "meaningful logical interrelation" between contexts. The two main goals of this project are 1) to provide users with an overview of the ontology's content; and 2) to shade some light in the direction of defining a notion of topicality in ontologies.

2. BENEFICIARIES AND IMPACT

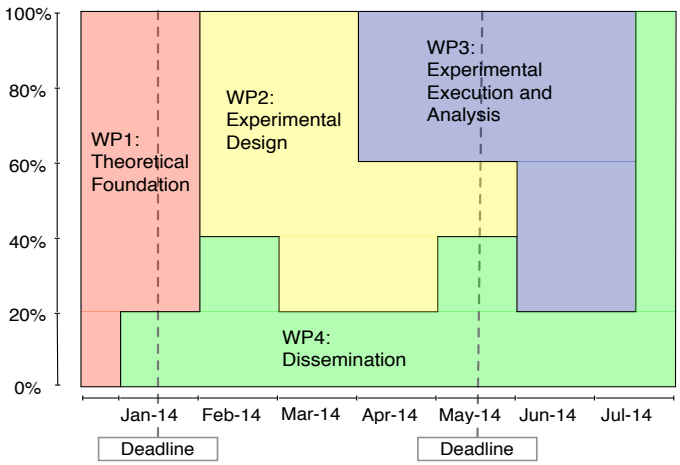
The main beneficiaries of the results obtained from this project are ontology engineers and ontology users. The development and mainainance of an ontology can present challenges, and modelling errors can emerge. Current supporting tools require the developers to actively search for errors. By determining meaningful logical interrelations between what modellers have identified as topics, it will be possible to reveal modelling errors, for example by identifying missing connections between terms, or unwanted aggregation of unrelated terms.

3. PROGRAMME

This project would last from December 2013 to July 2014.

This project is divided into 4 WorkPackages. The graph on the right is an enriched Gantt chart, where:

- Each block represents a WorkPackage; the respective positions show the logical dependency between the blocks: the lower ones support / inform the upper ones
- The x-axis values represent the time (in months) from the start of the project
- The y-axis values represent the distribution of the effort put in the active WorkPackages
- The dotted lines represent the points in time where the corresponding objectives will be met



The following describes the aims and deliverables of each WorkPackage:

WP1: Theoretical foundation

The aim of WP1 is to lay the foundation of COD, by: 1) formalising the notion of a "context", which is a set of annotated axioms that experts identify to belong to a same topic; 2) defining Context-preserving modularity; and 3) investigating its computational properties.

Deliverables: D1.1) Algorithm to compute the Context-preserving Decomposition of an ontology
D1.2) Paper to submit to IJCAR-14 (submission deadline: 15 January 2014)

WP2: Experimental Design

The aims of WP2 are to define suitable research hypotheses and design experiments that can confirm or reject them. To obtain meaningful results, an appropriate corpus of ontologies needs to be selected. The activities to perform in WP2 essentially are: 1) Generation of ontologies topically-labelled by domain experts; 2) Gathering of the corpus; this activity could require a direct contact with an institution where ontologies are developed.

Deliverables: D2.1) A set of research hypotheses.
D2.2) A tool for computing a Context-preserving Ontology Decomposition.
D2.3) A corpus of ontologies annotated with context-oriented labels.

WP3: Experimental Execution and analysis

This WP will deal with the execution of the experiments designed over the corpus selected, the data analysis, and the evaluation of the research hypotheses set in

Deliverables: D2.1) Results of the experiments carried out.
D2.2) Paper to submit to ISWC-14. (submission deadline: May 2014)

WP4: Dissemination

The results obtained will be presented at international conferences, primarily aiming at IJCAR-14 to be held in Vienna at the Summer of Logic.

[1] McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (1955). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence".

[2] Baader, Franz; Knechtel, Martin; Penaloza, Rafael (2009) "A generic approach for large-scale ontological reasoning in the presence of access restrictions to the ontology's axioms".