Future Generation Computer Systems 54 (2016) 423–434

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery

Rodrigo Bonacin [a,b,*], Olga Fernanda Nabuco [a], Ivo Pierozzi Junior [c]

[a] CTI Renato Archer, Rod. Dom Pedro I, 143, 6, 13069-901, Campinas, Brazil
[b] FACCAMP, Rua Guatemala, 167, 13231-230, Campo Limpo Paulista, Brazil
[c] Embrapa Agricultural Informatics, Av. André Tosello, 209, 13083-886, Campinas, Brazil

HIGHLIGHTS

- Describes interoperability issues in an ontology engineering process.
- Presents the design of a cross domain large ontology.
- Presents experiences of using the ontology in an information recovery scenario.
- Presents challenges and resources needed to work with domain specialists.

ARTICLE INFO

Article history:
Received 14 December 2014
Received in revised form
28 March 2015
Accepted 24 April 2015
Available online 21 May 2015

Keywords:
Semantic interoperability
Ontology engineering
Knowledge representation
Semantic web

ABSTRACT

Agriculture is both highly dependent on water resources, and impacting on these resources. Regardless of advances in the area, the impacts of water scarcity and climatic changes on agriculture, as well as the impacts of agriculture on water resources, remain uncertain. Potentially, collaborative systems can support the management and information sharing of multifaceted and large scale data sources, providing valuable and indispensable information for research. However, these solutions rely on semantic interoperability, the construction of complex knowledge representation models, as well as information recovery. This work describes interoperability issues in the engineering process of the OntoAgroHidro, an ontology that represents knowledge about impacts of agricultural activities and climatic changes on water resources. The paper presents representative scenarios and questions, and discusses the reuse and integration of concepts using knowledge visualization techniques. Experiments on the information recovery scenario point out the potential and limitations of the OntoAgroHidro.

1. Introduction

Climate changes impact our lives in a series of chained events, in terms of the effect on water supplies, agriculture/food production, urbanization, electric power generation, to cite a few. According to The United Nations Framework Convention on Climate Change [1] climate change means "a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods". Other important agencies and institutions have slightly different understandings (e.g., the IPCC [2]), however they agree that the planet is warming and extreme weather events can be observed.

Agriculture depends on climate, and the variation from extreme precipitation to long periods of drought affects crop and livestock production. Water is fundamental to life and the climate affects the quality and reliability of water, affecting local ecosystems or regional biomes. Agriculture is also one of the human activities that has the greatest impact on the environment. Nowadays, agriculture is responsible for more than seventy percent of the world's freshwater consumption [3]. It is also responsible for the water quality degradation process (e.g., eutrophication and pesticide contamination) and the physical effects on the soil and water

* Corresponding author at: CTI Renato Archer, Rod. Dom Pedro I, 143, 6, 13069-901, Campinas, Brazil.
E-mail addresses: rodrigo.bonacin@cti.gov.br (R. Bonacin), olgafernanda@gmail.com (O.F. Nabuco), ivo.pierozzi@embrapa.br (I. Pierozzi Junior).

http://dx.doi.org/10.1016/j.future.2015.04.010

bodies (e.g., erosion and aggradation). In this context, research studies have investigated how to mitigate reciprocal or unidirectional impacts of climate changes and agriculture on water resources. Research has highlighted, for example, new varieties of crops that are more resistant to drought or to wet seasons, new techniques to improve the effectiveness of water use, agricultural wastewater treatment techniques, as well as new crops that mitigate CO_2 warming by reconverting CO_2 into O_2.

However, the Brazilian Panel on Climate Change [4] have noticed gaps of information including the lack of good quality meteorological information based on complete, long-term data series and the lack of knowledge on current groundwater recharge. Studies from Assad and Pinto [5] point out that the Brazilian *Cerrado* and the semiarid Northeast regions are home to the largest collection of global warming resistant genes in the world. Nevertheless, genomes have to be built in order to find these genes. This is a typical scenario that requires multi, inter and transdisciplinary collaboration among scientists from several domains to improve research globally. Knowledge organization, integration and recovery technologies are crucial to enable such collaboration where scientists must access, trust and understand shared information.

Cross domain integrator systems enable us to manage complex information from multidisciplinary problems such as "the impacts of agriculture on water resources". This problem requires semantic interoperability of systems and data sources from multiple domains such as agriculture, hydrology, biology, chemistry, and economics, among others. In this sense, the use of formal and standardized models of knowledge organization and representation is highly recommended, as they are significant resources to improve interpersonal communication, knowledge recovery and semantic interoperability of information systems.

However, the semantic interoperability of these systems relies on open research issues such as reuse, alignment and mappings of ontologies. In this context, some typical difficulties of knowledge engineering processes include: existence of various media formats in which information is produced and distributed; the multidisciplinary nature of knowledge involving teams of professionals from various fields and specialties; and language and communication problems between experts due to different nationalities or schools of thought. Such difficulties are directly related to elements in the human processes of cognition, meaning and communication.

In this work, we propose the OntoAgroHidro as an ontology to represent knowledge about the impacts of climatic changes and agricultural activities on water resources. The objective is for the ontology become a component of Embrapa's research network system (AgroHidro), which aims to support integration and information sharing among a range of institutions and researchers. The advances in understanding the cause–effect phenomena, by the network scientists, depend on information recovered from a multitude of sources. The heterogeneity of such data sources creates a barrier to scientists in terms of making connections among multiple domains of information. These scientists have questions that frequently depend upon sophisticated information recovery mechanisms.

This work describes the interoperability issues in the engineering process and the application of the ontology in an information recovery scenario that stresses the multidisciplinary nature of the problem. The engineering process is based on reusing the existing knowledge representation models [6]. In this document, we present how the reused models were interconnected, starting from the analysis of the interoperability needs of the existing and planned data sources, the use of a core ontology as integration strategy, and the modeling of concepts that carry out the interconnection among the reused models. The paper then illustrates the OntoAgroHidro and discusses key interoperability issues using visualization tools and representative scenarios. Experiments on an information recovery study stress the potential of the proposed ontology, its limitations, and future challenges in the modeling process.

We expect to contribute with ideas about an ontology engineering process for semantic interoperability of multidisciplinary domains, as well as to present experiences from applying this process. The paper is organized as follows: Section 2 describes the problem and the background that support our work; Section 3 presents the ontology engineering process focusing on the reuse and interoperability aspects considered during the construction of the ontology; Section 4 illustrates the elicited interoperability requirements, the core ontology and the connection of the reused concepts/terms; Section 5 presents the application of the OntoAgroHidro in a study of an information recovery scenario; and, Section 6 concludes and presents the next steps of this research.

2. Problem and background work

Multidisciplinary domains impose challenges on modeling activities and interoperability issues. Section 2.1 details the problems and challenges of modeling concepts from agriculture and climatic changes, as well as of relating the concepts to construct a consistent model. A scenario is used to illustrate the multidisciplinary aspects of the AgroHidro domain. Section 2.2 presents projects that propose parallel efforts in the context of environmental studies and other related fields. This section also highlights the limitations and concepts from existing work that were incorporated into our solution.

2.1. *Challenges of modeling agriculture and climate changes and their impacts on water*

The weather is becoming extreme and volatile [7]. What are the consequences of this on agriculture and water availability and quality? Agriculture is both highly dependent on water resources, as well as impacting on them. To understand these phenomena, scientists have to analyze a huge amount of data from heterogeneous sources [8]. Many investigations on scientific collaboration systems aim to deal with the diversity and complexity of multidisciplinary domains combined with large scale data sources. These systems also require research on complex infrastructure, architecture and tools [9].

A collaborative solution may lead to a framework to acquire, organize, and describe not only (raw) data but also its meaning. For instance, there is no easy answer to the following question: "How is the quality of Brazilian water sources and rivers affected by crop mono-cultivation with intensified use of chemical fertilizers and pesticides?" One can start by analyzing the fact that farms are irrigated by nearby water resources. This analysis, however, depends on well defined terms, such as: "What are water resources?", "Is it good quality and potentially useful freshwater?", "What is water quality?", "What are chemical fertilizers?", "What chemical components are used in chemical fertilizers?", "Are they the same all over the world?", "Do they have alternative names?", "Are there different levels of water quality?". Each question derives a series of more specialized questions. As a consequence, instead of an easy answer, one can expect an exponential growth of questions. The answer depends on the access to different sources of data, systems and documents.

The semantic web proposes the use of knowledge representation languages to organize and understand the information produced and shared through the Web. Nevertheless, there are multiple proprietary solutions that use various incompatible models and languages. The implementation of proprietary solutions is usually faster and easier, however they result in islands of