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 information and long term interoperability problems. Ontologies using the Web Ontology Language (OWL) [10] can support modeling concepts of complex domains and enhance the interoperability of multiple sources of data [11].

There are several projects that specify metadata, semantic models and services to improve their representation systems in related disciplines. We highlight here some of the solutions that most closely supported our work. These solutions are all based on different subjects of study; however, they relate to our project in terms of using semantics to achieve interoperability, or using ontologies to describe water quality or water resources.

2.2. Background on related fields

The following projects partially deal with our domain or use a methodology related to ours. Reuse ontology is a common approach, which, based on other studies, improves the understanding of the domain, avoiding rework and promoting interoperability.

The Coral Reef Ecological Observatory Network (CREON), for instance, implemented a solution for information sharing and interoperability [12]. Its solution uses the Ecological Metadata Language (EML) to describe data generated by its members focusing on observational data. They also proposed the SOnet (Semantic Observations Network) as part of the Semantic Tools project. EML is an XML Schema, which describes ecological data using the resource conception from Dublin Core[1] (DC). It also provides features such as datasets, literature, software and protocols. The CREON includes the Global Biodiversity Information Facility, which uses the Darwin Core, an extension of DC for biodiversity information. The Long Term Ecological Research (from CREON) uses and provides data sets from ecological process at time scales spanning decades to centuries. The project also uses EML to define metadata of two other systems, the NBII (National Biological Information Infrastructure) and the FGDC (Federal Geographic Data Committee).

The SemantEco project [13] faces similar problems for dealing with distinct data sources, and integrates them under a proposal for the Semantic Ecology and Environmental Portal[2] (SemantEco). The SemantEco offers decision support tools that aim to help resource managers identify different environmental scenarios. They also consider the reuse of ontologies to improve the usability and interoperability of their system.

Xiaogang Ma et al. [14] employed semantic web tools to provide reliable information for the National Climate Assessment.[3] Their objective was to "increase understanding, credibility and trust" in the research conducted on climate change. They developed an ontology model of the Global Change Information System by applying a series of use cases to identify goals and other elements of the domain. They used software tools like CMapTools[4] to easily interact with the users and environmental scientists. They reused ontologies to achieve and improve interoperability as well as system usability.

During the construction of the OntoAgroHidro we studied several models available from related projects, such as Cuahsi[5] (Consortium of Universities for the Advancement of Hydrologic Science, Inc.) and SWEET[6] (Semantic Web for Earth and Environmental Terminology). The intention was to reuse the models (or part of them) to reinforce interoperability as well avoiding rework.

Cuahsi is a consortium of over a hundred universities and US organizations focused on the Hydrology domain. The main sources of information provided by the consortium are temporal series previously classified according to metadata described using tags and a controlled vocabulary. To overcome interoperability problems they developed their own information system called the *Hydrologic Information System* (HIS) [15]. The HIS is composed of three main functions: the *Central*, the *Hydro Servers* and the *Hydro Desktop*. The *Central* has a metadata catalog and an ontology that enables data registration and discovery. The servers have their own catalog of web services, the *WaterOneFlow*, registered on the *Central* to make their data available to all participants. *WaterML* [16] is an XML based language they call "the water communication language" used to retrieve information from the server, which provides location, time series, and variables in a standardized way. The Hydro Desktop is an interface that allows users to access data using a visual interface.

A total of 4090 concepts were modeled on the Cuahsi Ontology. The *Hydrosphere* is the *root* hierarchical concept, which has three sub-concepts defined as follows: (1) the *Physical* concept encompasses the physical water quality and quantitative terms, including temperature, density, area, pressure, and optics; (2) the *Biological* concept encompasses concepts related to water from the biological perspective, including biological taxa, indicator organisms, and biological communities; and (3) the *Chemical* concept encompasses concepts related to water from a chemical perspective, including organic, inorganic, stable isotopes, oxygen demand, nutrients, and radiochemical properties. The Cuahsi ontology presents a comprehensive representation of water quality, and was taken into consideration in the construction of the ontology to represent water quality measures. However, it did not model other domains that are crucial to the problem under analysis, for example, agricultural, territorial and climate concepts. Consequently, it had to be integrated with other models, which required format translations. Excel spreadsheets were used to represent their concepts, which did not conform to our requirements of using an ontology language to represent concepts.

JPL-NASA has developed a set of 200 ontologies with around 6000 Earth Science concepts [17]. The SWEET is structured according to nine top level concepts as follows: (1) *Representation* including concepts of Math, Science, Data, Time and Space; (2) *Process* including concepts of Physical Process, Mathematical Process, Chemical Process and Biological Process; (3) *Phenomena* including concepts of Ecological Phenomena and Physical Phenomena; (4) *Matter* including concepts of Living Thing, Material Thing and Chemical; (5) *Realm* including concepts of Ocean, Land Surface, Terrestrial Hydrosphere, Atmosphere, Heilosphere, Cryosphere and Geosphere; (6) *Property* including concepts of Binary Property, Categorical Property, Ordinal Property and Quantity; (7) *State* including concepts of Role, Physical, Chemical, Space and Biological; (8) *Human Activities* including concepts of Decision, Commerce, Environment, Research and Jurisdiction; and (9) *Relation* including concepts of Human, Physical, Time, Space and Chemical.

Each ontology of the SWEET set can be visualized (and reused) individually. The JPL-NASA classifies SWEET as middle-level ontology, where users can add domain specific components. In this sense, it provides a valuable contribution to the definition of concepts related to the AgroHidro problems. However, the core structure of SWEET (presented above) is not designed to address our problem and it does not define detailed concepts representative of our context. Nevertheless, the SWEET models were taken into consideration during the modeling process of our ontology by providing modeling solutions and structures for important concepts in the AgroHidro context, such as, the *water/body of water* concept and constructions to represent the *time* concept.

Another aspect to be considered is how to use ontologies to deal with semantic interoperability [18], document indexing, query answering and semantic searches [19–21]. Ontology has been a

---

[1] http://dublincore.org/.

[2] http://aquarius.tw.rpi.edu/projects/semanteco/.

[3] http://nca2014.globalchange.gov/.

[4] http://cmap.ihmc.us/.

[5] http://www.cuahsi.org/.

[6] http://sweet.jpl.nasa.gov/.

central concept since the early versions of the Semantic Web stack [22]. Ontologies extend the RDF models providing reasoning possibilities and semantic interoperability of the existing data sources. There are many studies on how to explore ontologies to improve semantic interoperability and integration; however scientific cooperative networks (e.g. AgroHidro) also demand complex architectural solutions to index and recover information from multiple data sources.

Li et al. [23] highlighted three core challenges for managing scientific research data: (1) to manage large quantities of heterogeneous data, (2) to support metadata-related services, and (3) to accommodate evolving and emerging knowledge. Aiming to deal with these challenges, the authors proposed an independent domain and extensible architecture based on ontologies for managing scientific digital data and resources. Their architecture is structured in four layers as follows: the *interface layer* with object, metadata, publishing, and search services; the *security layer*; the *business logic layer* including object and concept management, and reasoning services; and the *data access layer* including repositories, triple store, database and search index services. In the context of the current study, we explored interoperability and information retrieval requirements. These requirements also led to architectural and modeling issues as presented in [23]. However, the focus of our research is on modeling issues of multidisciplinary domains, including how to model core concepts that integrate the existing models, and how to explore the model's potential in an information retrieval scenario.

3. The ontology engineering process

The proposed methodology for ontology design emphasizes the reuse of existing models (*e.g.*, Cuahsi and SWEET), as well as the promotion of interoperability between models and systems. The entire development of the ontology was driven by scenarios. The first "macro scenario", which guided the construction of a core/base model, was, "the impact of both agriculture and climate changes on water resources". This scenario is quite ambitious, as it includes aspects from the climate change domain articulated with agriculture and hydrology. It delimits the scope of the study on fundamentals for basic definitions and for modeling the intersection of those domains.

During the development of the ontology, we interacted with both final users and with a smaller but representative group of domain experts. The final users, in our case, are the participants in the AgroHidro[7] network, which aims to support the integration and information sharing among a range of institutions and researchers. There were 45 key researchers from 12 Embrapa research centers and 8 Brazilian universities in the network. The "domain experts" are a small team focused on deploying the scenarios, transforming them based on concepts and relationships. Modeling tools, including yEd[8] and CMapTools, were used to facilitate understanding of the concepts and communication between domain experts and ontology engineers. Protégé 4,[9] and its plugins, was the ontology editor employed.

From this initial scenario, we defined two key questions as the starting point for our modeling activities: "What are the impacts of agriculture and climatic changes on water resources?" and "What are the impacts of water quantity and quality on agriculture (and *vice versa*)?"

An iterative ontology engineering process was adopted and first presented in [6]. This process can be briefly summarized in six steps as follows:

1. We started with the analysis of the problem and cooperative design with domain specialists. In this step, the CMapTools [24] and yEd[10] where used to analyze key questions (extracted from scenarios) with a domain specialist. These tools supported the collaborative construction of concept maps [25] to express important domain concepts. These concept maps were used as inputs for drafts of OWL models.
2. During the second step we modeled the Core Ontology concepts with a domain specialist using the concept maps as a starting point. A small set of core classes were used to structure concepts from various domains. They represent the reciprocal interrelationships among environmental transformation events, agriculture and hydrological resources.
3. In the third step we evaluated the refining needs, i.e., the specialization[11] of the core concepts. Papers, reports, standards, among other documents, that describe the core concepts of our ontology were collected and analyzed by domain specialists.
4. The reuse of the existing models was analyzed in this step. During this step, we investigated the existence of knowledge representation models that aim to model concepts of each (sub)domain identified in the previous step. Their purpose and format, among other aspects, were also studied in order to decide which models should be reused.
5. We then employed one of four reuse alternatives: Reuse "as-is" (i.e., the model was reused according to its original conception and format), Adaptation (i.e., the model was adapted before being imported into our ontology), Remodeling (i.e., the model was completely remodeled, including format translations) and Conceptual Reuse (i.e., the model was analyzed and concepts and solutions inspired the modeling tasks), as described in [6].
6. In the last step, the ontology was interactively evaluated and submitted to consistence checking. This included its logical validation (using the reasoner Pellet[12]) [26] and the validation with domain specialists in two workshops.

The engineering process was executed as described above. However, while [6] focused on reusing ontologies, in this paper we focus on the interoperability of models, systems and databases. Interoperability issues were addressed during the entire process, as shown below:

1. During the first step (problem analysis and conceptual Codesign), we considered two scenarios.
    a. *Analysis of the existing system and data sources*. The analysis goes beyond data posted by the participants directly into the network system. It must be a unifying/integrating information system with information from various systems and data sources. To achieve this, we started by analyzing the various *datatypes* that should be shared through the network. This analysis included the identification of Embrapa's internal systems and databases with valuable data for the network. In addition, we analyzed data sources from external organizations, for example, governmental institutions and partners. This analysis was performed in four phases: (1) data source identification and description; (2) analysis of the technology used for importing and exporting data; (3) syntactic analysis of the database models, metadata structures, service interfaces, and proprietary export models; and (4) semantic analysis of the shared data, including the terminological aspects, vocabularies and the knowledge organization systems of each data source.

---

[7] https://www.agropediabrasilis.cnptia.embrapa.br/web/agrohidro/home.

[8] http://www.yworks.com/en/products/yfiles/yed/.

[9] http://protege.stanford.edu/.

[10] http://www.yworks.com/en/products/yfiles/yed/.

[11] In this paper the term "specialization" refers to the mechanism of establishing subclasses as well as the respective relationships, properties and individuals.

[12] http://clarkparsia.com/pellet/.

   b. *Analysis of the planned system and data sources*. This scenario takes into account the design of integration architecture and solutions. To this end, we analyzed related initiatives and solutions (e.g., Cuahsi and CREON) and the possible data sources that should be considered within the Brazilian context. This analysis included the technical architecture aspects (e.g., service oriented solutions), as well as semantic models and integration languages. Both analyses resulted in "requirements for integration and interoperability" that must be considered during the ontology design.
2. In the second step, we modeled the Core Ontology concepts using key questions and the requirements analyzed in the first step as a guide. In this work, the Core Ontology played the role of unifying concepts from various reused models. The ontology should equalize and maintain the consistence of information from many data sources. Some systems "cross/use" diverse concepts from several domains represented in the ontology. The Core Ontology should contain upper concepts to accommodate these areas and also concepts to integrate the reused ones. To deal with this problem, we elaborated and discussed a set of typical questions (with domain specialists) that must be answered by the ontology.
3. During the next step, we evaluated if the ontology model detailing was adequate. It was necessary to verify if the concepts of the systems and the data sources were properly represented in the ontology. Thus, for each concept in the ontology we verified the need to specialize it (or not) according to the interoperability requirements (step 1).
4. After having detected the need for a more detailed analysis, we studied alternatives aimed at determining the eligible sources (models) for reuse. One key aspect (in addition to those presented in [6]) is to consider the models adopted by the existing or planned system and data sources. For example, in the new information systems, Embrapa will adopt the Brazilian profile for geospatial metadata (PMGB) [27]. Consequently, PMGB was reused and integrated with the existing models.
5. In the fifth step, we defined the reuse strategy as presented in [6]. In terms of interoperability, the following aspects are highlighted:
   a. The reuse "as-is" is the preferential alternative. One alternative adopted in our architecture is the reuse of the DC through services that access documents and other resources;
   b. In all of the other alternatives (Adaptation, Remodeling and Conceptual Reuse) the concepts were imported or remodeled into the ontology. The PMGB was reused according to the adaptation alternative for representing concepts related to space. The Cuahsi ontology was reused according to the remodeling alternative for representing water quality and quantity measures. The SWEET, according to the conceptual reuse alternative for representing body of water and time concepts. We modeled concepts to link the imported or remodeled concepts with the reused ones. These "linking concepts" were identified by means of key questions formulated by domain specialists. These questions stressed the interoperability requirements identified in the first step and represent scenarios where multiple information sources and systems should be accessed;
6. In the sixth step, we performed the (re)evaluation of the model with domain specialists using an ontology visualization tool, simplified models, and metaphorical representations. The ontology engineers presented the models and discussed them in workshops with the domain specialists. The specialists described macro scenarios as a way to verify if the ontology correctly represented the information sources and how the sources could interoperate on these hypothetical cases. The version of the ontology presented in this paper was presented in two workshops of the AgroHidro network. During these workshops, the domain specialists analyzed key concepts from the core ontology and suggested scenarios including the one presented in Section 5 of this study.

4. Interoperability requirements and the Ontoagrohidro

The overall OntoAgroHidro model represents approximately 8500 concepts (including classes and domain instances). It includes the core concepts of the domain modeled on the upper hierarchical levels of the ontology. In this section, we present the analysis of the information sources and planned systems (Section 4.1), the core of the OntoAgroHidro (Section 4.2), and the analysis of concepts that integrate the main domain/issues through visualization techniques (Section 4.3).

4.1. The analysis of the information sources

As proposed in Section 3, we started with an analysis of existing and planned systems and data sources. Fig. 1 presents an overview of the architecture with a synthesis of the key elicited interoperability requirements for the ontology design. Proceeding from the top to the bottom of Fig. 1, there are four planned applications that will make use of the ontology: semantic search mechanisms, knowledge visualization mechanisms, conceptual support tools, and expert systems. The Semantic Search mechanism will recover documents and use the ontology to improve the results (e.g., by disambiguation) and to recover personalized information (by matching the user profile with semantic indexed terms). This mechanism is still under construction; there is however a functional prototype that uses simple term relationships. The second tool is a visualization shell to explore the content of the data sources by navigating through the ontology terms. The ontology has been used to illustrate non functional prototypes of this application. The third application is the use of the ontology (in many visualization formats) as a tool to support discussions with domain specialists, and to guide the organization of the concepts in the network. As mentioned above, the ontology was presented in two workshops. The fourth application is the construction of an expert system to directly support the domain specialists in their research. This application is still in the high level definition phase and demands further improvements on the ontological model, such as an extensive description of rules and axioms.

As shown in Fig. 1, the OntoAgroHidro is a fundamental piece in an Integration Bus which integrates six types of data sources. From left to right, there is the document management system, which indexes digital documents from Embrapa using a specific metadata model and terminology. In parallel there is the library database system with more than 800,000 documents including digital documents and catalog entries. These documents also use an *in house* developed metadata for indexing. There are also internal and external systems in our architecture that will be exported by a service layer (adapters may be needed). In addition, there is external data exported by proprietary formats and protocols. Some of the external sources include offline documents with important data, for example reports and statistical series of hydrological data. Finally there are posts included directly in the network. These posts can be, for example, experiments or tagged documents.

The bottom of Fig. 1 describes the major interoperability requirements (selected from an extensive list) considered in the ontology. The requirements are grouped according to the sources of information. Each source of information uses its own model, which results in integration needs. The *Document management system* and the *Library database system*, for example, use metadata

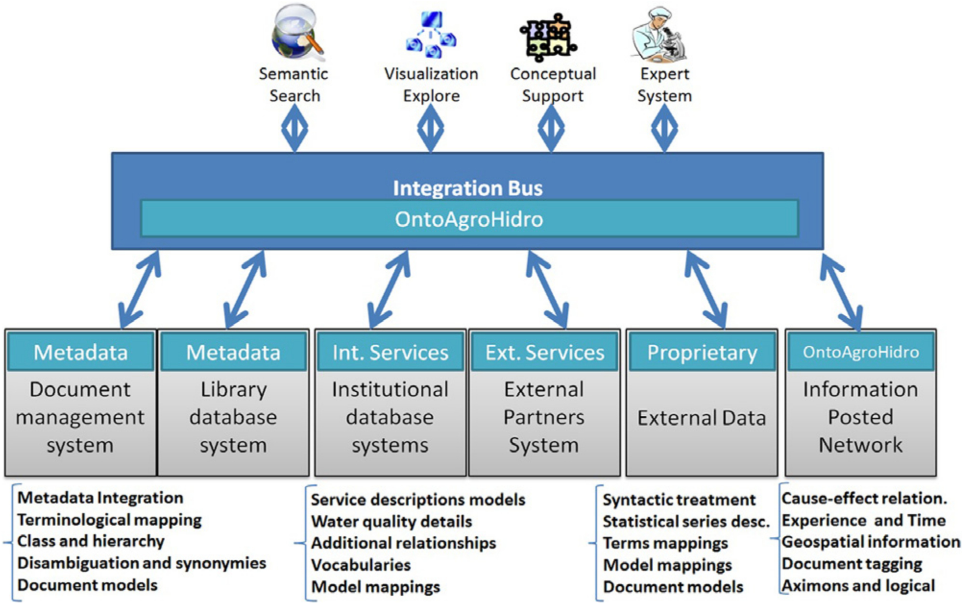

Fig. 1. Source systems and key interoperability requirements.

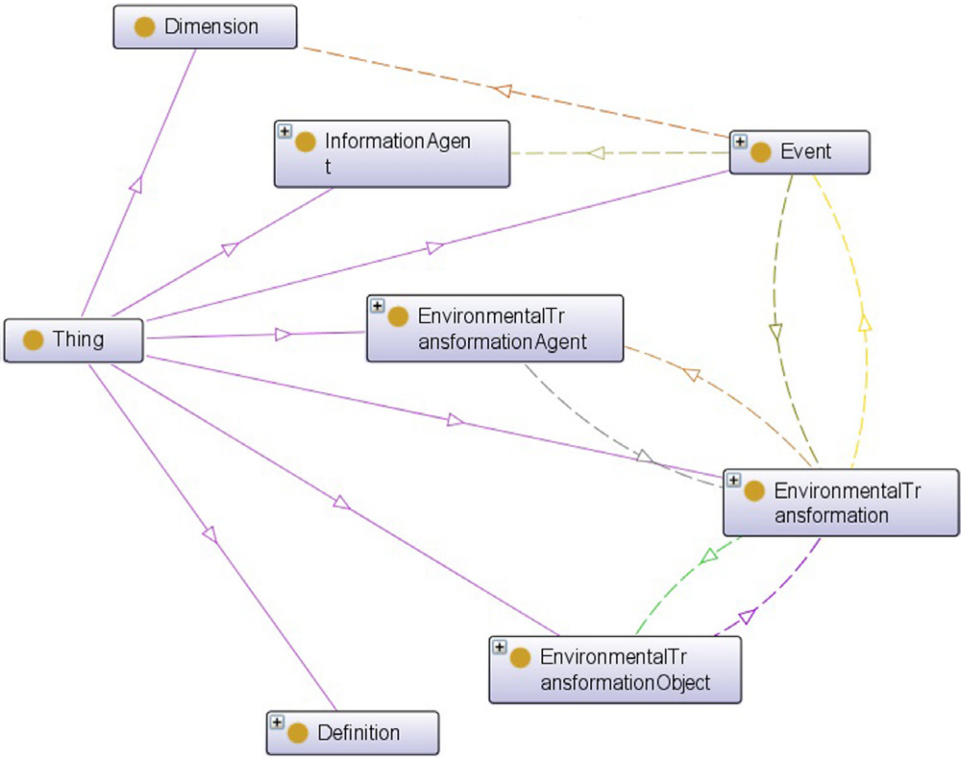

Fig. 2. OntoAgroHidro core concepts.

models to categorize documents. These models include basic information of the documents (e.g., authors, titles, summary), controlled keywords and the MARC (*MAchine-Readable Cataloging*) description. Consequently, these systems require integration such as metadata integration, terminological mapping, class and hierarchical descriptions, disambiguation and synonymies, and document models.

Despite the challenges, we used the elicited requirements as a guide to model the ontology. They were especially important in the definition of core concepts as well as the concepts that integrate the reused models.

4.2. Core ontology concepts and reuse

As shown in Fig. 2, the OntoAgroHidro is structured according to seven classes in the first hierarchical level (its first/previous version was presented in [6]). These classes, and respective properties and relations, represent a global conceptualization of the problem. This structure was constructed around the concept of environmental transformation, events and agents, which cause, or are objects of, these transformations.

The *EnvironmentalTransformation* (on the right of Fig. 2) is a core class that nominates a phenomenon that changes something in the environment. This concept is central to understanding how agriculture transforms the environment (in particular the water resources), and how the transformations on water resources (*e.g.*, those caused by climatic changes) can affect agriculture and food production in general, and vice-versa. The environmental transformation is structured according to two subclasses, one that covers all of the changes as a result of natural causes (i.e., a transformation caused by a natural phenomenon) and the other that covers all of the changes with anthropic causes (e.g., the transformations caused by anthropic influences including agriculture).

The *Event* class models events that cause environmental transformations. An event is characterized in *Space* and *Time* and is described by multiple *Dimensions*. For example, an event of agricultural expansion describes the increase of an agricultural activity in a certain place and at a certain time, causing multiples impacts on the environment, which result in environmental transformations. The *EnvironmentalTransformationObject* class represents the objects that are affected by a transformation. This transformation is caused by an agent modeled by the *EnvironmentalTransformationAgent* class. *ClimaticChanges* are considered in both cases, as a result of a series of environmental transformations as well as an agent that causes transformations.

WaterResource is a key subclass of the *EnvironmentalTransformationObject* class, since it is the main focus of studies on the AgroHidro network. *Agriculture* is another subclass of *EnvironmentalTransformationAgent;* it includes *Livestock* and *Crop*, as well as *AgriculturalManagement* and *ProductionSystem* subclasses.

The *Dimension* class (top of Fig. 2) describes the various aspects (dimensions) of an *event*. It includes the following subclasses:

- The *EnvironmentalDimension* class expresses environmentally related concepts subdivided into the *Hydrosphere*, *Pedosphere*, *Atmosphere,* and *Biosphere* subclasses. The *BodyOfWater* is a key subclass of *Hydrosphere* based on SWEET. It represents the accumulation of water.
- The *SocioeconomicDimension* class includes the social and economic impacts of an event. *Resource* is a subclass of this class, and the *WaterResource* indicates an intersection of water as hydrosphere and chemical element, with water as a resource (explored by humans) for food production, as well as an object of an environmental transformation. A *WaterResource* is composed of a body of water.
- The *SpatialDimension* class determines the location where an event occurs.
- The *TemporalDimension* class determines when a certain event occurs. The *SpaceTime* class is defined as an intersection of the *SpatialDimension* and *TemporalDimension*. The *SpaceTime* concept is essential to describing events and to characterizing other important concepts of the domain, such as biome.
- The *Agriculture* class represents agriculture as a dimension of the problem, including concepts related to agricultural activities (e.g., permanent crops, temporary crops, pastures and silviculture).

The *Definition* class covers concepts necessary to accurately define other concepts (mainly from *Dimension* hierarchy), as well as key concepts from related fields. The subclasses of *Definition* are:

- The *PhysicalDefinition* class models definitions from physics used in our ontology. It includes the following subclasses: *PhysicalProcess* (*e.g.*, Discharge, Evaporation, and Precipitation), *PhysicalUnit*, *WaterQualityPhysicalMeasure,* and *WaterQuantityPhysicalMeasure*. The majority of the concepts from physics were remodeled (*cf.* Section 3) from Chuasi and represent the physical aspects of the quality and quantity of the water bodies;
- The *ChemicalDefinition* class models definitions from chemistry used in our ontology. It includes the following subclasses: *ChemicalCompound* (*e.g.*, Inorganic, Nutrient, Organic, Radiochemical, Stable Isotopes), *ChemicalWaterIndicator* (i.e., a set of chemical compounds that indicate water quality), and *WaterQualityChemicalMeasure*. The majority of the concepts from chemistry were remodeled (*cf.* Section 3) by Chuasi and represent the chemical aspects of the quality of the water bodies;
- The *BiologicalDefinition* class models definitions from biology used in our ontology. It includes the following subclasses: *BiologicalCommunity*, *BiologicalTaxa* (*e.g.*, Benthic, Fish, Macrophyte, Nekton, Phytoplankton, Plant, and Zooplankton species), *IndicatorOrganism* (i.e., a set of organisms that indicates water quality), and *WaterQualityBiologicalMeasure*. The majority of the concepts from biology were remodeled (*cf.* Section 3) from Chuasi and represent the biological aspects of the quality of the water bodies.
- The *TimeDefinition* class models definitions for time related concepts. It includes the following subclasses: *TimeExpression (i.e.,* our subjective sense of time), *TimeMeasurement, TimeStandard*, *GeologicalTime,* and *SpaceTemporalExtension*. The majority of the time concepts were inspired (conceptual reuse) (*cf.* Section 3) by the SWEET ontologies.
- The *SpaceDefinition* class models definitions for characterizing geospatial concepts. It includes the following subclasses: *Location* (*e.g.*, altimetrical-bathymetrical, geographical position, geographical extension, and geopolitics), *SpaceMeasure* (i.e., standard measures of space), and *SpaceTemporalExtension*. The majority of the geospatial concepts were adapted (*cf.* Section 3) from the PMGB standard.
- The *DeviceDefinition* class models definitions for characterizing the devices used for collecting, analyzing, and measuring water quantity and quality parameters. This includes multiple device specifications and attributes;
- The *StatisticalDefinition* class models definitions for characterizing statistical concepts (e.g., sample, measure, and statistical series) used for analyzing water quantity and quality.

The *InformationAgent* is the last class of the first hierarchical level of the OntoAgroHidro. It models the agents that produce or contain information about environmental transformations. This class is specialized by subclasses that represent a *Person*, an *Institution,* or a *Project, i.e. entities* that produce, carry information, or can be considered information resources. The *InformationResource* subclass is linked to an external web service that describes objects using the DC metadata.

4.3. *Integration concepts and the visualization of the ontology*

Fig. 3 presents the visualization using Gephi 0.8.2[13] of the Cuahsi concepts reused into the OntoAgroHidro. Each "agglomerate" of the concepts in Fig. 3 is interconnected by relationships that cross the whiter spaces and concepts in the middle of the Figure. This visualization was used in validations in two workshops with domain specialists. We used a metaphorical representation of the *Victoria amazonica* aquatic plant species where leaves (concepts and individual agglomerates in the ontology visualization) are supported and interconnected by a submerged stalk (relationships and concepts used to connect other concepts in the ontology).

The connection of the reused models was guided by the interoperability requirements presented in Fig. 1. The main problem was to construct "bridges" among the concepts in various domains of the reused models. To illustrate this problem we present a typical question/scenario that the ontology would need to address: "*How would we define the quality of the water in the*

---

[13] https://gephi.org/.

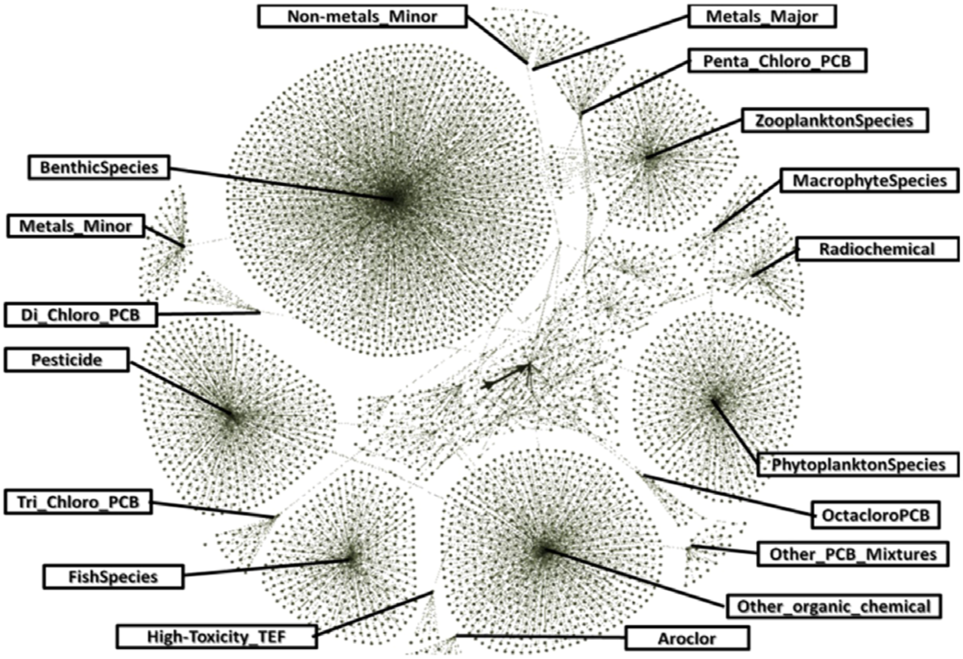

Fig. 3. Visualization of the cuahsi concepts reused in the ontology.

Atlantic forest biome in the soybean production regions during the last year?".

The answers can be based on recovered documents, results from the database system, or experiments posted in the network. In all cases, this question allows the use of several important concepts. When we ask for the "*… quality of the water …*". The "*water*" concept in the question above is not restricted to "H2O", but rather it is related to a "*BodyOfWater*". In our ontology we reused the concept from the SWEET ontologies, however the quality aspects of the water were reused from Cuahsi.

Fig. 4 iconically illustrates (using yEd 3.11) the *BodyOfWater* in OntoAgroHidro: number *1* indicates the position of this concept in an ontology overview; number *2* expands a node where the *BodyOfWater* is represented; and number *3* expands the concepts and the "nearest" related concepts. These "nearest" concepts include those related to "water quality indicators" and *Biome*, which are essential to answer the question presented above. Some key concepts necessary to answer the question are: (1) *Measures* (*QuantitativePhysicalMeasure*, *QualitativePhysicalMeasure*, *QualitativeChemicalMeasure*, *QualitativeBiologicalMeasure*) and Statistical Definitions (including samples and series), which define what we mean by "*… quality of …*"; (2) Child classes of *BodyOfWater* (e.g., a Lake) and water resource concepts, which define what we mean by "*… water …*"; (3) Biome, which defines what we mean by "… of the Atlantic forest biome …", that, in turn, is linked to agricultural crops production, which defines what we mean by "*… soybean production …*"; and (4) *Time* and *Space* definitions to contextualize the information, which define what we mean by "… regions during the last year …". In addition, the concepts are also linked to *InformationAgent* classes that identify the resources related to the ontology concepts in the data sources and systems.

5. Experiencing the Ontology in an information recovery scenario

In this section we describe an information recovery scenario, which involved dealing with concepts from the reused knowledge sources of multiple domains. The main objectives were to evaluate the ontology in terms of representativeness of the (cross domain) context and consistence, to analyze the potential and expressiveness of the model for indexing text based resources from multidisciplinary domains, and to investigate possible benefits of using the ontology in a search scenario when compared to syntactic search mechanisms.

Section 5.1 describes how the experiment was conducted, including how the data was collected and indexed, as well as how the results were evaluated by a domain specialist. Section 5.2 presents the results of the study. Section 5.3 discusses the strengths and limitations based on the analysis of the results.

5.1. Context and methods of the study

The study was performed on the agricultural research database of Embrapa[14] (BDPA—*Base de Dados da Pesquisa Agropecuária*), which included 813,937 documents from various sources (49 collections), and different formats (e.g., DVDs, pdfs, pages) and content (e.g., technical reports, papers, proceedings, books, news). With the aim of eliciting interesting scenarios, during the 2013 workshop of the AgroHidro network, we carried out an activity where the ontology and information recovery scenarios were evaluated and discussed by four domain specialists. The main criterion for defining the scenario was its expressiveness to represent domain concepts and real problems in the Brazilian context.

The selected scenario concerns sugarcane production in the São Paulo state. A key question was elaborated to characterize and summarize the main terms used in the scenario: "How is the expansion of sugarcane in the state of São Paulo affecting water quality?" This question expresses the following major aspects of the model that were tested:

- The question represents an important and current issue that must be considered by the AgroHidro researchers. Sugarcane production has largely expanded in recent years. The São Paulo state (southeastern Brazil) is now experiencing the most severe drought in its history, and the impacts of sugar cane production on water resources is also uncertain.

---

[14] http://www.bdpa.cnptia.embrapa.br/.

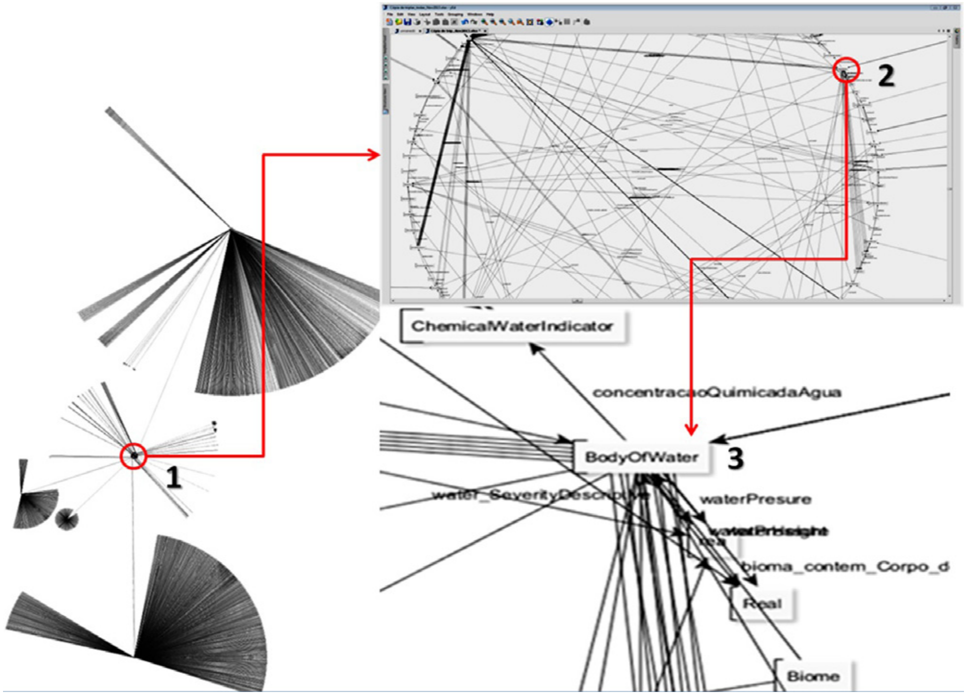

Fig. 4. The BodyOfWater concept in the OntoAgroHidro.

- The question is of interest from both economic and environmental perspectives. There is also content (reports and papers) produced by researchers from various institutions in the São Paulo state.
- The question explores the spatial dimension by delimiting the search to a specific state, as well as dealing with the concept of "expansion".
- The question explores agricultural terminology by specifying the type of production, which includes different denominations (e.g., *Saccharum spp*).
- The question explores the aspects of water quantity and quality.

After defining the scenario's principle question, the study was performed in four steps: (1) collection of necessary documents; (2) categorization of the collected data using the OntoAgroHidro; (3) execution of queries; and (4) analysis of the results with a domain specialist. Due to resource and time limitations, we filtered and collected the data to be categorized by the ontology (instead of categorizing the 813,937 documents from the entire database) using relaxed queries on the syntactic query interface. All texts containing the words "sugarcane" and "water", as well as their synonyms, were considered in the study.

In the second step, 393 documents were categorized using a manual-assisted process. Firstly, the keywords were automatically highlighted, and passages from the texts were extracted. A non-domain specialist manually selected classes and properties of the OntoAgroHidro that were related to the document. When necessary, new instances for related concepts were created. After that, an instance of the *InformationResource* class was then created and linked to the instances of the selected classes and properties, in addition to being linked to the URI's of the descriptors and documents.

The syntactic queries (third step) were performed directly in the search box of the web interface of the BDPA. Firstly, we started querying by using all of the keywords in the question (i.e., *expansion, sugarcane, state, São Paulo, affecting, water, quality*). We then removed one keyword, and executed the queries with all of the seven combinations. Another word from the query with the best results was then removed and the six possibilities were tested. We did this successively until four words remained, i.e., a total of 22 query variations were performed. Queries with less than four words were not considered, as precision began to deteriorate. The semantic query was performed using the SPARQL Query tab of Protégé 4.3. It was beyond the scope of this work to design an easy-to-use interface for querying the ontology, as well as coping with performance issues.

In the last step, the results of all the syntactic and semantic queries were randomly unified in a single list (removing duplicated entries). A domain specialist read and analyzed each document and marked the results as relevant or not-relevant. He also wrote a short explanation for each decision. Additionally, we analyzed results from Google through a search using the seven keywords (in Portuguese). The Google search returned 386,000 results. However, this comparison is limited by the fact that the search spaces are different (i.e., google indexes trillions of documents), the document types are different (e.g., Google includes web pages in general), and we are not able to determine relevant results that are missing from the Google results.

The results were collected and tabulated, and the precision, recall and F-Measure of each result was calculated (except for the results from the Google search). Finally, the results were individually analyzed to explore the potential and limitations of the ontology.

5.2. Results of an information recovery scenario

Table 1 summarizes the quantitative results as well as information retrieval measures for the studied scenario. The first column identifies the queries analyzed as follows: *Collect Queries* includes the results of the queries used to collect the documents for semantic indexing, including all of the relevant and irrelevant documents (categorized in the ontology), OntoAgroHidro contains the analysis of the results from querying the ontology; *Syntactic Query 1* was performed with all (seven) keywords; *Syntactic Query 2* is the query with best F-Measure for six keywords (*expansion, sugarcane, São Paulo, affecting, water, quality*); *Syntactic Query 3* is the query with best F-Measure for five keywords (*expansion,*

Table 1
Comparison of the results measured.

	Quantitative results			Retrieval measures		
	# Results	Relevant	Irrelevant	Precision	Recall	F-Measure
Collect Queries	393	28	365	0.07	1.00	0.13
OntoAgroHidro	29	24	5	0.83	0.86	0.84
Syntactic Query 1	1	1	0	1.00	0.04	0.07
Syntactic Query 2	1	1	0	1.00	0.04	0.07
Syntactic Query 3	7	3	4	0.43	0.11	0.17
Syntactic Query 4	13	5	8	0.38	0.18	0.24
Google 30	30	19	11	0.63	–	–
Google 60	60	33	27	0.55	–	–

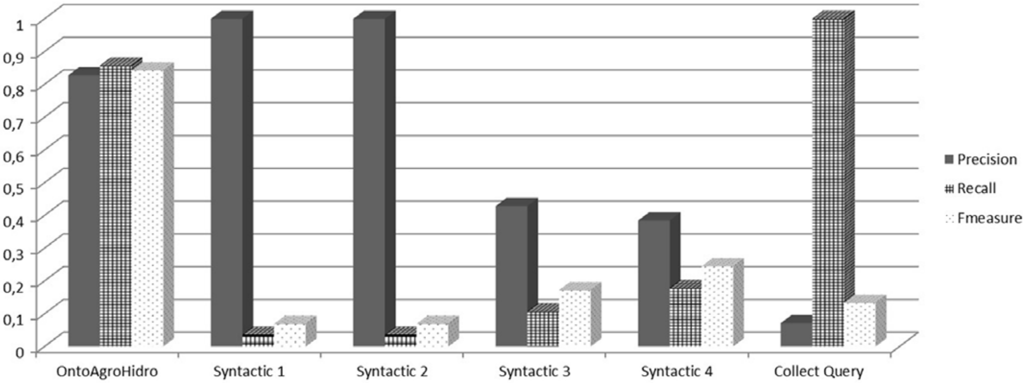

Fig. 5. Recovery measures of the performed queries.

sugarcane, São Paulo, water, quality); *Syntactic Query 4* is the query with the best F-Measure for four keywords (*expansion, sugarcane, São Paulo, water*); *Google 30* contains the analysis of the first thirty results; and *Google 60*contains the analysis of the first sixty results of a *Google* search.

The second, third and fourth columns present the total number of results analyzed, considered relevant and not relevant, respectively. The fifth column presents the precision, which is calculated according to the number of relevant results in the query (i.e., the interception of relevant results with query results) divided by the total number of results returned by the query. The sixth column presents the recall, which is calculated by dividing the relevant results by the total number of results in the dataset (there were 28 relevant results in our dataset). The seventh column presents the F-Measure, that is calculated by the harmonic mean of precision and recall, i.e., $F\text{-}Measure = 2 * (precision * recall)/(precision + recall)$.

As shown in Fig. 5, the OntoAgroHidro query returned the best *F-Measure* (*0.84*) by combining good precision (*0.83*) with good recall (*0.86*) values. The *Syntactic 1* and *Syntactic 2* queries returned only one result each. Both results were considered relevant, and consequently, their precision was *1.0* and their recall was *0.04*. The *Syntactic 3* query returned 7 results with 3 relevant, scoring *0.43* of precision and *0.11* of recall. The *Syntactic 4* query had the best *F-Measure* (*0.24*) of all of the 22 syntactic queries considered in the study. It returned 13 results with 5 relevant, measuring *0.38* of precision and *0.18* of recall.

Additionally, Table 1 presents the *Google* results. The 19 results of the first 30 were considered relevant resulting in *0.63* of precision, and 33 results of first 60 were considered relevant resulting in *0.55* of precision. We were not able to calculate the recall and *F-Measure* of *Google* queries, as is was not possible calculate the entire set considered relevant.

5.3. Discussion on potential and limitations

As mentioned above, we asked the domain specialist to write short statements indicating why he considered each document relevant or not relevant. These statements were used to analyze of the potential and limitations of the OntoAgroHidro. By analyzing each of the results, we can highlight the following strength factors that explain the superior results of OntoAgroHidro when compared to the syntactic scenarios.

Using the ontology, it was possible to identify content related to the São Paulo state that did not explicitly cite São Paulo in the text. São Paulo has 645 municipalities, 15 subdivisions (*mesoregions*) and thousands of body of water. We used a table from Embrapa with 178,000 geo-referenced instances that indentify bodies of water in Brazil. The syntactic search was not able to retrieve results that referenced the names of the bodies of water, municipalities, or subdivisions that belong to the São Paulo state but did not mention São Paulo itself. This aspect considerably increased the recall of the OntoAgroHidro results.

Using the ontology we were also able to disambiguate many terms during the assisted categorization. For example, some of the results recovered by the syntactic search were related to the misidentification of São Paulo as the location of the study, when it was, in fact, the city (or state) of the publication. This aspect considerably increased the precision of the OntoAgroHidro results.

The ontology modeled complex terms such as *affecting, expansion* and *quality*. By using the ontology we could also specify the relationships of these concepts. For example, sugarcane production is the *EnvironmentalTransformationAgent* that affects an *EnvironmentalTransformationObject* during an *Event* of agricultural expansion that results in changes in water quality measures (i.e., *WaterQualityBiologicalMeasure*, *WaterQualityChemicalMeasure* and WaterQualityPhysicalMeasure). The *affecting* keyword, in particular, did not produce relevant results on the syntactic search. When we removed this keyword from the search the F-Measure increased from *0.07* (Syntactic Query 2) to *0.17* (Syntactic Query 3).

The terms can be better contextualized by means of the ontology. For example, the word *water* in the query refers to the *body of water* concept that is affected by *sugarcane* production, instead of the chemical element, or water as a resource for *sugarcane* production. Two documents recovered by the syntactic searches referred to "how the quality of the *sugarcane* is affected by the concentration of the water in the soil".

The OntoAgroHidro recovered five results considered irrelevant by the domain specialist. There were also four results missing from the list. By analyzing the domain specialist's statements in these cases, we identified major limitations of the ontology as well as limitations of the assisted categorization process used in the study.

The ontology is limited in terms of detailing; the disambiguation capability must be improved by modeling specialized subclasses and new axioms. For example, the phenomenon of expansion and intensification were superficially represented in this version of OntoAgroHidro because the distinction between these concepts was not completely modeled. The ontology can also be expanded by considering the Agrovoc thesaurus.[15] This is an initiative of the Food and Agriculture Organization of the United Nations (FAO) to serve as a controlled vocabulary for agriculture and related areas. The Agrovoc is now a SKOS scheme published as a Linked Open Data (LOD) [28,29]. It is available in 21 languages with approximately 41,000 terms in the English language (representing 32 thousands concepts). The evolution of Agrovoc as an ontology is also being studied by the agricultural community [30]. The OntoAgroHidro concepts can be linked to Agrovoc terms (by associating concept identifiers to terms); however other types of reuse should also be studied with the objective of providing a more compressive model of the AgroHidro domains.

It is necessary to improve the categorization process in order to be more productive. The assisted process can be applied to categorize hundreds (or even thousands) of documents, however it is not scalable for millions of documents. Resources from simple assistances (e.g., highlighting related terms in the text) to the complex natural language processing algorithms (e.g., suggesting ready to use categorization) should be considered in order to improve this process. The categorization process should also be improved to increase precision. Categorization was performed by a non-domain specialist, resulting in erroneous classifications (3 of the 5 results were considered irrelevant by the domain specialist). However, it is usually unfeasible to have a specialist of each domain performing this task. Thus, the process could be assisted by tools that check for consistency and alert for possible alternatives. The construction of such tools, however, requires future research.

6. Conclusions and future work

Agricultural practices are fundamental to guarantee food safety and even alternative sources of energy in the near future. However, there is a lack of good quality information about the impacts of agriculture and climatic changes on water resources. The advance of web collaborative systems and semantic technologies has resulted in new possibilities for the research of complex interdisciplinary topics by means of scientific multidisciplinary networks. However, this kind of computational support requires complex solutions that require research in related areas such as knowledge representation, semantic interoperability and information recovery.

This paper presented how the OntoAgroHidro deals with semantic interoperability issues in the development of a network to share and recover information on the impact of climatic changes and agriculture on water resources. The engineers proposed the creation of a core ontology model, the reuse of existing representations, and the modeling of ontological constructions that integrates cross domain concepts. The article also includes an analysis of the data sources and systems to be incorporated into the architecture, as well as the integration of concepts from the reused representations by means of key questions and scenarios. Visualization techniques were used to explore these scenarios with the domain specialists. A study in an information recovery scenario was conducted to analyze the potential and limitations of the ontology.

The current version of OntoAgroHidro models the core domain concepts, as well as details important aspects including water quality, location and time concepts. However many concepts from the field of agriculture should be considered in order to provide a more precise and "complete" model of the problem. Reuse of agricultural thesauri (*e.g.*, Agrovoc) and other terminological resources, for example, could result in an improved and more detailed model. The OntoAgroHidro is also quickly increasing in size, which creates problems in terms of management and performance. Solutions that include the use of databases for large numbers of instances (triple store) as well as the use of federated ontology specification must be studied. Although it worked well in the study of document categorization and retrieval, as well as for conceptual discussions during the AgroHidro workshops, we have not used the OntoAgroHidro in an integrated system. Practical aspects of the integration of computational systems, detailing, reuse, and scalability will be considered in the next steps of this work.

Acknowledgments

This work is partially funded by the Brazilian Agricultural Research Corporation (Embrapa) (no. 01.12.01.001.02.04) from the Ministry of Agriculture, Livestock and Food Supply (MAPA). The authors would also like to thank the valuable contributions from participants of the AgroHidro project.

References

[1] United Nations, United Nations Framework Convention on Climate Change, 1992. http://unfccc.int/files/essential_background/background_publications_htmlpdf/application/pdf/conveng.pdf, Last visited 27th March 2015.

[2] Intergovernmental Panel on Climate Change, Climate Change 2007: Synthesis Report, 2007. http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr.pdf, Last visited 27th March 2015.

[3] N. Alexandratos, J. Bruinsma, World Agriculture Towards 2030/2050: the 2012 revision, in: ESA Working Paper No. 12-03. Rome. Food and Agriculture Organization of the United Nations, 2012.

[4] PBMC, Executive summary: impacts, vulnerability and adaptation to climate change, in: E.D. Assad, A.R. Magalhães (Eds.), Contribution of Working Group 2 to the First Assessment Report of the PBMC, 2013, COPPE. Universidade Federal do Rio de Janeiro, 2013, p. 28.

[5] E.D. Assad, H.S. Pinto, Global warming and the future scenarios Brazilian agriculture, Embrapa, 2008, p. 84. http://www.agritempo.gov.br/climaeagricultura/download.html, Last visited 27th March 2015.

[6] R. Bonacin, O.F. Nabuco, I. Pierozzi Júnior, Conceptualizing the impacts of agriculture on water resources, in: Proceedings of the Fifth International Conference on Management of Emergent Digital EcoSystems, ACM Press, New York, 2013, pp. 262–269.

[7] D. Schrag, Extreme weather and climate change, Environ. Harv. 5 (1) (2013).

[8] V. Vescoukis, N. Doulamis, S. Karagiorgou, A service oriented architecture for decision support systems in environmental crisis management, Future Gener. Comput. Syst. 28 (2012) 593–604.

[9] C. Pettit, S. Williams, I. Bishop, J.P. Aurambout, A.B.M. Russel, A. Michael, S. Sharmab, D. Hunter, P.C. Chan, C. Enticott, A. Borda, D. Abramson, Building an ecoinformatics platform to support climate change adaptation in Victoria, Future Gener. Comput. Syst. 29 (2013) 624–640.

[10] OWL Working Group, OWL 2 Web Ontology Language Document Overview (Second Edition), World Wide Web Consortium (W3C) Recommendation, December 2012.

[11] L. Ding, J. Michaelis, J. McCusker, D.L. McGuinness, Linked provenance data: a semantic Web-based approach to interoperable workflow traces, Future Gener. Comput. Syst. 27 (2011) 797–805.

[12] K. Jaroensutasinee, M. Jaroensutasinee, S. Bainbridge, T. Fountain, S. Holbrook, M. Nekrasov, CREON—integrating disparate sources of remote coral reef sensor data, in: Proceedings of the 12th International Coral Reef Symposium, Cairns, Australia, 9–13 July 2012.

[13] E.W. Patton, P. Seyed, P. Wang, L. Fu, F.J. Dein, R.S. Bristol, D.L. McGuinness, SemantEco: a semantically powered modular architecture for integrating distributed environmental and ecological data, Future Gener. Comput. Syst. 36 (2014) 430–444.

---

[15] http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus.

[14] X. Ma, J.G. Zheng, J. Goldstein, S. Zednik, L. Fu, B. Duggan, S. Aulenbach, P. West, C. Tilmes, P. Fox, Ontology engineering in provenance enablement for the National Climate Assessment, Environ. Model. Softw. 61 (2014) 191–205.
[15] D.G. Tarboton, D. Maidment, I. Zaslavsky, D. Ames, J. Goodall, R.P. Hooper, J. Horsburgh, D. Valentine, T. Whiteaker, K. Schreuders, Data interoperability in the hydrologic sciences, the cuahsi hydrologic information system, in: Proceedings of the Environmental Information Management Conference, 2011, pp. 132–137.
[16] Open Geospatial Consortium, OGC WaterML 2.0. OGC® WaterML 2.0: Part 1-Time series, 2012, pp. 137. https://portal.opengeospatial.org/files/?artifact_id=48531. Last visited 27th March 2015.
[17] R.G. Raskin, M.J. Pan, Knowledge representation in the semantic web for Earth and environmental terminology (SWEET), Comput. Geosci. 31 (9) (2005) 1119–1125.
[18] H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, S. Hübner, Ontology-based integration of information—a survey of existing approaches, in: IJCAI-01 Workshop: Ontologies and Information, 2001, pp. 108–117.
[19] A. Kiryakov, B. Popov, I. Terziev, D. Manov, D. Ognyanoff, Semantic annotation, indexing, and retrieval, J. Web Semant. Sci. Serv. Agents World Wide Web 2 (1) (2004) 49–79.
[20] H. Dong, W.A. Perth, F.K. Hussain, E. Chang, A survey in semantic search technologies, in: 2nd IEEE International Conference on Digital Ecosystems and Technologies, 2008, pp. 403–408.
[21] H. Hoang, A. Tjoa, The state of the art of ontology-based query systems: a comparison of existing approaches, in: Proceedings of the IEEE International Conference on Computing and Informatics, ICOCI06, 2006, Paper-Nr. 99.
[22] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, Sci. Am. 284 (5) (2001) 29–37.
[23] Y.F. Li, G. Kennedy, F. Ngoran, P. Wud, J. Hunter, An ontology-centric architecture for extensible scientific data management systems, Future Gener. Comput. Syst. 29 (2013) 641–653.
[24] A.J. Cañas, G. Hill, R. Carff, N. Suri, J. Lott, T. Eskridge, G. Gómez, M. Arroyo, R. Carvajal, CmapTools: a knowledge modeling and sharing environment, in: A.J. Cañas, J.D. Novak, F.M. González (Eds.), Concept Maps: Theory, Methodology, Technology, Proceedings of the First International Conference on Concept Mapping, Universidad Pública de Navarra, Pamplona, Spain, 2004, pp. 125–133.
[25] J.D. Novak, A.J. Cañas, The Theory Underlying Concept Maps and How to Construct and Use Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition (2008) available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf Last visited 27th March 2015.
[26] E. Sirin, B. Parsia, B. Grau, A. Kalyanpur, Y. Katz, Pellet: a practical OWL-DL reasoner, J. Web Semant. Sci. Serv. Agents World Wide Web 5 (2) (2007) 51–53.
[27] Comissão Nacional de Cartografia, PMGB—Perfil de Metadados Geoespaciais do Brasil, 2009, CONCAR, p. 194. http://www.concar.ibge.gov.br/arquivo/Perfil_MGB_Final_v1_homologado.pdf. Last visited December 2014.
[28] C. Caracciolo, A. Stellato, S. Rajbahndari, A. Morshed, G. Johannsen, J. Keizer, Y. Jacques, Thesaurus maintenance, alignment and publication as linked data, Int. J. Metadata Semant. Ontol. (IJMSO) 7 (1) (2012) 65–75.
[29] C. Caracciolo, A. Stellato, A. Morshed, G. Johannsen, S. Rajbhandari, Y. Jaques, J. Keizer, The AGROVOC linked dataset, Semant. Web 4 (3) (2013) 341–348.
[30] A.C. Liang, B. Lauser, M. Sini, J. Keizer, S. Katz, From AGROVOC to the Agricultural Ontology Service/Concept Server: an OWL model for creating ontologies in the agricultural domain, in: Proceedings of International Conference on Dublin Core and Metadata Applications (DC-2006). Manzanillo, Colima, Mexico, 3–6 October 2006.

Rodrigo Bonacin is a Senior Researcher at CTI Renato Archer, Brazil, and Professor at FACCAMP, Brazil. He received a B.Sc. and M.Sc. in Computer Science from UFPR, Brazil, and a Ph.D. in Computer Science from UNICAMP, Brazil, and Postdoctorate by Public Research Centre Henri Tudor, Luxembourg. His research interests include ontologies, human–computer interaction, medical informatics and organizational semiotics.

Olga Fernanda Nabuco is a researcher at Centro de Tecnologia da Informação Renato Archer. She has a Ph.D. degree in Mechanical Engineer from University of Campinas (Brazil) and Post Doc in Semantic Technologies applied to web services at Laboratoire d'Analise et Architeture des Systemes LAAS from CNRS, France. She is actually participating on projects based on ontology engineering defining domain and restrictions to climate changes impacts on water supplies and the other in cyber security. She is also chair of annual workshop IEEEWeb2Touch workshops, now a track of IEEE Wetice Conference. Dr. Nabuco main interests are on ontology engineering, semantic technologies and web applications.

Ivo Pierozzi Junior is Graduate in Biological Sciences (1981), Master degree (1985) and Doctor degree (1989) in Ecology from State University of Campinas (Unicamp). For more than a decade (1994–2009), he was charged of management positions on Research & Development and Science & Technology, including Assistant Director of Research and Development at Embrapa Satellite Monitoring, where had also worked with Operational Acridology and Applied Ecology. Currently is scientific researcher at Embrapa Agricultural Informatics working with Agricultural Terminologies, Knowledge and Information Organization and Representation.