Data-driven ontology evaluation based on competency questions: a study in the agricultural domain

Article · October 2016

CITATIONS
0

READS
67

3 authors:

Webert Araújo
Federal University of Minas Gerais
4 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Gercina Angela de Lima
Federal University of Minas Gerais
79 PUBLICATIONS **144** CITATIONS
SEE PROFILE

Ivo Pierozzi Junior
Embrapa Agricultural Informatics
22 PUBLICATIONS **41** CITATIONS
SEE PROFILE

Data-Driven Ontology Evaluation Based on Competency Questions: a Study in the Agricultural Domain

Webert Araújo
Gercina Lima
Ivo Pierozzi Jr.

1 Introduction

Climate changes and water shortages are real concerns that affect various spheres of our society. Brazil, for example, lived in 2015 one of its major water crisis, as water availability in the country is directly related to climate. Thus, research is needed in various fields of knowledge in order to understand climate changes and its impacts on water resources. The Knowledge Organization's contribution can relate to the modeling of Knowledge Organization Systems (KOS), which are systematic tools aimed at building abstract models of the real world through the representation of domain concepts. These instruments make it possible, among other things, a shared view on a specific area, and facilitate communication between people with different views through common vocabulary.

A KOS that has gained prominence in the Knowledge Organization field is ontology. An ontology can be defined as a shared model of reality, which is formally represented by classes, properties, relationships and axioms. Although ontologies have potential in knowledge representation, they must be tested for the guaranty of quality. The ontology evaluation is one way to ensure the quality of an ontology. Since the purpose of ontology evaluation is to verify the conformity of world model to the world modeled formally.

The guiding objective of this research is to evaluate a domain ontology using a data-driven ontology evaluation. The ontology intends to evaluate in this work was developed by specialists from Embrapa Campinas, and it is named OntoAgroHidro.

2 Background

The OntoAgroHidro was built using Web Ontology Language 2 and the ontology editor Protégé (version 4). The main purpose of the OntoAgroHidro is to represent knowledge about the impacts of climate changes and agriculture on water resources. As pointed out by Bonacin, Nabucco and Pierozzi Junior (2015), it is a quite ambitious scenario to model, insofar as it covers the domain particularities of climate changes associated with agriculture and hydrology.

The research is characterized as exploratory, qualitative and applied, starting with the exploration of the relevant topics to the research in the Information Science and Computer Science literatures. The theoretical and methodological foundation permeate

concepts on topics such as organization and knowledge representation, definitions of term and concept, ontologies and competency questions.

The literature review found that there were several proposals for ontology evaluation. These proposals can be classified into four categories:

1) those based on comparing the ontology to a gold standard (which may be an ontology itself), e.g. Maedche and Staab (2002); Gangemi et al. (2006);
2) those based on using the ontology in an application or task and evaluating the results, e. g. Porzel and Malaka (2004); Fernández, Cantador and Castells (2006);
3) those involving comparisons of the ontology against a data source (a collection of documents, for example) about the domain to be covered by the ontology, e.g. Brewster et al. (2004); Hlomani and Stacey (2013);
4) those where the evaluation is done by humans who try to assess how well the ontology meets a set of predefined criteria, standards, requirements, etc. e.g. Lozano-Tello and Gómez-Pérez (2004); Almeida (2009).

3 Methods

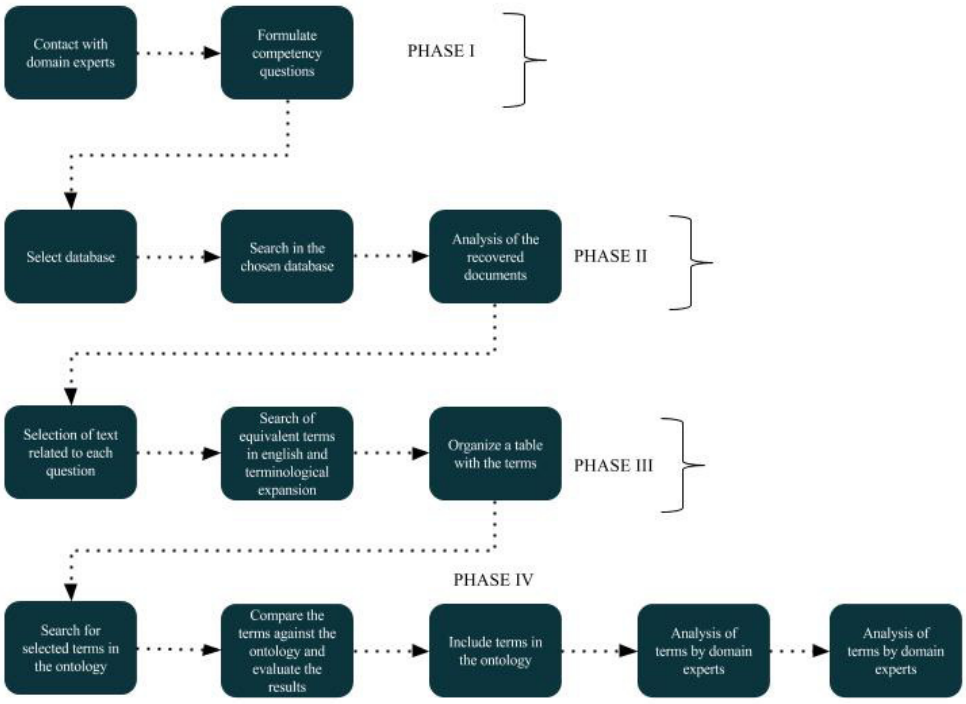
The research methodology is based on a proposal for a data-driven ontology evaluation developed by Brewster et al (2004). The proposal was adapted and has four steps: 1) definition of competency questions[1]; 2) selection of the collection of documents that will compose the data source for the ontology evaluation; 3) selection of terms related to the concepts of the competency questions; 4) evaluation and mapping the terms in the OntoAgroHidro.

From three competency questions formulated by a domain expert, we extracted the key concepts from each competency question and used the concepts to search for documents in a domain database with the purpose of identifying documents related to each concept. Then we extracted from the documents, terms associated with the concepts from the competency questions and other terms from the AGROVOC and NAL thesaurus and created a list with all terms. Finally, we compared the ontology concepts to the list of terms with the purpose of identifying how much from the list the OntoAgroHidro represents. Figure 1 shows a framework with each one of the phases performed.

---

[1] *Question 1:* Does organic crops cause some impact on water quality of our basins?
Question 2: There will be water shortage in water reservoirs for supply to the population of the State of São Paulo, Brazil in 2016?
Question 3: How deforestation nearby draw-wells can affect the water quantity and quality of water from our rivers?

Figure 1 - Phases for the ontology evaluation



4 Results and Discussion

Based on three competency questions which led to 10 concepts we recovered 60 documents from a specific database in the agricultural domain (BDP@[2]). We extracted 130 terms from the 60 documents and the AGROVOC and NAL thesaurus. The 130 terms were compared against the OntoAgroHidro.

The evaluation of the ontology against the corpus found out that the OntoAgroHidro represents 68% of the concepts related to competency questions. Detailing the result, it was discovered that the ontology represents 78% of the concepts related to the competency question number 1, 56% of competency question number 2 and 42% of the competency question number 3. Table 1 presents in more detail the data, which shows the number of concepts represented by OntoAgroHidro in each competency question.

[2] https://www.bdpa.cnptia.embrapa.br/consulta/

Table 1 – OntoAgroHidro Representativeness

Questions	*Nº of terms*	*Nº of concepts*[3]	*Nº of concepts represented by OntoAgroHidro*	*OntoAgroHidro representativeness*
Question 1	66	50	39	78%
Question 2	40	25	14	56%
Question 3	24	7	3	42,85%
Total	130	82	56	68,29%

In a qualitative analysis, it was observed that the ontology is representative with regard to concepts related to the subdomain "water resources". We also find out that the OntoAgroHidro must be worked on to represent relevant concepts from the domain in a more specific way. Because most of the concepts represented by the ontology are due to the lack of specificity in the field of representation.

5 Conclusions

The results obtained in this research, so far, have been helpful because it was possible to achieve the proposed objective. We could implement a proposal for ontology evaluation based on a text corpus by adaptations to the context of research and to the ontology evaluated in this study.

The methodology presented in this work differs from most existing methods in the literature; it makes use of an approach that uses competency questions in combination with a data source. The results achieved are considered sufficient for identification of positives and negatives aspects of the OntoAgroHidro. Although it is believed that an expansion of the study scope can prove useful to detect other characteristics of the evaluated ontology and thus suggest improvements.

References

Bonacin, R., Nabuco, O. F., & Junior, I. P. (2015). Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery. *Future Generation Computer Systems.*

Maedche, A., Staab, S. (2002). Measuring similarity between ontologies. *Proceedings of CIKM*. vol. 2473.

[3] The number of concepts is based on an intellectual analysis of how many concepts the list of terms represents.

Gangemi, A. et al. (2006). A metaontology based framework for ontology evaluation and selection . *Proceedings of 4th International Workshop on Evaluation of Ontologies for the Web*, Edinburgh, UK.

Porzel, R., Malaka, R. (2004). A task-based approach for ontology evaluation. *ECAI Workshop Ont. Learning and Population.*

Fernández, M., Cantador, I., Castells, P. (2006). A tool for collaborative ontology reuse and evaluation . *Proceedings of 4th International Workshop on Evaluation of Ontologies for the Web*, Edinburgh, UK.

C. Brewster, et al. (2004). Data-driven ontology evaluation. *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal.

Hlomani, H., Stacey, A. D. (2013). Contributing evidence to data-driven ontology evaluation: Workflow ontologies perspective. *Proceedings of the 5th International Conference on Knowledge Engineering and Ontology Development*, Vilamoura, Portugal.

Lozano-Tello, A., Gómez-Pérez, A. (2004). Ontometric: A method to choose the appropriate ontology. *J. Datab. Mgmt*., 15(2), 1–18.

Almeida, M.B. A proposal to evaluate ontology content. *Applied Ontology*, Birmingham, 4(34), p. 245-265.