Ontologies and Domain Specific Language to Support the Decision Making in Agricultural Sustainability, SustenAgro case.

John Garavito, Dilvan Moreira, Katia Regina, and Ivo Pierozzi

Institute of Mathematical and Computer Sciences , São Paulo University,
Avenida Trabalhador São-carlense, 400 São Carlos - SP, Brazil

Abstract. Agricultural systems have the need of to measure their sustainability, an approach to meet this goal is the use of Indicators of Sustainable Development (ISDs), which was used to define the SustenAgro method, which purpose is provide sustainability assessment in sugar-cane production systems at the center-south of Brazil. This paper we present the representation of this method trough of two ontologies, the sugarcane ISDs ontology and the decision support ontology which represent the domain knowledge and an Domain Specific Language (DSL) that manages the knowledge, these elements support the software tool entitled SustenAgro presented in this paper and makes it capable to work with Semantic Web architecture.

Keywords: Indicators of Sustainable Development, Sustainability Assesment, Decision Making, Sustainable Ontology, User Interface Ontology, Domain Specific Language

1 Introduction

The Sustenagro Project was led and executed by Embrapa Environment, and generated the knowledge base and the indicators which support the assessment process and the development of SustenAgro Software, the objective of this paper is presents the semantic web based solution for support the sustainability assessment process in sugarcane and represent the agriculture sustainability knowledge in a computable format.

Indicators represent the critic information about the system, which is possible to take immediate problems or prevent future problems, the base of this paper was an research that identified and defined the indicators of sustainability[3], these indicators were used as conceptual raw material.

We utilize an agile method to define the ontology, we adapt this method to facilitate this work with a team. We made frequent meetings to adapt the knowledge and we represent it in a conceptual map. This strategy allows define the main concepts by experts without knowledge about the tools to create ontologies and to facilitate the communication.

These ontologies are part of the SustenAgro Project, that integrates the development of a computational system based on semantic web technologies whose objective is evaluate the sustainability in production of sugarcane in center-south of Brazil.

The agricultural systems involve environmental, social and economic aspects, which require a full understanding and inter-disciplinary field, is given complexity of the phenomenon arises the need to maintain a conceptual basis to organize and represent the concepts used by the team of experts and the systems computer.

The agricultural sustainability assessment integrates the environmental, social and economic dimension of an agricultural production system, in the case of SustenAgro project the production system is sugarcane and it was modeled and systemized, allowing to measure and improve the sustainability in the farms and plants related with sugarcane production and others sub products like sugar, bioethanol, yeast, bagasse and electric energy.

The SustenAgro Ontology support concepts that are constantly changing, for example in the process of "sustainability assessment" the indicators and indexes are continuously redefined, for this reason a software feature the flexibility.

The hypothesis that we validated was if SustenAgro Ontology will support the conceptual and technologically representation the domain that constantly changing, the methodology for evaluating sustainability and a software system information retrieval.

The definition of an sustainability assessment method in agriculture is a

O fornecimento de um metodo de avaliação da sustenabilidade em agricultura is a latent requirement. Whereby researchers of Environmental Embrapa developed a sustainability assessment method entiled SustenAgro Method focused on a culture and particular region of Brazil, that allows you to integrate knowledge related to sustainability, provide metrics and generate recomendations to serve strategies to improve sustainability, besides enabling the information base for the formulation of public policies.

2 The problem.

O problema abordado nesta pesquisa foi o design de uma arquitetura software que permita se adaptar e se reconfigurar às mudanças do conhecimento em sustentabilidade no sistema produtivo de cana-de-açúcar, suportando a reconfiguração dos modelos de dados e das visualizações deles por meio de interfaces gráficas.

Este problema foi identificado em varias ferramentas software da Embrapa Meio Ambiente que precisam ser reconfiguradas às mudanças do domínio de conhecimento, no gerenciamento e na apresentação da informação.

Uma característica essencial deste problema é que a informação é complexa e muda continuamente.

3 Architecture.

Tendo em conta a complexidade dos dados do sistema software, foi necessário usar um formato de representação de conhecimento que forneça uma estrutura bem definida e que seja adaptável às mudanças do domínio, por esta razão foi escolhido o formato de ontologias da web semântica, que permite definir, organizar, relacionar e inferir conhecimento.

As ontologias fornecem suporte na representação e organização de conhecimento, mas na definição de procedimentos e configurações especificas de um sistema carece de flexibilidade, pelo que procurou-se uma alternativa para complementar as ontologias; identificou-se que uma DSL, fornece o suporte na definição de processos, de configurações e de logica computacional.

sistema software abrange tecnologias da web semântica que permitem representar o conhecimento de uma forma estruturada por meio de ontologias da web semântica.

4 Methodology.

Due to the diversity and the changing nature of the indicators, the construction of a methodology for the evaluation sustainability and the requirement to have a massive system of data collection, was necessary to analyze the technological possibilities to provide an architecture and information retrieval system to deal with said requirements. It was decided to implement an information retrieval system based on triplestore, which needs the development of a titled ontology.

The development of ontology is responsive type by techniques of rapid prototyping of coverage and increased complexity, starting with the most relevant components of the model to the experts and incorporating each one of the other components by validations, this method was cyclical obtaining in each cycle ontology prototypes.

Among them are the conceptual maps that allow a focused communication in the field of specialist, in order to represent knowledge of a visual way and we also have the computational models in semantic web formats that allow communication with experts in modeling of knowledge.

After having the conceptual model well-defined, modelers represented this model in OWL-DL standard and popular data storage system, after that the expert built questions that were asked in the system, and it generated the expected results, then follows a validation phase, integration and adjustment, which ended with a reliable prototype that represents an ontology sector, this process will be repeated until you have all sectors of interest and the required integrity.

The methodology includes the following steps that will represent the domain knowledge, the process corresponds to a cyclical methodology, which will be added sectors according to the maturation of the terms and the need of the information contained:

1. Definition of entities
2. Definition of ratings

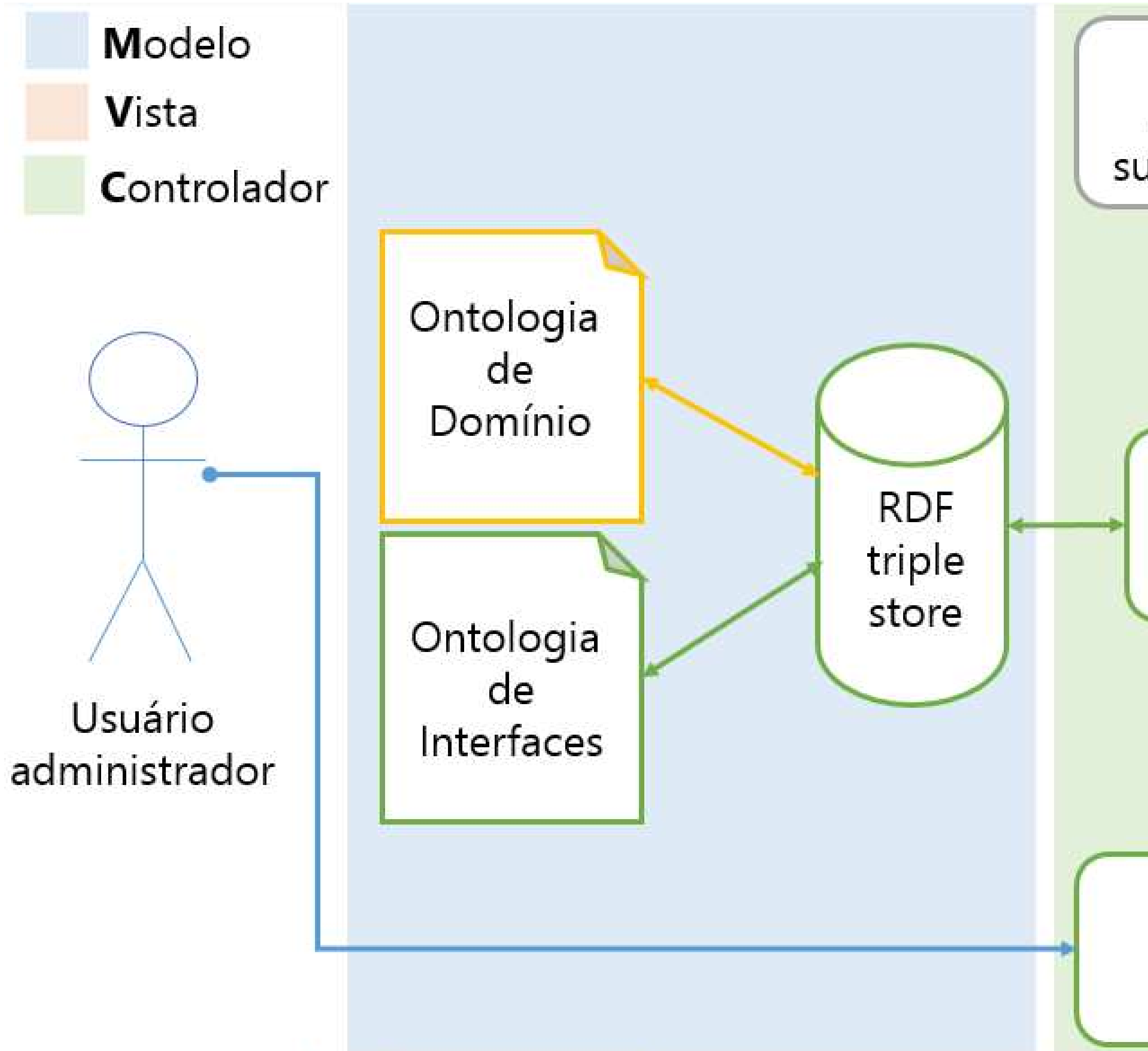

Fig. 1. Arquitetura do sistema

3. Definition of semantic relations
4. Definition the rules and axioms
5. Implementation in OWL
6. Instantiation of individuals
7. Construction of questions
8. Validation through consultation
9. Adjustments and integration

This process is not necessarily linear, as moderators can return to the previous steps as necessary..

The figure 1 represents how is the methodology and how we work cyclic and circular with it.

The modules that were addressed in the modeling are, in order of development:

1. Module attributes and data from production plants of sugarcane
2. Module assessment of sustainability indicators in sugarcane systems
3. Module representation of sustainability assessment methodology
4. Module georeferencing connection with the supply of natural resources data.

5 Development and Tests

5.1 Overview

SustenAgro Ontology represents the critical concepts in the evaluation of sustainability in the production of sugar cane system in the state of Sao Paulo, this ontology was developed by a team that includes specialists in the area of sustainability in agriculture and experts in semantic web technologies, and tries to represent the system by means of sustainability indicators, which allow adding, quantify and simplify information about complex phenomenon so that trends become more apparent and significant in order to improve the process of understanding and communication[?]. According to [?]] the best indicators are those that simplify relevant information, making them clearer phenomena.

Indicators of SustenAgro Ontology, are classified in three dimensions standards of sustainability: environmental, social and economic[1];]; these dimensions are subdivided into smaller compartments called guiding attributes for production systems that bring together related identifiers, for example the environmental dimension has the soil properties, water and climate.

The indicators are composed of elements called indicator components that initialize the status indicator, also have link with the concept of sustainability threshold for assessing the value of an indicator if is instantiated more sustainable or less sustainable in a given production system.

In the conceptual map of Figure 2 can be seen a schematic of the main existing concepts in the ontology developed and how the indicator element is the center of this model:

Indicators of Ontology have a basic format and they all are following describes the subject:

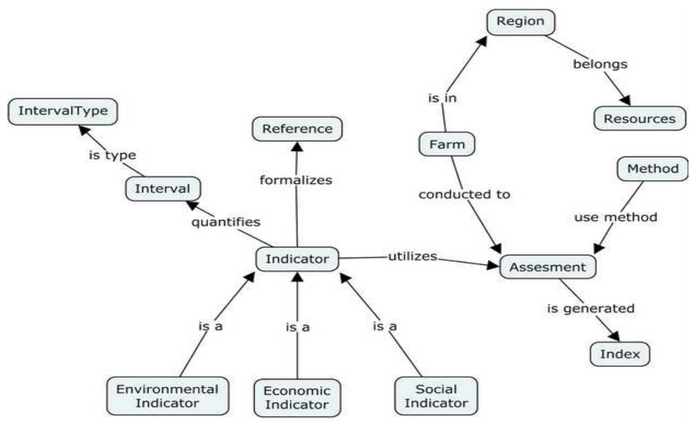

Fig. 2. Mapa conceitual da metodologia de desenvolvimento da ontologia

Title: Important be quite explicit to facilitate understanding and need, defining clearly the variables involved.

Description: It should be clear in quantitative / semi-quantitative terms (There are many dimensions of the indicators, ex).

Data source: Must be where in the literature, this indicator was taken.

Sustainability threshold: the threshold is a minimum and maximum acceptable point in the range of sustainability, which will tell you whether a certain indicator according to the data are more sustainable or less sustainable.

Management measure - recommendation: for some indicators are suggested management recommendations to improve the sustainability, if this indicator is less sustainable, management measures indicate what factors must be analyzed to improve sustainability.

Rationale: are added technical justifications (with references) to allow a better understanding of the information that the user must lift, giving information than finger and because it is included.

The instances of the indicators of ontology will be used by the evaluation methodology that will generate sustainability indexes, which will represent the sustainability of the agricultural system evaluated.

According to this methodology, has developed a conceptual map that represents the concepts and domain relationships and allows communication with the experts:

In figure 3 it can be seen Dimension element and the three subelements, it's a base of concepts the dimensions whose each indicator are allocated.

Each dimension was divided into attributes and indicators such as presents itself in the following figures:

Conceptual map of the environmental dimension in Figure 4this dimension is about weather and soil attributes and is important to production system:

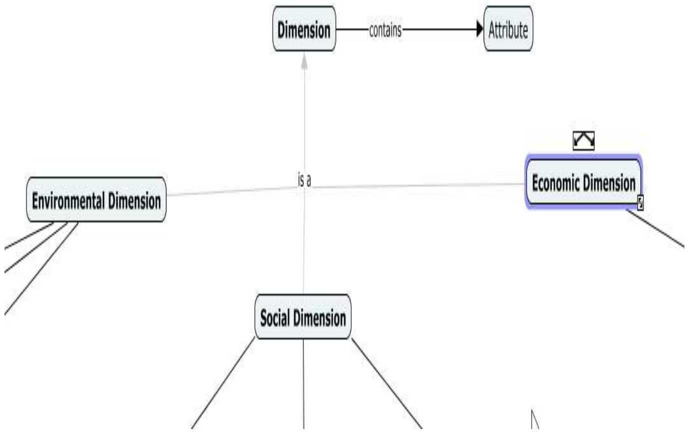

Fig. 3. Mapa conceitual das Dimensões do sistema

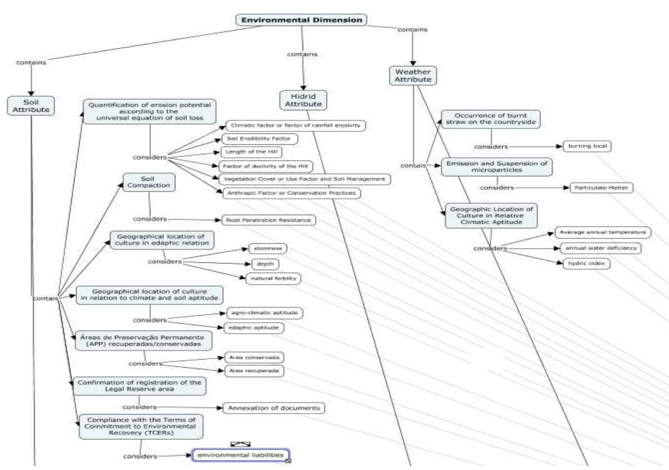

Fig. 4. Mapa conceitual da Dimensão Ambiental

Conceptual map of the social dimension in Figure 5, in this dimension the focus is in how the society Works and be with the production system:

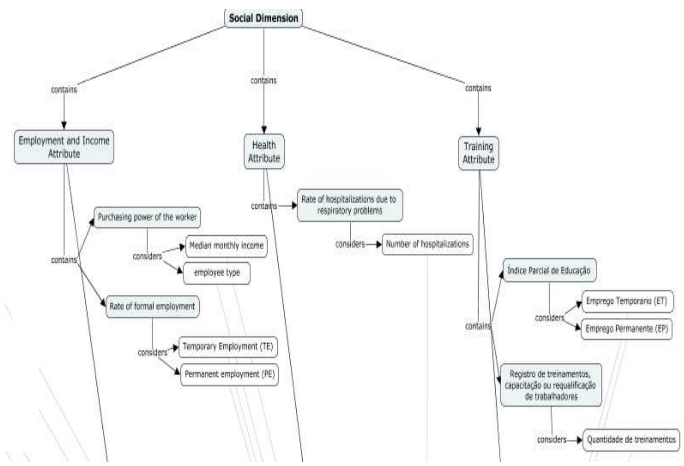

Fig. 5. Mapa conceitual da Dimensão Social

Conceptual map of the economic dimension in Figure 6, is important to know that the production system needs to be sustainable in three dimensions and not only in one dimension, the economic dimension sometimes is most analyzed but is not the most important:

Each of these dimensions are linked through the fundamental concepts of ontologies that are represented in the figure 7.

These elements can generate the sustainability indices that are the indices for each dimension and attribute.

After modeling the phenomenon through the conceptual map, became a translation to OWL-DL standard to achieve its interpretation by computers, the following images show the model made by Protégé tool.

Figure 8 represents the dimension class in Protege:

Figure 9 represents the attribute class in Protege:

Figure 10 represents the indicator class in Protege:

Figure 11 represents the indicator componets class in Protege:

Figure 12 represents the base elements of ontology in Protege:

5.2 Re-use of ontologies

AOntologies are formal representations of knowledge that can be reused according to the needs of each ontological development in the case of SustenAgro Ontology have consulted sources of ontologies in the agricultural sector and on the assessment of sustainability, and were selected two developments the "Agricultural ontology Service "which is a model for defining ontologies for agriculture

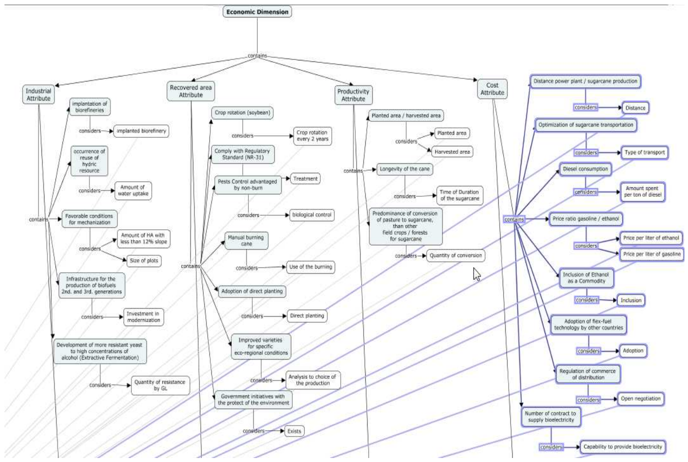

Fig. 6. Mapa conceitual da Dimensão Econômica

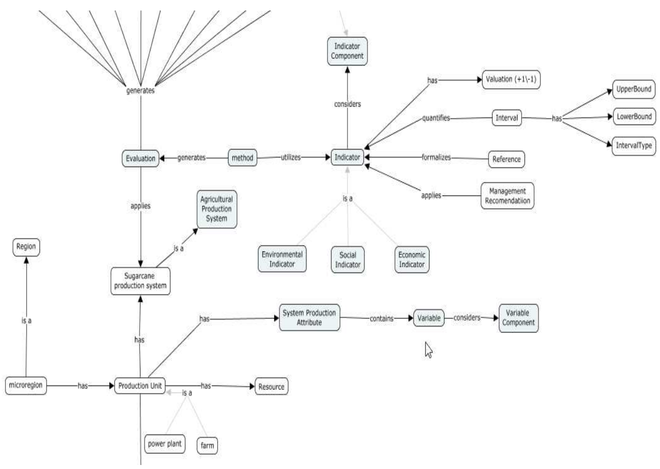

Fig. 7. Mapa conceitual dos conceitos fundamentais

Fig. 8. Classe Dimension

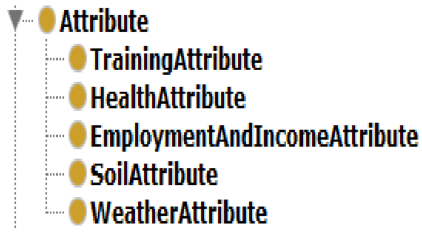

Fig. 9. Classe Attribute

Fig. 10. Classe Indicator

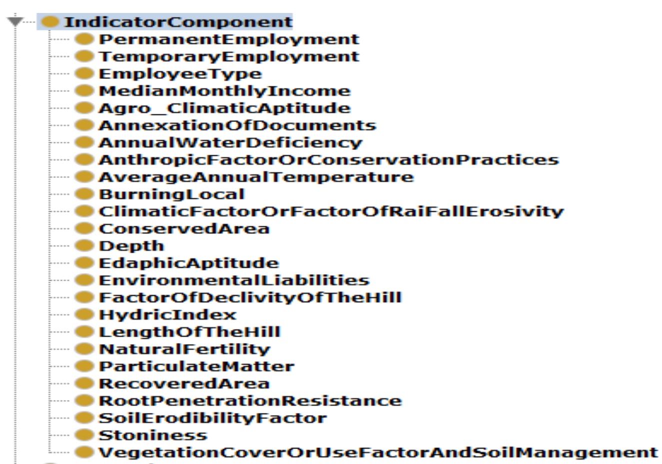

Fig. 11. Classe Componentes de Indicator

Fig. 12. Elementos fundamentais da Ontologia

based on AGROVOC which is a major agricultural vocabularies developed by FAO, and the ontology of sustainability evaluation ISD-Economics ontology [**?**]] that provides sustainability concepts, and according to research it was concluded that there is an ontology that integrates the concepts of sustainability assessment in agriculture. These two proposals state of the art will be integrated into SustenAgro Ontology in the following versions of the ontology, since the goal of this first version is the identification and definition of the fundamental concepts of sustainability assessment in the production of cane sugar system.

5.3 Technological aspects

There is a requirement to establish a standard format that define the terms used in this area, so that people and computer systems that will use the model have access to the concepts involved. These concepts are joining the expert knowledge and the knowledge of ontólogo to define a means of communication between stakeholders and computer systems.

To meet the above requirement was decided to implement the SustenAgro Ontology through semantic web technologies that enable implement triplestore information retrieval systems, taking into account these characteristics was selected the Parliament system that supports the SPARQL query language.

6 Related work

Sustainable Development in this context, is the base to serve the present needs without compromise the future generations to supplement their needs[2], taking into account the sustainability dimensions: environmental, social and economic [1]. Due to complexity of this process is necessary to define metrics that quantify the sustainability.

6.1 Agricultural Area

This part describes the pappers that were relevant to the knowledge in agriculture.

It is necessary for a standardized ontology using techniques which make it multilingual so that it can be used without the need for adjustments to researchers from anywhere.

The papper[?] describes a technique that uses the AGROVOC multilingual Theasaurus for conversion into an ontology and also covers the concepts of agriculture that are a focus of the work, so we analyzed the techniques used to adapt to our ontology.

6.2 Sustainability Area

This part describes the pappers that were relevant to the knowledge in the area of sustainability.

For evaluation of sustainability requires working with indicators that measure how each practice is sustainable. In the paper[?] describes an sustainable analysis framework of Amazonas state and is developed an ontology of these indicators that work in conjunction with the framework. The last foundation for this work was extremely important for the development done. Our ontology is based on agricultural system, the focus on sustainability help to amplify the knowledge and how to make it computational.

The ontology must be flexible to some extent to accept changes according to the need of the specialist. The concepts presented by[?] gave an idea of how to make flexible ontology in the area of sustainability so that it can support new concepts and still not miss the formality or allow the ambiguous concepts. In this way, we adapt this concepts and make it in ours.

To describe the sustainability of a clearer and adaptively is important to understand several indicators that are used in several different places.[?] makes a comparison between sustainability indicators going through in most of it, analyzing global metrics and interpreting sustainability among nations, through a clear notion of sustainability to support knowledge used in the ontology.

6.3 Ontologies Area

In this part of the pappers that were relevant to the knowledge in the area of development of ontologies in general are described.

For a good ontology is important that the concepts are well defined and for this and necessary to use techniques to detect semantic conflicts.[?] defines how to assemble senarios in order to detect conflicts or problems that may have been made in the creation of ontology, the techniques described can be considered as a kind of reasoning that detects situations that was not explicit in the ontology. With this and other pappers related to sustainability we can disambiguate and ensure the concepts are consistent and follow the expert needs.

For development of an ontology is important to understand the concepts involved and needs to create an ontology.[?] describes about ontologies in computer science and generalized terms that are needed for development of ontologies, these terms were extremely important for all stages of creation and analysis of ontology. We do patterns in links between classes to try make more consistent the ontology.

6.4 Conceptual basis of ontology

[3] was taken as the basis of knowledge and was chosen by the domain expert. Was taken all the knowledge base that was used as the indicators and the needs of each and still had part of integrated sustainability directly with the part of agriculture geared to the location of the state of São Paulo. It was analyzed with all the parts described in the dissertation by selecting indicators previously validated by a group of experts.

6.5 Related Technologies

In order to eliminate ambiguity in semantic understanding and making mining the relationship between the concepts of knowledge in agriculture, should be combined with the knowledge among heterogeneous databases.

It proposes a Agricultural Knowledge Grid[?] which was built with three layers "Resource Layer", "Semantic Layer" and "User Layer", has been applied to "Semantic Extension on Retrieval", "Knowledge links" and "Experience" to deepen "Agricultural Knowledge Services".

AGROVOC is a controlled vocabulary that covers all areas of interest FAO and consists of 32,000 concepts available in 21 languages, this tool is used by researchers, librarians and information managers to index, retrieve and organize data in agricultural information systems.

[?] is an initiative that serves as a reference in structuring and standardization of agricultural terminology in multiple languages for use systems in agriculture, the purpose of this technology is to achieve more interoperability between agricultural systems.

7 Results

Among the results of this research we can emphasize the development of the conceptual map which established a means of communication between domain experts and specialists of ontologies, thus achieving the implementation of the concepts of ontology in computer language.

This approach was chosen because it better fit into the context where it was being developed. Experts in the field of sustainability had not domain in ontology development tools and so would be costly teach them about it, making the conceptual map a more common means of communication throughout the team and when in the ontology needs experts knowledge they were introduced the necessary tools.

SustenAgro Ontology is the most relevant results in computational terms because it allows represent the expert's knowledge in a standard modeling elements, relations, axioms and rules of the modeled system, this representation is flexible to change that is needed in this project since the domain is still in construction also supports information retrieval technologies.

This ontology was developed using OWL-DL. This was selected because it is a language created to define ontologies on the Web and also for a better understanding of the team that developed. Since there was a need for a data recovery system, it was essential that the ontology could adapt the semantic web technologies to optimize processes and adapt to the reality of sustainability.

In addition to the ontology and the conceptual map the ontology was also adapted into a data recovery system type triplestore, which is the system used in the semantic web. Was chosen this type of system because besides the team's expertise in the technology, sustainability in itself already has some semantic baggage, which introduced into a scope own optimizes the analysis of sustainable processes and can communicate with other data systems, providing a greater range of concepts and analyzes.

This base was made through the triplestore Parliament, this was already pre-configured on a server facilitating own use. The ontology was adapted to it and so tests were made based on data collected by experts to investigate the completeness and consistency of this ontology.

For the evaluation of ontology were made consultations with experts to create them questions (paths), to answer certain concepts to see if this ontology was adjusting the nature of the data and that the very experts expected the ontology.

After this evaluation some adjustments were made to fix some flaws that had and to generally optimize the ontology.

8 Discussion

One of the difficulties in modeling both the concept map as the ontology is to reach a consensus of critical concepts, which are the binding more than one area of expertise, for example in the modeling of the interface elements as the elements between the indicators and the assessment methodology, or between the elements of the plant variables and indicators, this is natural as each expert has a different view of the phenomenon.

Another difficulty is that the semantic web technologies for knowledge representation are under development, which leads to incompatibilities of the various technologies, for example in the development of information retrieval system there are no software components that facilitate communication with the server part frameworks. And to validate this ontology, the experts have some difficult to define how the ontology have to work in some places to make the validation complete and make the ontology more reliable at the concepts.

9 Conclusions

In defining the models must define what the purpose of it, which will guiding each of the discussions that arise when modeling is a complex phenomenon that can be interpreted in different ways by the experts.

As to the technological development is important to note that the semantic web storage technologies provide more functionality beyond the information retrieval, so it is reasonable that the efficiency is lower in this case is necessary to assess the requirements and determine the most coherent architecture , which can be hybrid architecture.

Its important to define how the evaluation techniques will be used, because without a good validation, all of the methods may be invalid if it's not have a validation well defined. *(é importante para definir como serão utilizadas as técnicas de avaliação, porque sem uma boa validação, todos os métodos podem ser inválidos, se não é têm uma validação bem definida.)*

References

1. Johanna Alkan Olsson, Christian Bockstaller, Lee M Stapleton, Frank Ewert, Rob Knapen, Olivier Therond, Ghislain Geniaux, Stéphane Bellon, Teresa Pinto Correira, Nadine Turpin, et al. A goal oriented indicator framework to support integrated assessment of new policies for agri-environmental systems. *Environmental Science & Policy*, 12(5):562–572, 2009.
2. Gro Harlem Brundtland, World Commission on Environment, and Development. *Our common future*, volume 383. Oxford University Press Oxford, 1987.
3. Bruno Oliveira Cardoso. Avaliação da sustentabilidade da cana-de-açúcar no estado de são paulo: uma proposta metodológica e de modelo conceitual. Master's thesis, 2013.