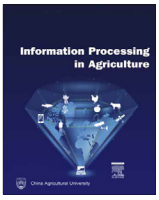
Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 1 (2014) 66–72

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



Semantic representation of proved and disproved statements extracted from scientific papers: Meat science case study

Piotr Kulicki [a], *Robert Trypuz* [a,*], *Rafał Trójczak* [a], *Jerzy Wierzbicki* [b], *Alicja Woźniak* [b]

[a] *The John Paul II Catholic University of Lublin, Al. Racławickie 14, 20-950 Lublin, Poland*
[b] *Polish Beef Association, ul. Kruczkowskiego 3, 00-380 Warszawa, Poland*

ARTICLE INFO

Article history:
Received 27 December 2013
Received in revised form
28 June 2014
Accepted 30 June 2014
Available online 22 July 2014

Keywords:
Ontology
Scientific law
Livestock
Beef

ABSTRACT

The article presents an ontology called "Science" designed for representing proved and disproved statements extracted from scientific papers from the database of the *ProOptiBeef* project. We argue that the proved and disproved statements are the best characterization of the *content* of the papers they are extracted from. We show that ontological representation of proved and disproved statements not only facilitates the access to the content of the papers and increases the precision of a semantical search, but also allows for an automatic generation of new statements from the ones already coded in the ontology.

1. Introduction

It is rather obvious that observations and experiments are prevailing scientific activities and that they are directed toward discovering and formulating scientific laws.[1] A scientific law is a proposition which asserts a constant relationship between things and, more precisely, between the qualities of the objects or between events in which objects participate [10]. For instance, in the *beef production and consumption* domain, which is our scope of interest, scientific laws may concern the relations between an animal's genotype and phenotype or may assert the dependency between a process, quality of beef and consumers' assessment (e.g. aging and tenderness).

Statements expressing scientific laws are crucial elements of scientific papers. We believe that they are the best way to describe and represent content of papers. They contain information important for practitioners and industry. They can be also used in a search as keywords are used.

It is worth noting that scientific papers contain *proved* and *disproved* statements and both types of statements are equally important. The first ones express accepted scientific laws, whereas the other ones are either rejected scientific laws[2] or negatively verified (falsified) hypotheses.

Taking the above into account we claim that an explicit representation of proved and disproved statements can be

* Corresponding author. Tel.: +48 81 445 40 43.
E-mail address: trypuz@kul.pl (R. Trypuz).

[1] This paper is an extended version of [11].

Peer review under the responsibility of China Agricultural University

Production and hosting by Elsevier

[2] They were accepted in the past or are still accepted by some groups due to the lack of knowledge about new results.

http://dx.doi.org/10.1016/j.inpa.2014.06.002

very useful for browsing the content of scientific papers, provided that the statements are represented in a computer-readable manner. For instance, let us imagine that the main result of a paper is expressed in the statement '*Beef slaughtering season has influence on beef fat acids profile*.' If it is extracted from a scientific paper and represented properly, as it is shown in the following sections, it will enable us to find the paper (and similar ones) by queries like:

- 'Find papers which confirm or reject that beef slaughtering season has influence on beef fat acids profile.'
- 'Find papers in which it is asserted that beef slaughtering season has influence on something.'
- 'Find papers in which it is asserted that beef fat acids profile depends on something.'

In order to represent scientific laws it is important to have a representation of the domain which the scientific laws concern. The representation should contain a vocabulary of terms relevant to the domain with a precise specification of their meanings. Ontologies are recognized as valuable tools for obtaining that [5,6]. Thus, our representation of accepted and rejected scientific laws is based on a domain ontology called "OntoBeef Domain" which specifies the meaning of terms relevant to the domain of *beef production and consumption*.[3]

We shall also show that the ontological representation of proved and disproved statements is useful for the discovery of new knowledge. Properly designed ontological provers are able to generate new proved or disproved statements automatically from the ones already coded in the ontology.

To summarize, in the research presented in this paper we are aiming at an ontology based expert system in which knowledge from the literature on beef production and consumption is expressed in terms of ontology, to facilitate search for relevant papers and to enable gaining new knowledge.

Our Science ontology contributes also to the realization of a more general idea of integration and description of research resources (scientific papers, researchers and research activities) facilitating their access and dissemination (see e.g., [18,19]). Our approach allows for a comparison of the content of scientific papers which take into account the classes of entities they concern, their consistency, etc. It also helps to transfer knowledge from the world of science to the real-world practice by providing a simple and short note about scientific achievements contained in the scientific papers.

2. Methods

2.1. Previous works

The OntoBeef Domain and its applications have been presented in [12,16]. It is an OWL ontology currently consisting of 2408 classes, 96 object properties and 13128 axioms.[4] It is more structured than the well known agricultural thesauri such as AGROVOC [2]. Compared to other agricultural ontologies such as ATO, VT, ATOL [9,15,4,8] it focuses on concepts related to beef, but at the same time covers a wider spectrum of phenomena.

The OntoBeef Domain is based on an OWL version of DOLCE [14] as a foundational ontology. Let us mention here the four top classes of DOLCE important for this paper. DOLCE distinguishes: (1) endurants (i.e. entities 'wholly' present at any time of their existence, e.g., food, fodder, bovinae, meat), (2) perdurants (i.e. entities that 'happen in time', e.g., feeding, eating, slaughtering, fermentation), (3) qualities (i.e. entities that can be perceived or measured, e.g., tenderness, salt distribution, sex, weight) and (4) regions (i.e. quality spaces providing values used for measuring qualities, e.g., pH scale, color scale, weight scale).

So far the OntoBeef Domain has been used as the basis for two applications. The first one is a Java desktop application named Oxpecker (see [12], Section V). It facilitates the access to the database of the descriptions of scientific articles which were collected in the ProOptiBeef project. The database contains descriptions of around 2800 articles. It collects information about the structure of papers (title, author(s), abstract, etc.) and some content data, such as definitions of terms, material samples and proved and disproved statements extracted from the papers. The Oxpecker uses the annotation of articles by IRIs of the OntoBeef Domain classes to find the articles which best match the user's query in a natural language.

The other OntoBeef Domain-driven application is a web application, which initially served as a thesaurus.[5] It enabled the users to search for categories and had Linked Open Data connections to four external ontologies. Recently the application has got some functionalities of the Oxpecker and for each OntoBeef Domain class it provides the scientific laws which govern its instances. The last of these functionalities is the main subject of this paper and is described below.

2.2. Science ontology

Most of the papers collected in the ProOptiBeef database have proved and disproved statements attached. The statements were extracted from the articles by a group of experts who analyzed the papers within ProOptiBeef. The OntoBeef Domain was applied to represent them. For this purpose we have created a new ontology, which we called Science.[6] It imports the OntoBeef Domain ontology and uses its classes to represent proved and disproved statements.

Currently Science provides a taxonomy of scientific laws and their ontological characterization.[7]

The types of scientific laws adopted in the project were taken from the works of Polish philosophers of science and nature: Ajdukiewicz and Krajewski [1,10]. The distinctions

[3] It is already known that in food and agricultural Science ontology serves not only as lingua franca for researchers in the area but it is also used as a powerful tool for scientific information especially for efficient and user friendly search support [17,12,16] and expert systems [13].

[4] http://onto.beef.org.pl/ontobeef/owl/Domain.owl

[5] http://onto.beef.org.pl/ontobeef/.

[6] http://onto.beef.org.pl/ontobeef/owl/Science.owl.

[7] It lacks the possibility of representation of many aspects important for complete characterization of scientific laws; for instance, a representation of methods, observations, material samples and experimental results is currently missing in it. We plan to develop a richer ontology of science in future works.