Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 275–278

Evaluation of beef production and consumption ontology and presentation of its actual and potential applications

Rafał Trójczak, Robert Trypuz and
Przemysław Grądzki
The John Paul II Catholic University of Lublin
Faculty of Philosophy
Al. Racławickie 14, Lublin, Poland
Email: trypuz@kul.pl

Jerzy Wierzbicki and Alicja Woźniak
Polish Beef Association
Ul. Kruczkowskiego 3, 00-380 Warszawa
Email: jerzy.wierzbicki@pzpbm.pl

***Abstract*—The paper concerns beef production and consumption ontology (*OntoBeef*) and its applications. It is presented the three-stage *OntoBeef* evaluation process with a special focus on description of interaction of ontologists with domain experts. We also describe Linked Open Data (LOD) philosophy and show how links between *OntoBeef* and four other ontologies were established. We also present the components of *OntoBeef*-driven information system, a technology used to its creation and its functionalities. In particular we describe thesaurus component of the information system incorporating LOD connections.**

I. Introduction

THIS paper is a continuation of the one presented a year ago during WEO-DIA (FedCSIS) 2012 workshop (see [1]). In WEO-DIA 2012 paper we have described motivations to create a beef production and consumption ontology—which we call *OntoBeef* in short—and a project ProOptiBeef[1] within our research was carried out. In particular the paper contained information about the methodology of building the ontology, its content and possible applications. We have presented there ontological choices made while building the ontology and their justification. We have also shown how *OntoBeef* is used for browsing a database of articles (being indexed by the ontology concepts). In this paper we shall describe the way in which *OntoBeef* was further validated by the domain experts. We shall also present the initial stage of *OntoBeef*-driven application built by us—in particular we shall focus on its technological and functional sides.

The structure of the paper is the following. In Section II we present the three-stage *OntoBeef* evaluation process. In Section III Linked Open Data philosophy is introduced. In that section we show how links between *OntoBeef* concepts and the concepts of four other ontologies were established. Finally, in Section IV we present the components of a built by us ontology-driven information system, a technology used to its creation and its functionalities.

[1]http://www.prooptibeef.pl/

II. *OntoBeef* Evaluation

OntoBeef ontology has been evaluated by seven experts: Prof. Krystyna Gutkowska (Institute of Rural and Agricultural Development PAS), Prof. Zenon Nogalski and MSc eng. Maciej Borzyszkowski (University of Warmia and Mazury), Prof. Agnieszka Wierzbicka, Dr eng. Marcin Gołebiewski, Dr eng. Eliza Kostyra and MSc eng. Rita Rakowska (Warsaw University of Life Sciences). The experts has been invited to three-stage evaluation process. In the first stage of the evaluation the experts have been choosing the concepts to be validated, in the second stage of it they have been evaluating labels assigned to each concept and finally they have been assessing the ontological relations between concepts. In what follows, we shall describe in details each of the three stages.

A. *Concepts choosing*

At the first stage the experts have been asked to choose among all 2344 concepts these which belong to their domains of interest. In figure 1 we can see the screen of the application used to support the process.

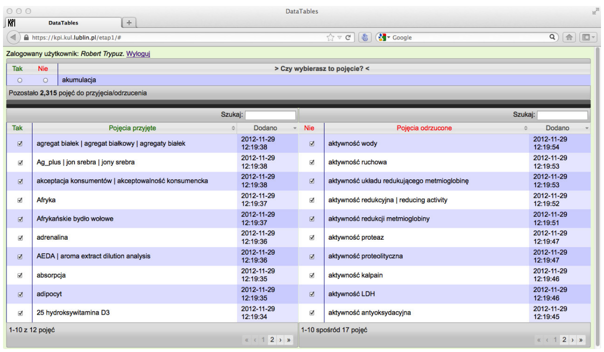

Fig. 1. First stage of *OntoBeef* evaluation

Concepts were displayed one by one. Each of them was represented by the sequence of synonymous names. An expert could choose between “Yes” (I do accept the concept and want to take care of it later on) and "No" (I don't accept the concept). Accepted concepts have been gathered in the table on the left and rejected concepts in the table on the right (see figure 1). An expert could always change his mind by unmarking the checkbox close to the chosen or rejected concept. It is worth noting that one and the same concept could be chosen by two or more experts (in fact there were many concepts validated by more than one expert; e.g. two experts shared even 675 concepts).

B. Labels evaluation

When all the concepts have been distributed to the experts, the second stage of the evaluations process began. During the stage the experts were asked to make an order in the labels assigned to each concept. In figure 2 we can see the screen of the application used to support the process.

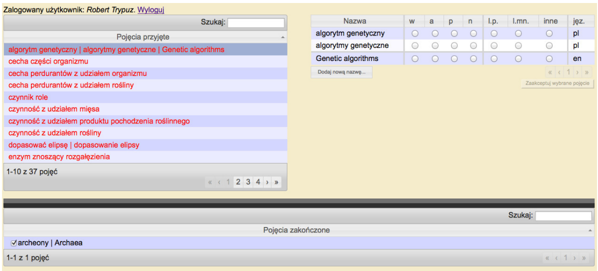

Fig. 2. Second stage of *OntoBeef* evaluation

Experts assessed each label by assigning to it one of the four label properties: "proper" (w), "adequate" (a), "common" (p) and "wrong/hidden" (n). They identify also the grammatical category of labels ("singular" or "plural") and validated their language. Experts might also add a new synonymous label to the concept in any language. Aforementioned label properties: `proper_label`, `adequate_label`, `common_label` and `wrong_label` are instances of `owl:AnnotationProperty` and stay in relation `rdfs:subPropertyOf` to `rdfs:label`. Additionally proper label is a sub-property of adequate label. It is assumed the each class has to have exactly one proper label per language[2]. Since many concepts have been chosen by two or more experts during the first stage, we could expect that the second stage of the evaluation will lead to the emergence of conflicts. And in fact after the work of the experts has been done we found that 654 concepts meet a conflict of labels, e.g. that the same label for the concept had different characteristics (from the set of properties: proper, adequate, common and wrong) or more than one label has been recognized by experts as "proper". Most of the conflicts were solved by using the following criteria:

1) in the case of labels in the singular and the plural forms determined as "proper", the singular remained "proper" whereas the plural one(s) became "adequate";
2) in case of the full name and its acronym were determined as "proper", the full name was selected as "proper" and it acronym as "adequate";
3) in the conflict between more than two experts, the choice proposed by majority was accepted;
4) in case the conflict could not be solved otherwise, the label which is more common in the Google search resources was chosen as "proper";
5) indication of experts, which is not consistent with the original meaning of the term definition has been ignored.

C. Evaluation of the ontological relations

The final stage of the evaluation process was analysis of the ontological relations in *OntoBeef*. In figure 3 one can see the screen of the application of this stage.

Fig. 3. The final stage of *OntoBeef* evaluation

An expert had an access to all the previously chosen concepts of the ontology and to the labels proposed by all experts during the second stage. For each concept the following information has been provided: its labels, ancestors, children, siblings and some ontological connections (e.g. "participation" or "parthood") with other concepts. An expert could submit a comment or suggestion on concept's labels, ancestors, etc, by clicking on "Zgłoś sugestię" button (see figure 3) – after clicking it a new window with a space for typing comment appears. All visited concepts have been colored yellow and unvisited yet—red. The result of this evaluation stage was 494 received submissions, which have been then analyzed and applied by ontologists.

D. Lesson learned

The evaluation process of *OntoBeef* by domain experts was only partially successful. During the second stage of the process (see section II-B) experts added many correct labels to the concepts, what helped us then to establish connections between *OntoBeef* concepts and LOD ontologies (see section III). But it is also true that many added labels were simply wrong, mostly because misunderstanding of the real concepts references (e.g. to the concept possessing a label "child" some expert added "calf"). Most of the remarks submitted during the last evaluation stage (344 out of 492) concerned labels.

[2]Thus `proper_label` is much like `skos:prefLabel`. Similarly `adequate_label` and `common_label` correspond to `skos:altLabel` and `wrong_label` to `skos:hiddenLabel`.

They were constructive and improved *OntoBeef* quality. The rest of them (i.e., concerning ancestors, children, siblings and some ontological connections) were rather missing the point. One of the reasons of this state of affairs is that domain experts supporting ontologist in the project were not trained in ontological thinking and did not feel competent enough to suggest changes in the *OntoBeef* structure. Our earlier experience shows that much better results gives direct face-to-face cooperation of ontologists with domain experts. But this way of processing engages more people, is more time-consuming and as such is obviously more expensive.

III. *OntoBeef* and Linked Open Data

After *OntoBeef* was finally validated, its concepts have been linked with other (lightweight) open ontologies. At least two meanings of "Linked Data" are known. In the first one the phrase means a method of knowledge creation and sharing[3]. In the second meaning "Linked Data" refers to "collection of interrelated datasets on the Web"[4]. Of course both definitions are compatible; Linked Data as a collection of interrelated datasets is brought about by many working agents acting according to Linked Data as a method. Linked Open Data is Linked Data which is released under an open license. (Semantic) Web visionary Tim Berners-Lee provided the following set of requirements which a data should possess to be called Linked Data[5]: 1) to be available on the web; 2)to be available as machine-readable structured data; 3) to be coded in some of open standards from W3C (e.g. RDF) to identify things; 4) to be linked to other people's data. It is also strongly suggested to register data at some open data catalogue (e.g. The Data Hub[6]), what in practice leads to a few more technical requirements (e.g. that HTTP URI of a piece of data should be dereferenceable – see [2]). *OntoBeef* has been linked with four thesauri: AGROVOC, General Multilingual Environmental Thesaurus (GEMET), National Agricultural Library's Agricultural Thesaurus (NAL), and STW Thesaurus for Economics (STW). Interlinking process has been done in two steps. In the first step for each thesaurus we have created database with two column table "concept number – label" . The same representation has been created for *OntoBeef*. Then by SQL query we have selected the concepts from thesauri and *OntoBeef* which have the same labels in common. In the second step ontology experts have validated the quality of the automatic connection of concepts and removed the wrong connections where needed. Finally *OntoBeef* has 797 links with AGROVOC, 211 links with GEMET, 546 links with NAL and 119 links with STW.

IV. Applications

OntoBeef and its connections to other ontologies are a good starting point for building an ontology-driven information system (IS, in short). An ontology-driven IS is IS in which "ontology profitably "drives" all aspects and all components" of it [3, section 3].

[3]http://aims.fao.org/standards/agrovoc/linked-open-data
[4]http://www.w3.org/standards/semanticweb/data
[5]http://www.w3.org/DesignIssues/LinkedData.html
[6]http://datahub.io

In figure 4 we find four components of IS currently being developed by us within ProOptiBeef project. Semantic Oxpecker was described in the proceedings of FedCSIS 2012 (see [1, section V]). Components: "theses representation and search" and "interface to the database of results of experiments" are under development. Thesaurus component has been already created and will be described in section IV-B.

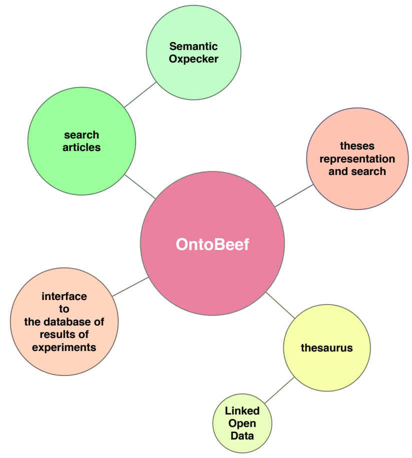

Fig. 4. Four components of information system based on *OntoBeef*

Before we started developing our IS, we have formulated a number of requirements a technology used to create the application should satisfy. The four most important of these conditions are: a) to run as the web application; b) to support OWL 2 (in which *OntoBeef* is formulated); c) to guarantee smooth application performance; d) to be flexible enough to accommodate new functionalities.

Based on our experience we chose three options initially: 1) JavaScript with jOWL framework; 2) Java Enterprise Edition ("Java EE" or "JEE", in short) with Jena framework; 3) Java EE with OWL-API framework [4]. All these technologies satisfy the first condition. Jena framework does not support OWL 2. jOWL framework does not satisfy the third condition, because it requires downloading ontology each time user's computer reloads application page. This takes time and distracts smooth application performance. Finally only OWL API framework meets all the requirements.

A. *OWL API*

In [4] we read that OWL-API is "a high level Application Programming Interface (API) that supports the creation and manipulation of OWL Ontologies". Its first version has been released in 2003. The last version 3.4.3 (which we are currently using) has been released in 2013. OWL-API is open source project managed by people from University of Manchester, written in Java programming language. It is worth noting that OWL-API was used for the development of components of a widely used ontology editor Protégé.

OWL-API allows to use a variety of notation: RDF/XML, OWL/XML, Turtle, Manchester and others. It supports the use of reasoners. It also "includes validators for the OWL 2 profiles – OWL 2 QL, OWL 2 EL and OWL 2 RL" [4].

B. Thesaurus component

Thesaurus component was developed within Java EE plus OWL-API framework. We shall now present how the web part of thesaurus component is running. Java EE consists of a large number of elements. In our application we use only a few selected, namely: servlets technology, JavaServer Pages (JSP) and JDBC technology. The first one handles the low-level web operations such as handling request and response, reading and writing HTTP headers. JSP allows to create the HTML pages in Java with great ease (in comparison with servlet technology). Access to the database is implemented through JDBC technology. It enables to abstract away from particular database technology (in our project MySQL RDBMS was adopted).

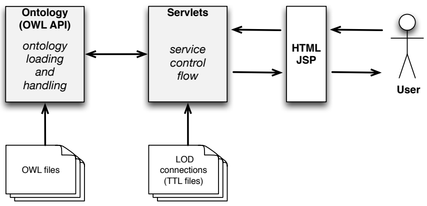

Fig. 5. The control flow diagram for thesaurus component of application

In figure 5 we can see the control flow diagram for the thesaurus component. When a user invokes the web application by writing URL address or by clicking any link on thesaurus component's page, a user's browser is sending HTTP request to a server where application is running. The server (in our case Apache Tomcat web container) receives request and runs appropriate servlet. If necessary the servlet retrieves information about the concept from OWL file by OWL-API. When the servlet obtains data about the concept it also checks Linked Open Data connections stored in TTL files. After obtaining all the necessary information, HTML page is sent to the end user. From the end user perspective the application looks as presented in figure 6. For each concept – in this case "beef"[7]– in ontology the component displays its labels: common and adequate (among them the proper ones indicated by the green color), the list of domain experts who validated the concept, the ancestors, the children, the sibling concepts and some other ontological relations as for instance parthood and participation. An end user can search for a concept. It is worth noting that a concept can be found also by typing a wrong label which are assigned to the concept (however they are invisible for the user). Application enables also registration and after log in allows reporting suggestions and comments considering concept labels, ancestors, children, siblings and other ontological properties. In the top application bar (see figure 6) there is LOD part, which displays LOD connections. In figure 6 we can see that beef class is linked with three thesauri. By linking *OntoBeef* with other resources we get for instance translations of labels to 22 languages, definitions and some related (RT), border (BT) and narrower (NT) terms to the searched one.

[7]See: http://onto.beef.org.pl/domain/concept/201

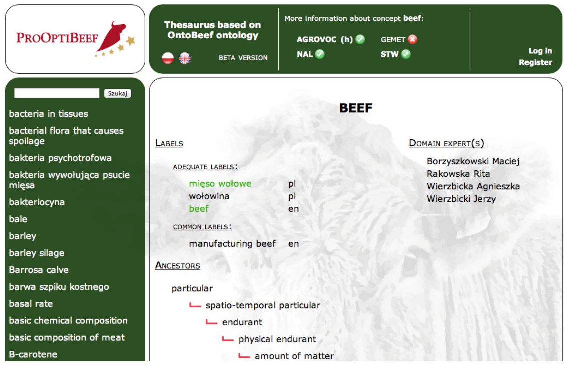

Fig. 6. *OntoBeef* thesaurus component

V. Conclusion and Perspectives

In this paper we described how *OntoBeef* ontology was validated by the domain experts. We presented the thesaurus component of *OntoBeef*-driven application. We also described the technological and functional aspects of our application. Finally the ongoing work was also described. We are very happy to notice that the researchers and practitioners working in the domain of beef production and consumption in Poland recognize the impact an ontology (as an artifact and as a methodology) had on the way they think about their domain and on the quality of their communication. There is still a lot of ontological work to be done in the field. For instance the beef sector has a lot of local carcass cuts systems which are in part mutually incompatible. We believe that their ontological implementation in *OntoBeef* will be the first step towards their comparison and integration

Acknowledgment

Research was realized within the Project no. WND-POIG.01.03.01-00-204/09 Optimizing of Beef Production in Poland According to "from Fork to Farm" Strategy co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme 2007 – 2013.

References

[1] P. Kulicki, R. Trypuz, and J. Wierzbicki, "Towards beef production and consumption ontology and its application," in *Proceedings of the Federated Conference on Computer Science and Information Systems*, 2012, pp. 483—488.

[2] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, ser. Synthesis Lectures on the Semantic Web. Morgan & Claypool Publishers, 2011.

[3] N. Guarino, "Formal ontology in information systems," in *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June*, I. P. Amsterdam, Ed., 1998, pp. 3–15.

[4] M. Horridge and S. Bechhofer, "The OWL API: A Java API for OWL Ontologies," *Semantic Web Journal*, no. 2(1), pp. 11–21, 2011.