Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 483–488

ISBN 978-83-60810-51-4

Towards beef production and consumption ontology and its application

Piotr Kulicki
and Robert Trypuz
Faculty of Philosophy
John Paul II Catholic University of Lublin
Al. Raclawickie 14, Lublin, Poland
Email: {kulicki,trypuz}@kul.pl

Jerzy Wierzbicki
Polish Beef Association
ul. Kruczkowskiego 3
00-380 Warszawa
Email: jerzy.wierzbicki@beef.org.pl

***Abstract*—We present a current state of development of an ontology of beef production and consumption developed within the interdisciplinary project *ProOptiBeef*, which focuses on increasing the level of innovation of beef sector in Poland. The ontology is intended to improve communication between specialists from different disciplines and expressing the results of the project in an unambiguous and machine readable way. We present the methodology applied to creating the ontology and several specific problems that occurred during its development. The ontology has been used as a component of semantic search engine, Oxpecker, which is described in the paper.**

I. Introduction

THE MAIN role of ontologies is to provide a clear understanding of terms used by different parties involved in information exchange [1], [2]. They are of special importance especially in those fields where the parties essentially differ. This is exactly the case in the domain of *beef production and consumption* where we find experts in: marketing and consumer research, economics of consumption, sensory analysis, chemical research, development of green areas, cattle feeding and farming, evaluation of material of animal origin, technology and chemistry of meat. That is why it was not surprising that the need of ontology was quickly recognised by the leaders of project *ProOptiBeef – Optimising beef production in Poland according to strategy "from fork to farm"*[1]. The aim of the project is an attempt to increase the level of innovation in Polish beef sector through comprehensive research and development in the field of beef quality. Within the scope of ProOptiBeef there are experimental tasks such as for instance:

- controlled studies and observations of cattle fattening,
- analysis of the biochemical properties and evaluation of technological properties of meat,
- sensory analysis,

and more theoretical activities such as:

- comprehensive study of the literature on beef and
- creating interactive knowledge base containing information extracted from the literature and results obtained experimentally in the project.

[1] http://www.pzpbm.pl/wolowina/

The expected users of the results of the project are researchers working in the related fields, agricultural advisors and people directly involved in the process of beef production and consumption, such as: agricultural producers, processors, traders and consumers. All of them are expected to have the access to results of the project via Internet.

Taking into account the fact that the amount of information in the project is growing fast and that the information comes from different sources and research domains, the leaders of the project decided to create beef ontology covering categories and terms relevant for the project. The ontology is intended to help in expressing the results of the project in an unambiguous way. Moreover, the ontology can be used as a component of a system for searching information in a database of scientific articles created within the project and an expert system gathering the results of the project and revealing them to the public.

The methodology of building the ontology, its content and application is the main subject of this paper. In Section II we refer to other works on related subjects which form the state of the art in the area. In Section III we sketch the methodology that was implemented in the process of creating our ontology. In Section IV we present selected problems, ontological choices and lessons learned during the work on the ontology. In Section V we show how our ontology can be used for a system browsing a database of articles. Finally, in Section VI we summarise the contributions of the paper and point out at further works.

II. Related works

The work presented in this paper is in a close relation to other projects concerning the creation of unified conceptual systems for agriculture and food science and their use in the domain of scientific information. The specific character of our project, in which we have to cover a wide range of problems of a different nature from the whole chain of beef production and consumption, makes it necessary to create a new ontology but we make use of the existing ontologies and thesauri which overlap with our domain of interest.

The largest, most widely recognised and with the biggest common area of interest with our project among them is the AGROVOC thesaurus, created and maintained by FAO (see e.g. [3])[2] based on the global public domain bibliographic database AGRIS[3]. Concepts represented in AGROVOC cover a wide area of agriculture science and related disciplines in multilingual settings. From the point of view of the project discussed in the present paper AGROVOC is not satisfactory as a statement of common vocabulary for two main reasons.

Firstly, AGROVOC misses technical terminology which is specific for its area, e.g. meat processing technology and consumer satisfaction study. Secondly, the concepts are connected by relations: 'broader'/'narrower' and 'associated with'. Since the meaning of those relations is far from being precise, their use is an important drawback from the perspective of formal ontology. It is especially important for us that the former relation contains two ontological relations: subsumption and parthood (and their converses) which, in our approach, have to be distinguished. We shall address this problem later in more details.

Among other available resources it is worth to mention several domain specific ontologies covering concepts that are relevant to beef production and consumption, i.e. e.g.: Animal Trait Ontology (ATO, [4])[4], Chemical Entities of Biological Interest (ChEBI)[5] Reproduction Ontology (RepO)[6]. The overlap between those ontologies and our domain of interest is however much smaller than in the case of AGROVOC.

The ideas of semantic web and utilisation of formal ontology were, in the recent years, applied on a large scale to scientific information in general and, to some extent, in the specific area of agricultural sciences. From the former let us mention two papers [5], [6]. In both of them ontology of metadata is used for representation of information about articles from a selected journal from agricultural domain ('Food, Nutrition and Agriculture Journal' and 'The Cuban Journal of Agricultural Science' respectively). The representation of domain terminology is limited to a restricted vocabulary listing possible subjects of indexed articles. AGROVOC and Wordnet are used to connect terms occurring in users' queries with the restricted vocabulary of subjects of articles.

In contrast to those works in our approach we focus on the development of a genuine domain ontology. Utilisation of the ontology is just one of the possible benefits of having it.

III. *OntoBeef Library* CREATION METHODOLOGY

In the project several ontologies have been developed. They are gathered under the name "*OntoBeef Library*" (or just "*OntoBeef*" for short).

A. OntoBeef Library

OntoBeef Library is planned to be composed of four ontologies: Domain, Papers, Conceptualisation and Science (see figure 1). The first three ontologies from that list have been already developed and are currently being validated by experts.

[2]http://www.fao.org/agrovoc
[3]http://www.fao.org/agris
[4]http://www.animalgenome.org/bioinfo/projects/ato/
[5]http://www.ebi.ac.uk/chebi/
[6]http://www.sabre-eu.eu/Portals/0/ATO/RepO.zip

Domain ontology, provides a knowledge about beef, its production and consumption. It consists of 2310 classes, 95 object properties, 90 individuals. It contains also 2747 `owl:subClassOf` axioms.

Conceptualisation is a reification of Domain, i.e., it is a metaontology in which Domain classes, data and object properties become instances of Conceptualisation classes. It consists of 5 classes, 8 object properties, 3 data properties and 2496 individuals.

Papers provides mainly metadata knowledge about documents, i.e., their authors and structure. Additionally it refers also to Conceptualisation in order to describe the content of papers by assigning reified classes to each document. It consists of 39 classes, 67 object properties, 19 data properties and 36297 individuals.

Science ontology is intended to be a description of problems, methods, data and theses. We plan to start it's development in the near future.

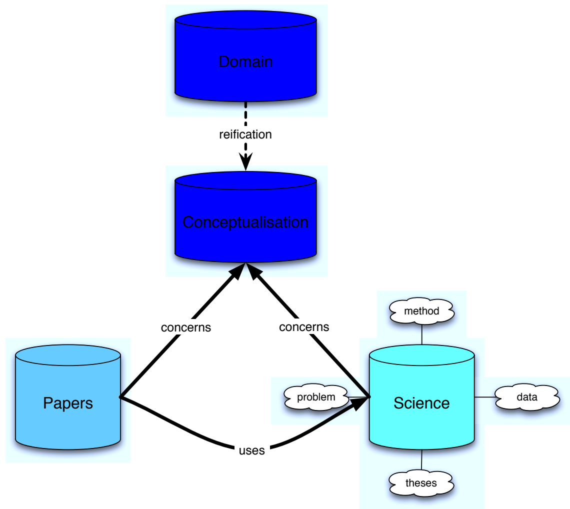

Fig. 1. OntoBeef Library

B. DataBase of articles descriptions

A starting point for creating *OntoBeef* was created in the project a database of articles. The database contains descriptions of around 2300 articles. Articles included in the database were originally published in English or Polish, their descriptions in the database are in Polish. It was done by 61 researchers working in eleven groups. Experts were responsible for selecting papers relevant for the domain of beef production and consumption and describing them according to the scheme of DB. All the articles have been divided into two groups: (A) articles presenting experiments and their results and (B) review articles (see table I). Among the database columns there were

- those referring to parts of the structure of each article, i.e., its title, author(s), abstract, keywords and bibliography and

- those describing the content of each article, e.g.: definitions, theses proved and theses disproved.

TABLE I
"X" IN THE ROW AND COLUMN MEANS THAT FILLING THE DB COLUMN FOR THE ARTICLE OF CERTAIN TYPE (I.E., A OR B) WAS REQUIRED.

	DB columns	A	B
1	Author(s)	X	X
2	Title	X	X
3	Keywords	X	X
4	Abstract	X	X
5	Summary	X	X
6	Aim and scope	X	
7	Research material	X	
8	Research methodology	X	
9	Main results	X	
10	Discussion	X	
11	Important conclusions	X	
12	Defining the issues and concepts		X
13	List of chapters and characterisation of the major issues		X
14	Summary of chapters		X
15	Recapitulation of chapters		X
16	Key positions in the bibliography	X	X
17	Personal summary of the key issues	X	X
18	Theses proved	X	X
19	Theses disproved	X	X

C. Selection of relevant terms

As a starting point of development of Domain ontology, the relevant keywords have been selected as follows. First a TXT file has been prepared for each article. It contained the content of columns: 2-5, 18 and 19 from DB (see table I). It should be stressed here that the content of the aforementioned columns was expressed in Polish. Then Polish tagger TaKIPI [7] has been used for describing words in the files by attaching tags representing morphosyntactic descriptions of words and disambiguation. Finally TF-IDF (term frequency - inverse document frequency) weight has been used to assess importance of every word (taking into account only its base form) in each article in relation to a whole collection of papers. 15 most relevant words for each article selected by TF-IDF have been intersected with the paper keywords. As the result the relevant keywords for each article have been selected.

It was assumed that the obtained keywords denote entities relevant for the beef production and consumption domain. It appeared that most of them are general names and just a few of them are individual names.

It was further assumed that each general name corresponds to the one class in OWL and each individual name to the one instance in OWL. Previously selected names became "values" of *datatype properties* of classes and instances, i.e. to each class and instant its name has been assigned by datatype property `rdsf:label`.

D. Ontologisation

The first step of ontologisation of the obtained list of notions was to identify synonymous notions appearing among keywords chosen for ontology. Such synonymous notions may have appeared in the database of articles since both authors of original papers and experts preparing database entries may have used different phrases for the same concepts (there were no formal restrictions on vocabulary of keywords).

Domain OWL classes, having at this stage different but synonymous labels, have been merged into one class. The obtained "merged" classes have obviously a few (more then one) labels assigned. All labels attached to one and the same class make up a sort of a synset structure [8].

Then all the classes have been structured according to the DOLCE ontology. DOLCE [9] is a widely recognised and used top level ontology. Among others successful applications of DOLCE let us list the following:

- it was used to create a conceptual structure *Semantic Content Model* in METOKIS project[7],
- DOLCE-UltraLite is an ontological module of e-learning project *Language Technology for eLearning*[8],
- DOLCE-Lite-Plus was used as a fundament of *Fishery Ontology Service*[9].

We have started that stage of the work with recognising an appropriate top DOLCE category for each class of DomainİIn other words, we have made decisions whether the instances of each Domain class are endurants, perdurants, qualities or regions. Endurants are "entities that are 'in time', they are 'wholly' present (all their proper parts are present) at any time of their existence." [9] (e.g., `Beefburger`, `Bovinae`, `Slaughterhouse`), whereas perdurants "are entities that 'happen in time', they extend in time by accumulating different 'temporal parts'" [9] (e.g., `Slaughter`, `Sterilisation`, `Feeding`). Qualities are particulars "we can perceive or measure" [9] (e.g., `Sex`, `Tenderness`, `Colour`). Regions are quality spaces. Each region provides values used for measuring qualities (e.g., `pH`, `Carcass grade`, `Colour space`).

The following steps consisted of constructing a taxonomy within the main categories (the relation of subsumption was used there) and connecting classes with relations other then subsumption. That step of ontologisation was performed by a team of specialists in applied ontology with the support of domain experts. The final validation of ontology and complementation with concepts important for the domain and absent in the initial list of notions is still in progress.

IV. ONTOBEEF PRESENTATION

Domain ontology of *OntoBeef* is grounded in DOLCE. In section III-D we have pointed out the four main top classes of DOLCE subsuming together all classes of Domain. In this section we would like to present how Domain is structured. Of course it is impossible to describe all 2310 classes of Domain and relations in which they remain. Let us just focus on a few interesting ontological choices and distinctions made by us while creating this ontology. It should be stressed that some of them may be rather obvious for ontologists but their application in the ontology of beef production and consumption is pioneering and beneficial for the field.

[7]http://metokis.salzburgresearch.at
[8]http://www.let.uu.nl/lt4el/
[9]http://www.fao.org/agris/aos

A. isa (subsumption), instantiation and parthood

All of the three concepts listed in the title of this paragraph appeared to be indistinguishable for the experts working in the field. Thus it was essential to make first a clear distinction between

- classes, i.e. ontological objects which are used to classify things (e.g., `Institution`), and instances, i.e. objects which are classified by classes and which cannot classify things (e.g., `Japan`)
- subsumption, which is a relation holding between classes and mainly used to express inheritance (e.g., `Institution` subsumes `Country`), and parthood, relation between a whole and its parts (e.g., `Shortloin` is part of `Hindquarter`)

After working with domain experts, we can report that the distinction between subsumption and parthood within dynamic entities, i.e. perdurants caused a particular difficulty, e.g. experts obdurately claimed that class `Stunning` is a subclass of `Slaughter`.

However, from the point of view of intended applications, such as an ontology-driven information search described in Section V, the distinction plays a crucial role. It is especially important that subsumption is transitive and a subsumed class inherits the properties of its superclasses.

B. Amount of matter vs. physical object

To understand the meaning of terms it was very helpful to refer to the unity criteria.

Among the entities with no unity such as `Chemical compound`, `Mineral`, `Isotope` there are two important categories subsumed by DOLCE class `Amount of matter`: `Amount of matter artifact` and `Amount of matter plant`. `Amount of matter artifact` contains entities being produced (in a general sense) intentionally by humans. Its subclasses are for instance: `Cereal mixture`, `Marinade`, `Milk`, `Oil` and `Uncountable food`. `Uncountable food` subsumes: `Beef stew`, `Pate` and `Roast beef`. `Amount of matter plant` subsumes classes of entities which are plants with no unity, for instance: `Lucerne`, `Grass`, `Tomato` and `Grain`.

Among physical entities with unity we have

- `Material Artifact` and its subclasses such as: `Straw Bale` and `Countable food` (`Steak` and `Burger`),
- `Organism` (among its subclasses are for example `Plant` and `Animal`),
- `Natural Thing` (among its subclasses are for example: `Part of organism`, `Part of plant`, `Inanimate thing`)
- `System as artifact` (`System sous vide`)

One should notice that in Domain there are two classes for plants with and without unity. This choice expresses the way in which domain experts think about plants. For instance in the context of feeding animals or studying nutritional properties "tomato" would mean uncountable mass of stuff, whereas in other context, for instance growing plants, it would rather refer to entities with unity. Thus, Domain ontology shows the usefulness of the DOLCE class `Amount of matter`.

C. Role

Roles are not necessary, relationally dependent and non-physical endurants [10]. `Forage` is an example of role. Domain ontology allows us to express that instances of one class can play a certain role. For example, instances of `Grass` may play a role of forage. This entails that cows do not really eat forage but grass playing a role of forage.

Among other classes `Role` subsumes `Role played by amount of matter` (examples of subclasses: `Allergen`, `Enzyme` and `Forage`) and `Role played by physical object` (examples of subclasses: `Culinary element` and `Consumer`).

From the perspective of the semantic search inter-linking roles with the entities which may play it provides a significant improvement while the expander using "friends" is applied (see section V).

The nature of roles is a subject of discussions in the community of applied ontology (see e.g. [11]) and we believe that examples from the area of agriculture can be an interesting reference for different positions in the matter.

D. Phase of the life of cattle and perdurants

Among perdurants it is worth mentioning a class `Phase of the life of cattle`. Examples of its subclasses are: `Phase of the life of bull` and `Phase of the life of cow`. Exemplary subclasses of the first class are `Life of bull` and `Young bull`, whereas the second one—`First bearing a calf` and `Heifer`. That makes our ontological choice concerning the meaning of such terms as "young bull" or "heifer" clear. The terms refer to dynamic entities, i.e. they are phases of the life of cattle. In Domain it is also true that between some phases the parthood relation holds. For instance, a phase of life of bull has as a temporal part young bull phase.

E. Qualities and Regions

Most relevant and contributing to better understanding of the process of measurement of beef properties are the choices made on the line of distinction qualities-regions.

It appeared that the same qualities of beef such as for instance belonging to the class `Tenderness` are measured in different ways and the results of the measurement are expressed in different scales. One can find in different papers the statements that a piece of beef is crumbly. Ontology forces one to make explicit which region space is used for stating "crumbly value". It gives also a conceptual framework for the comparison or inter-linking conceptual spaces.

We have also made an interesting choice concerning `Breed` being a quality of `Bos taurus`. In Domain it has values in the region `Bos taurus breed region`. The region has currently 49 instances such as for example: `Angus`, `Brahman`, `Japanese black`, etc.

Different parties involved in the project have sometimes different view on qualities. For instance `Sex` of an animal is by some experts (biologists) understood in such a way that it may take one of two values on the basis of congenital primary sexual characteristics, while some other (zootechnicians) accept here five possible values on the basis of actual hormone levels. This issue forms an interesting question for the process of creating an ontology based on experts of a certain domain, namely how to resolve a problem of inconsistent experts' opinions.

V. Oxpecker for ontology-driven information search

One of the advantages of using a formal applied ontology, that can be recognised already at the present stage of our project is a semantic search engine. A JAVA desktop application we developed within the project is called Oxpecker. Oxpecker allows the user to:

1) "translate" the database (see section III-B) whose records relate scientific papers with the keywords into an OWL ontology Papers
2) reify the Domain ontology into Conceptualisation
3) link Papers to Conceptualisation
4) create the so-called expansions of Papers
5) select a suitable expansion
6) perform a semantic search on the basis of selected expansion

The translation process 1 employs a simple ontology of documents developed for the sake of the current project. This ontology is populated with the data from the database. The reification process (functionality 2) is required because of the syntactic restrictions of the OWL language. Functionality 3 identifies the instances of the `owl:Class Keywords` in the Papers with the `rdfs:label` properties of Conceptualisation classes. The most innovative aspect of Oxpecker is related to the idea of the ontology expansion. Briefly speaking, instead of just retrieving papers whose keywords syntactically match a user's query we allow him/her to extend his/her search in two dimensions:

1) all papers whose keywords contain any synonym of any word used in the query are retrieved.
2) all papers whose keywords contain a word that labels (in the sense of `rdsf:labels`) an "expanded class" (from the reified OntoBeef ontology) whose label is used in the query are retrieved.

So any expanded search retrieves not only the papers that are related to the keywords from a user's query (see figure 2), but also those papers that are related to the "expanded labels" of the classes in the OntoBeef ontology. Each expansion model may be construed as a kind of a homomorphic image of the latter ontology. The user of Oxpecker may define any expansion he or she likes but we provide him with a number of predefined expansions:

1) basic expansion (see figure 3) in which each class is expanded by its every subclass (in the sense of `owl:subClassOf` property)

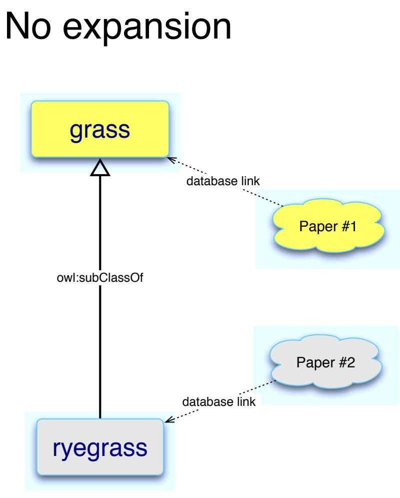

Fig. 2. No-expansion model of search

2) "parents" expansion (see figure 4) in which each class is expanded by its every superclass (in the sense of owl:subClassOf property)
3) "friends" expansion in which each class is expanded by any class that is related (in the OntoBeef ontology) to it by any owl:ObjectProperty
4) "grandparents" expansion in which each class is expanded by its every superclass and every superclass of the latter superclass (in the sense of owl:subClassOf property)
5) "parents and friends" expansion in which each class is expanded by its every superclass and then each such superclass is expanded by any class that is related (in the OntoBeef ontology) to it by any owl:ObjectProperty

As the reader may expect each expansion contains the basic expansion.

The three figures 2–4 explain the way the papers are linked to keywords in the "no-expansion" model, the basic expansion and the "parents" expansion. A query on ryegrass returns:

1) paper #2 in the no-expansion and basic model of search
2) papers #1 and #2 in the "parents" model of search

Similarly, a query on grass returns:

1) paper #1 in the no-expansion model of search
2) papers #1 and #2 in the basic and "parents" model of search

VI. Conclusions and further works

We presented the objective of *ProOptiBeef* project and the use of ontology within the project. The main part of ontology under the name Domain is currently in the phase of validation, so we are not yet able to open it to the public. Thus, we presented methodological background of the ontology and some ontological choices that we have made and some problems that occurred during its development, that we find interesting.

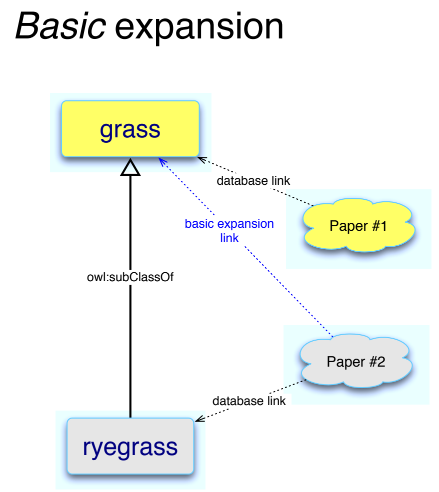

Fig. 3. Basic expansion model of search

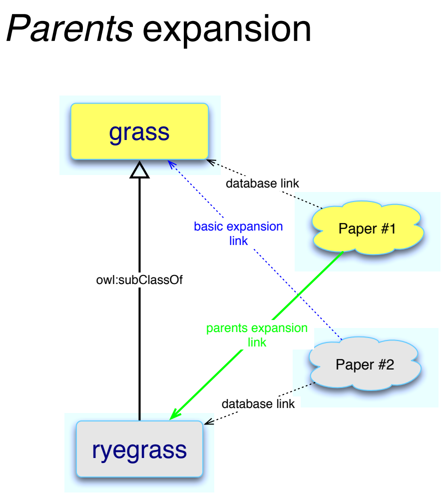

Fig. 4. "Parents" expansion model of search

Already in the current phase of its development our ontology has proven to be useful as a vital component of semantic search engine Oxpecker, which can be used for finding papers, placed in the database of scientific literature relevant to beef production and consumption chain, relevant to a user's query.

Further plans include the use of ontology as a component of expert system of the domain and publication of the ontology via a linked open data mechanism.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions, which were very helpful in improving the paper.

Research was realised within the Project no. WND-POIG.01.03.01-00-204/09 Optimising 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] N. Guarino and P. Giaretta, "Ontologies and knowledge bases: Towards a terminological clarification," in *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, N. Mars, Ed. IOS Press, Amsterdam, 1995, pp. 25–32.
[2] W. Borst, "Construction of engineering ontologies for knowledge sharing and reuse," Ph.D. dissertation, University of Twente, 1997.
[3] M. Sini, B. Lauser, G. Salokhe, J. Keizer, and S. Katz, "The AGROVOC Concept Server: rationale, goals and usage," *Library Review*, vol. 57, no. 3, pp. 200–212, 2008.
[4] L. M. Hughes, J. Bao, Z. Hu, V. G. Honavar, and J. M. Reecy, "Animal trait ontology (ato): The importance and usefulness of a unified trait vocabulary for animal species," *Journal of Animal Science*, vol. 86, pp. 1485–1491, 2008.
[5] *Ontology-based navigation of bibliographic metadata: example of the Food, Nutrition and Agriculture Journal*, 2007, International Conference on Solid Dielectrics, Winchester, Hampshire UK, 8-13 July 2007.
[6] K. Vila and A. Ferrandez, "Developing an ontology for improving question answering in the agricultural domain," in *MTSR 2009, CCIS 46*, M. S. F. Sartori and N. Manouselis, Eds. Springer, 2009, pp. 245–256.
[7] M. Piasecki, "Polish tagger TaKIPI: Rule based construction and optimisation," *Task Quarterly*, vol. 11, no. 1–2, pp. 151–167, 2007.
[8] G. Miller, R. Beckwith, C. Fellbaum, and D. Gross, "Five papers on WordNet," Princeton University, Tech. Rep., 1993. [Online]. Available: ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps
[9] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari, "WonderWeb Deliverable D18, Ontology Library (final)," LOA-ISTC, CNR, Tech. Rep., 2003.
[10] C. Masolo, L. Vieu, E. Bottazzi, C. Catenacci, R. Ferrario, A. Gangemi, and N. Guarino, "Social roles and their descriptions," in *Proceedings of the Ninth International Conference on the Principles of Knowledge Representation and Reasoning (KR2004)*, C. W. D. Dubois, Ed. Whistler, Canada: M.A. Williams, June 2004, pp. 267–277.
[11] G. van Heijst, A. T. Schreiber, and B. J. Wielinga, "Roles aren't Classes: a Reply to Nicola Guarino," *International Journal of Human and Computer Studies*, vol. 46, pp. 311–318, 1997.