Food Product Ontology: Initial Implementation of a Vocabulary for Describing Food Products

Maxim Kolchin and Dmitry Zamula
University ITMO
(National Research University)
Saint-Petersburg, Russia
{kolchinmax, zamula.dmitry}@gmail.com

***Abstract*—Semantic technologies are becoming an important part of the current Web, more and more information are being published with the semanticly enriching markup such as RDFa and microformats, and more services started consuming this information. The food product field is not an exception, the manufactories and retailers, goverments and instituations looking for ways to publish their data and maximize re-use of the data. The main component of any semantic dataset are ontologies that it uses. The paper describes the initial version of ontology for describing food products that aims to provide the all interested parties with a common vocabulary for better integration and collaboration.**

***Keywords*—Ontology Engineering, Semantic Web, Ontology Building.**

I. Introduction

Semantic technologies are becoming an important part of the current Web, vast amount of data has already been and is being published in forms of (a) data on the Web enriched with a semantic markup such as Microformats, RDFa[1] and Microdata[2], and (b) Linked Data. According to the paper [3] more than 12% of web pages have structured data in form of these markup formats and the number is growing each year.

Each of these forms for publishing structured data uses some schema for describing structure and semantics of the data, such schemas are integrating parts of them. RDFa format and Linked Data method use ontologies created with OWL[4] language, Microformats and Microdata use a fixed set of vocabularies.

At this paper we present a common ontology, vocabulary or language for the food product domain that aims to help the manufactures, retailers, governments and institutions to publish their data related to this domain in a way maximising reuse of the data.

A comprehensive technical description and the Food Product Ontology itself are available at https://github.com/ailabitmo/food-ontology repository. To download and/or view the ontology, use the short (http://purl.org/foodontology#) or the full (https://raw.github.com/ailabitmo/food-ontology/master/food.owl) links.

A. Related Work

So far there were various works on representing food and related things as ontologies. Paper [6] presents an attempt to create a food ontology for sharing knowledge between different stakeholders. Another paper[5] presents a food ontology used in design and development of the food-oriented ontology-driven system (FOODS) and overviews the state of food classification and nutrition standardisation systems. In several works [7][8][9][10] ontologies representing recipes were created. Another effort[11] that was presented in LIRMM ontologies publishing platform is a food ontology built from scratch and used in Open Food Facts[1] project. Open Food Facts project aims to create a free, open and collaborative database of food products from the entire world.

The ontology that is presented in this paper built as an extension of widely used standardised ontology for product, price, store and company data, called GoodRelations[12]. Many very well known search engines support retrieving information from Web pages that have structured data and use GoodRelations to describe this data, and providing better representations of such pages in the search results. GoodRelations is used to describe structured data in RDFa and Microformats.

B. Structure of the Paper

Section 2 outlines some use cases that the ontology should support. Section 3 provides more details of the food product domain and the ontology itself. Section 4 evaluates the built ontology with SPARQL queries that can be built using it. Section 5 presents current application of the ontology. At the end we conclude our work and discuss the future work.

II. Use cases

In this section we outline several use cases where an ontology for food products and their ingredients can be used.

Use case A: A Web resource on a retailer's Web site that usually represents a particular kind of food product for sale, either for wholesalers or to end users, or both; they can offer concrete instance or it may be that it's only said that such instance exists. In general, such Web resources have only name, description, image and price of a food product.

Use case B: A manufacturer's Web site generally provides more information about their products than a retailer's one does, and the web resources are not instances of particular food products that can be purchased, but sort of specifications of

[1]Open Food Facts - http://openfoodfacts.org/

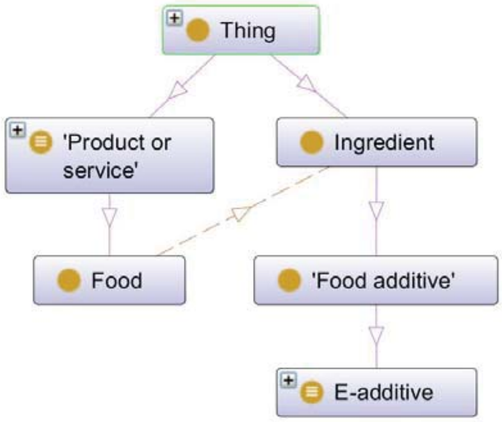

Fig. 1. The Food Product Ontology classes

food products. The specifications might has name, description, images, information about the food value and so on.

Use case C: An institution's Web site like the Food Standards Agency's one (http://www.food.gov.uk) publishes list of approved E-additives that can be used in food manufacturing. Usually the list has label and brief description of the additives.

III. Domain Capture

In the following section, we give an overview of the relevant conceptual entities and types of relationships. A visualisation of the domain capture in the form of an entity-relations diagram is shown in Fig. 1. Note that certain limitations of OWL, namely the lacking support for relations with a higher parity than two, require modeling workarounds that introduce new conceptual elements, which are only introduced in the later ontology coding stage. We use bold characters when introducing and defining a conceptual element, and underlining to refer to the particular definition of a word or group of words defined elsewhere.

A. Conceptual Overview of GoodRelations

Before we start describing the Food Product Ontology, we briefly survey the GoodRelations.

The goal of GoodRelations is to define a data structure for e-commerce that is

- industry-neutral, i.e. suited for consumer electronics, cars, tickets, real estate, labor, services, or any other type of goods,
- valid across the different stages of the value chain, i.e. from raw materials through retail to after-sales services,
- and syntax-neutral, i.e. it should work in Microdata, RDFa, RDF/XML, Turtle, JSON, OData, GData, or any other popular syntax.

This is achieved by using just four entities for representing e-commerce scenarios:

- An agent (e.g. a person or an organization),
- An object (e.g. a camcorder, a house, a car,...) or service (e.g. a haircut),
- A promise (offer) to transfer some rights (ownership, temporary usage, a certain license, ...) on the object or to provide the service for a certain compensation (e.g. an amount of money), made by the agent and related to the object or service, and
- A location from which this offer is available (e.g. a store, a bus stop, a gas station,...).

This Agent-Promise-Object Principle can be found across most industries and is the foundation of the generic power of GoodRelations. It allows you to use the same vocabulary for offering a camcorder as for a manicure service or for the disposal of used cars.

The respective classes in GoodRelations are

- **Business Entity** for the agent, i.e. the company or individual,
- **Offering** for an offer to sell, repair, lease something, or to express interest in such an offer,
- **Product or Service** for the object or service, and
- **Location** for a store or location from which the offer is available.

Authors of GoodRelations selected the following prefix and URI for the ontology:

```
@prefix gr: <http://purl.org/goodrelations/v1#> .
```

On Fig.2, a subset of classes from GoodRelations, that are primarily used in the Food Product Ontology, are shown.

The Food Product Ontology added several different classes and properties to GoodRelations: classes representing specific concepts from the food product domain such as a Food and an Ingredient, and properties such as energy per 100 gram and carbohydrates per 100 gram.

At the next subsections we describe the classes and properties of the ontology in more detail.

B. Main Concepts

The GoodRelations already has everything to describe a product, its price and specification, but without any relations to a particular domain. Our ontology defines a few new classes and properties related to the food product domain.

Food: any food product (i.e. Vanilla Yogurt). The <u>Food</u> is a subclass of <u>Product or Service</u> class from the GoodRelations that represents any product. The product can be, firstly, an <u>actual product</u>, like a Neapolitan pizza that I bought today or a concrete package of milk in a supermarket. Secondly, certain <u>product makes or models</u>, in example a make of Coca Cola. The name of a food product can be annotated with <u>gr:name</u> or <u>rdfs:label</u> properties.

Ingredient: any food or additives (i.e. salt or E385 additive) that can be used to prepare a food product.

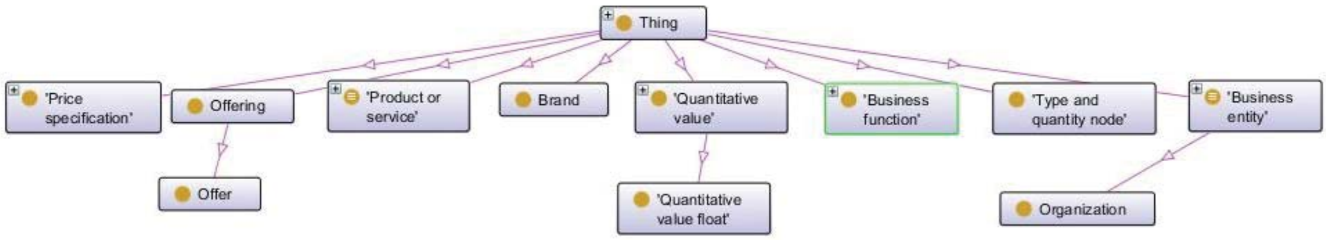

Fig. 2. A subset of GoodRelations' classes used in the Food Product Ontology

This class is a subclass of Thing which is the parent class of any class in OWL language. The rdfs:label property is used to describe a label of an ingredient. Food and Ingredient classes are related to each other through the contains ingredient object property (see Fig. 1) that will be presented later.

Carbohydrates Per 100g: a data property which represents amount of carbohydrates (in grams) per 100 grams a food product has.

Energy Per 100g: a data property which represents amount of energy (in KJ) per 100 grams.

Fat Per 100g: a data property which represents amount of fat (in grams) per 100 grams.

Proteins Per 100g: a data property which represents amount of proteins (in grams) per 100 grams.

Contains a GMO: a data property which says does a food product contains a genetically modified organism (GMO) or not.

Ingredients List: a data property which represents a list of ingredients separated by commas. The ingredients can have very different labels and usually have the same labels as on the package.

Properties such as carbohydrates per 100g, energy per 100g, proteins per 100g and fat per 100g have added to the ontology to simplify a food product representation in RDF. At the same time, you also can use the properties that have the same meaning, but are more descriptive: carbohydrates, proteins, fats and energy. These properties are object properties that semantically equal to the corresponding data properties, but their values are instances of Quantitative value float class.

Quantitative value float: An instance of this class is an actual float value for a quantitative property of a product. This instance is usually characterised by a minimal value, a maximal value, and a unit of measurement. Examples of the unit measurement: G - gram, KG - kilogramme and etc.

All properties above are the properties of Food class and its children.

C. Food Additives

At this work we also created two subclasses of Ingredient class to represent food additives which have standardised labels, such as E-additives, and which don't have ones.

Food additive: any food additive (i.e colour) that might be synthetic or natural.

E additive: a E-additive or a food additive having an E-number (i.e. E102 - Tartrazine). E additive is a enumeration of actual E-numbers, each of this numbers has an URI, in example, E102's URI is http://purl.org/foodontology#E102. E additive class includes 516 unique E-numbers.

D. Food Product Categories

Another important thing added in the ontology is the categories of food products (i.e. Confectionery, Eggs and eggs products). The food product categories were mapped to the ontology from the food category system defined by the CODEX Alimentarius[2]. The CODEX is organization established by World Health Organization and Food and Agriculture Organization of the United Nations. The category system consists of 16 top categories and more than 300 subcategories, with maximum depth equal to 4.

Food Category: the parent class for the food product categories, so each category class is a subclass of Food Category class.

Each food category has rdfs:label and rdfs:comment data properties. Below a set of triples given to illustrate the Cheese analogues category:

```
@base <http://purl.org/foodontology#> .

<CheeseAnalogues>
    a FoodCategory ;
    rdfs:subClassOf <CheeseAndAnalogues>
    rdfs:label "Cheese Analogues"@en ;
    rdfs:comment "Products that look like cheese,
                but in which milkfat has been partly
                or completely replaced by other fats.
                Includes imitation cheese, imitation
                cheese mixes, and imitation cheese
                powders."@en .
```

On Fig. 3 the all 16 top categories are illustrated. The categories have a lot of subcategories and subsubcategories, in example, on Fig. 4 the Dairy products and analogues category is shown, it has 37 subcategories and the maximum depth equals to three.

[2]The Food Category System from The CODEX Alimentarius - www.codexalimentarius.net/gsfaonline/foods/index.html

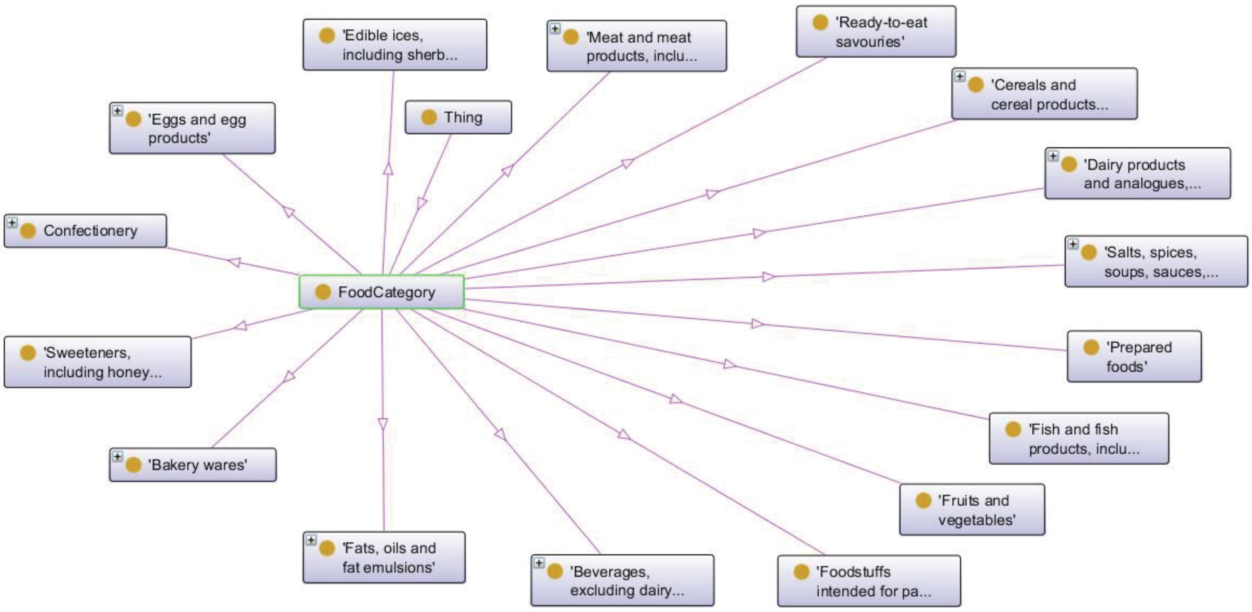

Fig. 3. Top Food Categories

E. An Example of Food Product

Below an example of food product is provided. The example is written in Turtle[13] syntax of RDF. The food product presented in the example is "Sunnyside Farms Plain Nonfat Yogurt"[3] that has zero fat, 21 (in gram) proteins, 21 (in gram) carbohydrates and 170 (in KJ) energy, also it contains the E432 additive. We use the following prefix for the ontology:

```
@prefix food: <http://purl.org/foodontology#> .
```

The example:

```
@base <http://example.org/foodproducts/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix gr: <http://purl.org/goodrelations/v1#> .
@prefix food: <http://purl.org/foodontology#> .

<4607046575613>
    a food:Food ;
    gr:name "Sunnyside Farms Plain Nonfat Yogurt" ;
    gr:hasEAN_UCC-13 "4607046575613" ;
    food:fatPer100g "0"^^xsd:double ;
    food:proteinsPer100g "21"^^xsd:double ;
    food:carbohydratesPer100g "21"^^xsd:double ;
    food:energyPer100g  "170"^^xsd:double ;
    food:ingredientsListAsText "cultured pasteurized
    grade a nonfat milk, grade a milk protein
    concentrate, maltodextrin, tapioca starch" ;
    food:containsIngredient [
        a food:Ingredient ;
        rdfs:label "Maltodextrin" .] ;
    food:containsIngredient food:E432 .
```

As you can see from the example the URI of each product is generated from the product's barcode, e.g. URI of the product is http://example.org/foodproducts/4607046575613

IV. Evaluation

In the following, we show by means of examples of queries in SPARQL how the ontology can be used to query for respective product and services data on the Semantic Web.

The queries extensively use classes and properties from GoodRealations such as gr:includesObject, gr:offers and others, details of these concepts you find in GoodRelations' specification - http://www.heppnetz.de/ontologies/goodrelations/v1.html.

A. Example 1: Who sells dried fruits and on which Web pages can I get more information on respective offerings?

This transforms into the following SPARQL query:

```
SELECT ?business ?uri
WHERE {
    ?business gr:offers ?offering .
    ?offering gr:includesObject ?TypeAnQuantityNode .
    ?TypeAndQuantityNode gr:typeOfGood ? product .
    ?product a food:Food ;
             a food:DriedFruit ;
             rdfs:seeAlso ?uri .
}
```

B. Example 2: Which offers of dairy-based drinks (i.e. chocolate milk or cocoa) exists, what is the price, and where can I find the offering on the Web?

This transforms into the following SPARQL query:

```
SELECT ?offering ?uri ?price ?currency
WHERE {
```

[3]Sunnyside farms Plain Nonfat Yogurt - http://www.fooducate.com/app#page=product&id=C1AFBFA8-5B41-11E2-AD7C-1231381A4CEA

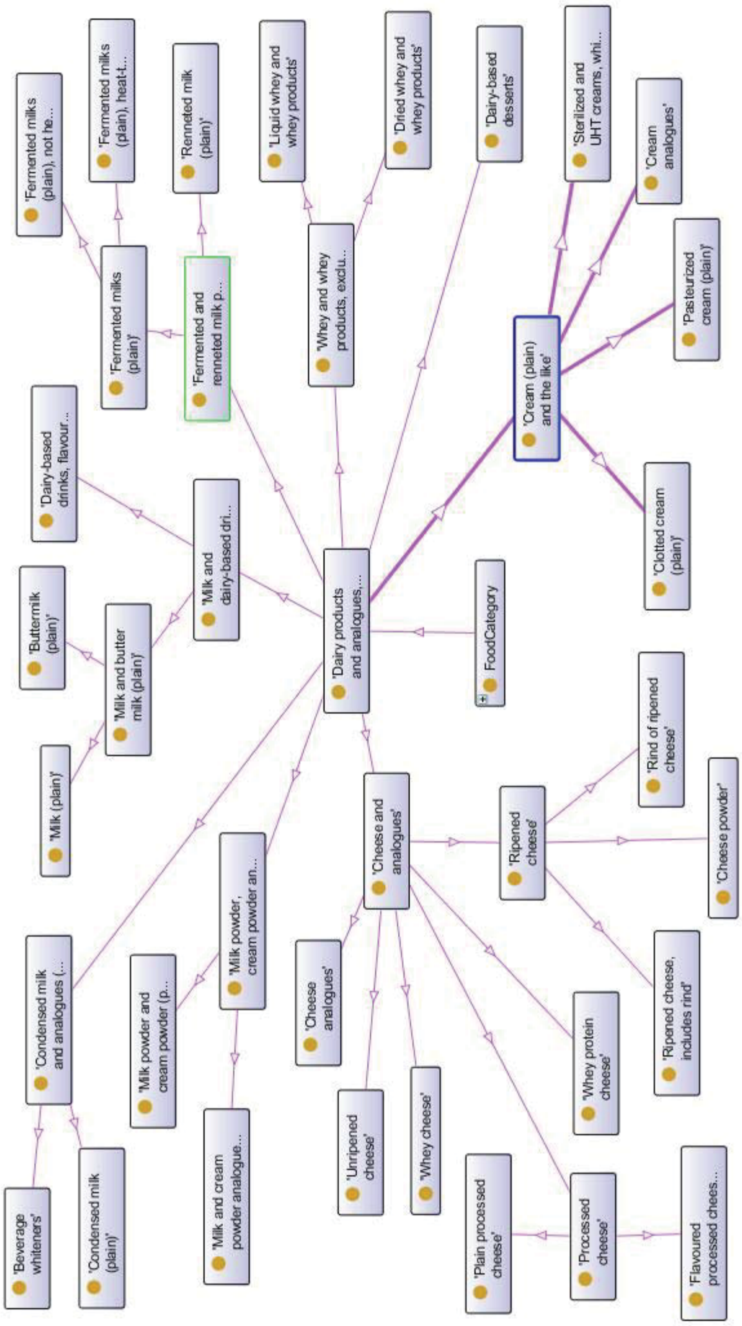

Fig. 4. Dairy products and analogues category

```
    ?offering gr:includesObject ?TypeAndQuantityNode .
    ?TypeAndQuantityNode gr:typeOfGood ?product .
    ?product a food:Food ;
        a food:DairyBasedDrinks .
    ?offering gr:hasBusinessFunction gr:Sell ;
        rdfs:seeAlso ?uri ;
        gr:hasPriceSpecification ?priceSpecification .
    ?priceSpecification a gr:UnitPriceSpecification ;
        gr:hasCurrencyValue ?price ;
        gr:hasCurrency ?currency .
}
```

The result is as expected (note that offerings that do not have a price specification do not appear in here due to the structure of the query and the semantics of SPARQL).

C. Example 3: Who sells bread and bakery wares without E-additives?

This transforms into the following SPARQL query:

```
SELECT ?business ?uri
WHERE {
```

```
    ?business gr:offers ?offering .
    ?offering gr:includesObject ?TypeAndQuantityNode .
    ?TypeAndQuantityNode gr:typeOfGood ?product .
    ?product a food:Food ;
        a food:BreadAndOrdinaryBakeryWares ;
        rdfs:seeAlso ?uri .
    ?eadditive a food:EAdditive .
    FILTER NOT EXISTS {
        ?product food:containsIngredient ?eadditive
    }
}
```

V. Application of the Ontology

The Food Product Ontology initially built for MneMojno[4] project and extensively used to represent all information about food products. MneMojno is a mobile application that provides a user with additional information about a food product that can't be found on the package, but can help select more better products, such information as a ten-point rating based on different criteria, explanation of the labels, laboratory results and etc. MneMojno is not just a mobile application, it also includes a back-end server, an RDF store, crawlers and a data management tool (OntoWiki[5]).

The Food Product Ontology is not the only one used in the project, there are two ontologies that are used to represent advices about selection of food products and an ontology for technical purposes. In total, the repository consists of more than 500 thousands triples that were crawled from different sources, refined and reviewed manually.

More information about the MneMojno project can be found in paper [16].

VI. Conclusion & Future Work

We described several use cases where an ontology can be used for describing food products on the Web and coded a respective ontology using a small subset of OWL ontology language, so that our ontology should scale very well on mainstream Semantic Web infrastructure available today. Our ontology is very flexible, while moderate in size, and supports value intervals plus existential quantification while posing only minimal requirements on the reasoning support of the ontology management infrastructure. Also, it should be compatible with some pragmatic reasoning support for SPARQL.

As future work (1) an integration with a common ontology, which represents food in general and doesn't related to food products, is possible to enrich data about food product by interlinking existent data from datasets containing lots of facts about food, i.e. DBpedia[6], Wikidata[7] and others related datasets. (2) A localisation of the ontology to different languages or even make it multilingual one using the Universal Networking Language (UNL)[8] that is presented in paper [14]. The approach to use the UNL in ontology engineering is presented in paper [15]. (3) More advanced representation of nutrition information such as saturated and trans fat, vitamins, cholesterol and etc.

References

[1] B. Adida, M. Birbeck, RDFa 1.1 Primer - Second Edition, W3C Working Group Note 22 August 2013, Web: http://www.w3.org/TR/2012/NOTE-rdfa-primer-20120607/.

[2] I. Hickson, HTML Microdata, W3C Working Draft, 25 October 2012, Web: http://www.w3.org/TR/microdata.

[3] C. Bizer, K. Eckert, R. Meusel, H. Mhleisen, M. Schuhmacher, J. Vlker, "Deployment of RDFa, Microdata, and Microformats on the Web A Quantitative Analysis", *in Proc. ISWC 2013*, Oct. 2013, in press.

[4] D. McGuinness, F. Van Harmelen, OWL web ontology language overview, W3C Recommendation, 2004, Web: http://www.w3.org/TR/owl-features.

[5] C. Snae, M. Bruckner, "FOODS: A Food-Oriented Ontology-Driven System", *in Proc. DEST 2008*, Feb. 2008, pp. 168-176.

[6] J. Cantais, D. Dominguez, V. Gigante, L. Laera, V. Tamma, "An Example of Food Ontology for Diabetes Control", unpublished.

[7] S. Mota, B. Agudo, "ACook: Recipe adaptation using ontologies, case-based reasoning systems and knowledge discovery", unpublished.

[8] G. Kalem, *Semantic web application: Ontology-driven recipe querying*. Atilim University, 2005.

[9] Linked Recipe Schema, Web: http://linkedrecipes.org/schema.

[10] Semantic Diet Web site, Recipe Ontology, Web: http://semanticdiet.com/ontology.event.

[11] LIRMM ontologies publishing platform, Food Ontology, Web: http://data.lirmm.fr/ontologies/food#.

[12] M. Hepp, "GoodRelations: An Ontology for Describing Products and Services Offers on the Web", *in Proc. EKAW 2008*, Oct. 2008, pp. 332-347.

[13] D. Beckett, T. Berners-Lee, E. Prud'hommeaux, G. Carothers, Terse RDF Triple Language, W3C Recommendation Candidate, 19 February 2013, Web: http://www.w3.org/TR/turtle.

[14] H. Uchida, M. Zhu, T. Della Senta, *UNL: Universal Networking Language An Electronic Language for Communication, Understanding, and Collaboration*. Tokyo: UNU/IAS/UNL Center, 1996.

[15] J. Cardeosa, C. Gallardo, L. Iraola, A. Miguel, D. la Villa, "A New Knowledge Representation Model to Support Multilingual Ontologies. A case Study", *n Proc. SWWS 2008*, Jul. 2008, pp. 313-319.

[16] D. Zamula, M. Kolchin, "MneMojno - Design and Deployment of a Semantic Web Service and a Mobile Application", *in Proc. FRUCT 14*, Nov. 2013, in press.

[4]MneMojno - http://mnemojno.ru

[5]OntoWiki - http://aksw.org/Projects/OntoWiki

[6]DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. Web site: http://dbpedia.org

[7]Wikidata is a free knowledge base that can be read and edited by humans and machines alike. Web site: http://www.wikidata.org

[8]UNL is a computer language that enables computers to process information and knowledge. It is designed to replicate the functions of natural languages. Web site: http://http://www.undl.org/