Ontology-Based Model for Food Transformation Processes - Application to Winemaking

Aunur Rofiq Muljarto, Jean-Michel Salmon, Pascal Neveu, Brigitte Charnomordic, Patrice Buche

Ontology-based Model for Food Transformation Processes - Application to Winemaking

Aunur-Rofiq Muljarto[1,4], Jean-Michel Salmon[2], Pascal Neveu[1], Brigitte Charnomordic[1], and Patrice Buche[3]

[1] MISTEA Joint Research Unit, UMR729, F-34060 Montpellier, France,
aunur.muljarto@supagro.inra.fr
[2] Unite Expérimentale de Pech Rouge, UE0999, Pech Rouge, France
[3] IATE Joint Research Unit, UMR1208, F-34060 Montpellier, France
[4] Dept. of Agroindustrial Technology, Brawijaya University, Malang 65145, Indonesia

Abstract. This paper describes an ontology for modeling any food processing chain. It is intended for data and knowledge integration and sharing. The proposed ontology (Onto-FP) is built based on four main concepts: Product, Operation, Attribute and Observation. This ontology is able to represent food product transformations as well as temporal sequence of food processes. The Onto-FP can be easy integrated to other domains due to its consistencies with DOLCE ontology. We detail an application in the domain of winemaking and prove that it can be easy queried to answer questions related to data classification, food process itineraries and incomplete data identification.

Keywords: ontology-based model, data integration, food processing, winemaking

1 Introduction

Researches on technological aspects in the field of food production have grown rapidly in recent years, mainly driven by consumer demands for food products that are safe, high quality and more sustainable [1]. The increasing complexity of technological aspects and the accumulation of heterogeneous data from research activities emerge new challenges related to data and knowledge organization, particularly to answer the following objectives. Firstly, providing researchers with a scientific tool for large scale data integration. It is important not only for presenting data in standard format, but the more important thing is, it will provide possibilities for further analysis by applying available knowledge [2]. Specific applications such as food traceability can only be achieved by applying data integration. Secondly, sharing data and knowledge between various stages of operations. Typically, data and knowledge are separated and reside on each section of food production chain. An expert on a particular section may not know the data and knowledge on the others. Therefore exchanging data and knowledge helps to coordinate the independent entities and increases efficiencies by greatly reducing redundancies [3]. Thirdly, providing feasible solution to address the problem of incomplete data and information. By giving proper ways to work with incomplete data and information, it will increase the quality of the whole result [4].

Developing methods intended for decision support with those objectives is fairly complicated work due to the nature of data in the field of food production chain. Despite the large amount of data collected, there are issues that require further investigation. A first key issue is their various terms, data schemes and formats used. Most of the data collected relate to the product characteristic, process conditions and other influential factors which come from experiments. Experimental data are stored and conveyed in various formats with their own schemes, such as in simple text files, csv and excel work sheets, complex word documents, laboratory reports, image files, etc., [5]. A second issue concerns the heterogeneous sources of the data. In food production chain, data are scattered at several locations and come from a variety of stages such as from the cultivation, harvesting, transformation process, and distribution of products to consumers [6]. These two issues are major obstacles to be resolved, especially for the purpose of data integration and data sharing. A knowledge layer that acts as a backbone for data and knowledge integration should be provided to overcome these issues.

Recent studies show that the ontology-based model is a flexible solution for building that knowledge layer. Compared to the previous knowledge management methods, ontology-based approach has more advantages in acquisition and creation, integrating different data sources, and interoperability among different systems [7].

In this paper, we propose a food processing chain ontology (Onto-FP) to achieve the objectives mentioned before. This ontology focuses on transformation processes by taking into account their key characteristics, i.e., food product transformation, temporal factor of operations, and data organization. To our knowledge, our proposal is the first ontology-based model describing explicitly material transformation combined with a sequence of operations. This ontology was initially built for specific domain, i.e., winemaking. However, with the need to be applied to other similar food transformation process, this ontology has been further developed to be more generic.

This paper is organized as follows: Section 2 gives a brief state of the art on food domain ontologies and process design; Section 3 is dedicated to the proposed food production chain ontology that includes how the ontology building process was carried out, detail core elements and relationships; Section 4 provides a brief description how the Onto-FP can be integrated to an upper ontology, particularly the Dolce ontology; Section 5 demonstrates some practical uses of the proposed ontology in the domain of winemaking; Finally, in Section 6, conclusions are drawn and further works are outlined.

2 Related Works

2.1 Food Ontology

The need for ontologies has increased in food sectors due to the need of knowledge expression and sharing. Some works described the food world from a general point of view such as FOODS [8], which contains specifications of food ingredients, substances and nutrition facts. Contrary, the others developed ontologies for very specific food commodities, e.g., wine classification [9], potato [6], and fish production [10]. Another part of the food domain that recently attracts the attention of researchers is ontologies for food traceability. It is driven by a growing interest in developing systems for food supply chain. Some essential works in this domain have been carried out, such as The Food Track and Trace Ontology (FTTO) [11] and TraceALL [12].

The ontologies mentioned above have proven useful and can be applied to their domain. However there are some important things that have not been presented in those ontologies. Most of them describe the classification concepts and their relationships in the target domains, but how food product transformations and processes associated with the transformations as well as important factors such as temporal aspects of operations have not been widely studied.

2.2 Ontology-based Model for Process Design

To date, an ontology related to food transformation processes has not been widely discussed. However, since food transformation processes are mainly derived from chemical processes, related works in the domain of chemical process design can be used to support this work.

A widely known ontology that currently become reference in the domain of chemical engineering is OntoCAPE [13]. It is a formal, heavyweight ontology. In this ontology, the design, construction, and operations of chemical plants are considered as the major engineering activities. Furthermore, OntoCAPE provides chemical engineering concepts needed for describing structural and phenomenological details of the chemical process. Based on this ontology, a framework for work process modeling in the chemical industries has been developed [14]. This framework comprises an iterative modeling procedure, an extensible modeling language for work processes (including temporal aspects of chemical operations), and software tools for its practical application.

3 Proposed Model

Food production chain is a complex system with many variations. Therefore, building an ontology that can be used to represent different food production chain is a challenging task. The ontology should be fairly generic to be easy to use. However, that ontology should also provide flexibility for different varieties of food production chain.

3.1 Ontology Building Process

To date, several methodologies related to the ontology development and maintenance have been proposed in the literature. [15] have discussed comprehensively about the three most well known generic methodologies, i.e., TOVE (Toronto Virtual Enterprise), Enterprise and Methontology. In the domain of agriculture, [7] proposed a general method for the construction of agricultural ontology. The main stages of this method are determination of ontology purpose and scope, collection and analysis of domain information, key concepts and relationships identification, formalization, confirmation an evolution, and ontology evolution.

Based on this development approach, the ontology building process of Onto-FP has been done by a series of activities : (1) Defining ontology purpose and scope through meetings with experts; (2) Knowledge acquisition by using various techniques, such as brainstorming, interviewing with experts in the domain of food processing, literature searching, etc; (3) Conceptualization by accurately selected domain relevant concepts and relations according to the purpose; (4) Formalization by using ontology language and semantic web technology (RDF, RDFS, OWL DL, etc); (5) Confirmation and evaluation of ontology by using an automatic reasoning tool provided by an ontology development tool (Protege) in order to check its correctness and logical relationships between concepts.

3.2 Ontology Core Elements

Onto-FP is purposed mainly for providing researchers a framework for constructing their knowledge in semantic way. The domain is specified into food processing, where raw materials are processed into final products. It covers the wide range of activities, started from harvesting, preparation, intermediate processes and final processes.

Main Concepts. Four main classes have been defined to represent general food transformation process, i.e., *Product*, *Operation*, *Attribute* and *Observation*. Figure 1 shows the complete hierarchy of these classes. A short description for each class is provided in the following paragraphs.

- *Product*
  Product class represents an abstract model of the different types of food products. The product taxonomy used in the Product class is based on the product transformation stages. Thus, as shown in Figure 1, the second level of this class comprises *HarvestingProduct* that represents raw material entering particular food transformation process, *IntermediateProduct* represents semi or unfinished products, *FinishedProduct* is an abstract model of final products and *ServiceProduct* models all materials or products that are used by operations to transforms raw material or intermediate product during process flow.

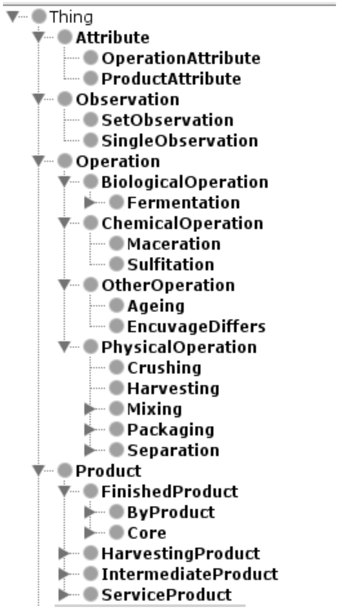

Fig. 1. Food processing domain hierarchy.

- *Operation*
  Operation is class that conceptualizes the knowledge related to the process activities. The second level of this class corresponds to the general classification of operations including physical, chemical and biological operations (see Figure 1). The third level represents more specific operations widely known in the domain of food processing, such as fermentation, separation, mixing, maceration, crushing, etc. This level can be added depending on the needs of selected domain.
- *Attribute*
  Attribute class models all the characteristics of product or operation. The aim of this class is to store information about all features or qualities belonging to Product or Operation class. *ProductAttribute* is a class that represents all attributes of *Product* while *OperationAttribute* is used to model *Operation*'s attributes. For the purpose of generality, more specific attributes are defined as an instance of *ProductAttribute* or *OperationAttribute* rather than as a new class. For instance, volatile acidity as an attribute of finished wine is declared as an individual, not as a sub-class of ProductAttribute. By using this approach, this ontology becomes more stable and flexible enough to cope with the rapid changes on qualities used in product and operation, which commonly happen due to new innovations on sensors and observation methods.
- *Observation*
  The class of Observation is a conceptualization of an abstract model of activity where an instance of Attribute class is measured. Observation can be a single observation (*SingleObservation* class) that means one time only

measurement or a set of observation (*SetObservation* class) where multiple measurement to the particular attribute of product or operation are needed (see also in Figure 3). An instance of SetObservation will always have at least two SingleObservation instances. The class of SetObservation is very important to model a series of observations, such as fermentation temperature that is commonly measured not only once but several times in a given operation.

General Relations between Elements. Generally, relations among core elements can be grouped into four categories, relations between : (i) two or more products, (ii) products and operations, (iii) two or more operations and (iv) relations related to data observations. Figure 2 shows the first three general relations, while Figure 3 shows relationships between Observation, Product, Operation and Attribute.

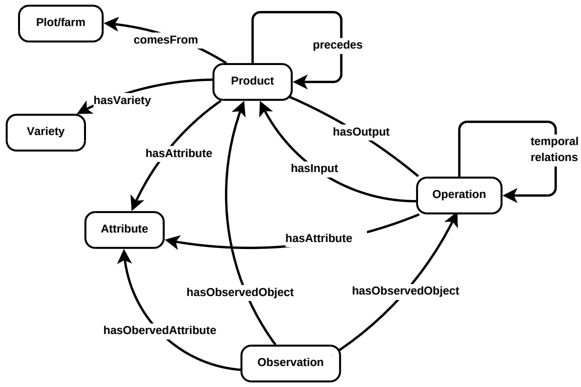

Fig. 2. General relationships between elements.

– *Product transformation*
  Product transformation is one of the key elements in food production. Raw materials are processed into intermediate products to subsequently be further processed into the finished products. In the domain of food processing, raw materials are crops that are harvested from farm. These products can be in different varieties. In most cases, varieties will determine what kind of operations that should be considered and what type of final products that may be produced. Therefore, the concept of variety is important (see Figure 2).
  There are two relations proposed concerning product transformation, i.e., product to product and product to operation relationships. The first relation aims to list all existing products and arrange them according to a particular sequence in a food transformation chain. Here, a transitive object property called *precedes* and its inverse *succeeds* is used to model this relation. The second relationship is inspired by the basic theory of material balances, where there are links between products as input, operations as processor, and other

products as output. The object properties of *hasInput* and *hasOutput* are used to model this condition (see Figure 2). By using this representation, product changes due to a particular operation can always be identified.

– *Temporal relations of operations*
Temporal factors are important issues in food transformation processes. They become important considerations because they will affect the attributes of the raw materials, semi-finished products, as well as the final products. Representing temporal factors or dynamic aspects of a particular domain such as food transformation will help researchers or related actors to deal with problems of prediction, planning and data explanation [16]. Temporal relations also can be used to check the consistencies of the set of operations. Semantic temporal relations between operations can be described by the Ontology of Time for the Semantic Web proposed in [17]. This ontology relies on the interval representation of time developed by Allen [16]. Furthermore, this ontology has been refined and listed as W3C Working Draft since September 2006. According to this ontology, there are two subclass of *TemporalEntity*, i.e., *Instant* and *Interval*. "Intervals are, intuitively, things with extent and instants are, intuitively, point-like in that they have no interior points" [17]. Using this definition, it can be stated that Operation class is a subclass of Interval. Thus it will inherit the two important properties of Interval, i.e., *hasBeginning* and *hasEnd*. It can be also inferred that operations can use object properties of interval to represent semantic relation between operations. Table 1 shows all possible relations between two operations and corresponding examples.

Table 1. Temporal relations between operations

Relations	Inverse	Notation	Examples from winemaking
Before*(A,B)*	After*(B,A)*	$A < B$; $B > A$	Crushing occurs after destemming
Meets*(A,B)*	MetBy*(B,A)*	$A\ m\ B$; $B\ mi\ A$	Draining is started immediately after maceration
Overlaps*(A,B)*	OverlappedBy*(B,A)*	$A\ o\ B$: $B\ oi\ A$	Malolactic fermentation can be started before alcoholic fermentation is finished
Starts*(A,B)*	StartedBy*(B,A)*	$A\ s\ B$; $B\ si\ A$	Maceration is started when alcoholic fermentation started
Finishes*(A,B)*	FinishedBy*(B,A)*	$A\ f\ B$; $B\ fi\ A$	Sulfitation finished malolactic fermentation
During*(A,B)*	Contains*(B,A)*	$A\ d\ B$; $B\ di\ A$	Alcoholic fermentation occurs during maceration
Equal*(A,B)*		$A = B$; $B = A$	Extraction of ethanol is started and finished at the same time as alcoholic fermentation

– *Flexible data organization using attribute and observation*
Recording chronological changes in the products and operations during the process flow is very important. For the specific application of food traceabil-

ity, data and information regarding product transformation are main sources for identifying potential causes if some problems arise. Therefore, data must be organized in such a way that represents the actual condition of the product or operation that has observed. Here, we proposed a simple approach based on natural relationships that occur between Product/Operation, Attribute, and Observation. Figure 3 shows how this approach is represented.

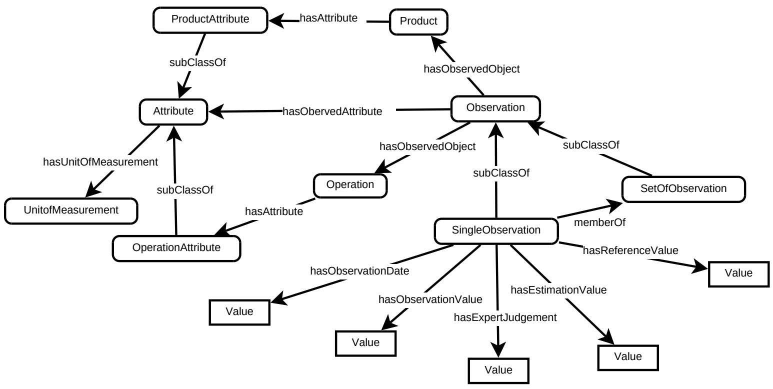

Fig. 3. Data organization using Product/Operation Attribute and Observation.

The Observation class plays an important role to represent the attribute measurement process. An instance of single observation has always relations with an instance of product or operation through the object property of *hasObservedObject* as well as with an instance of product/operation's attribute using *hasObservedAttribute* object property. It is a natural representation where an observation is done on a particular attribute belonging to a given object (product or operation) at a specific time. Another important thing regarding observation is how various data types have to be represented. For a single observation result, data can be quantitative values provided by a sensor or based on estimation, or qualitative values based on expert judgments. Thus, as seen on Figure 3, object properties such as *hasObservationValue*, *hasEstimationValue*, and *hasExpertJudgment* are added to the single observation class.

4 Integration into an Upper Ontology

The Onto-FP was built initially for a specific domain (i.e., winemaking ontology). From the development perspective, it can be stated that it follows a bottom-up approach where the process started by defining the most specific concepts in the domain of winemaking. This approach results an ontology which is in accordance with the specific conditions of the domain being modeled. However, there is a possibility that this ontology will be difficult to be modified and integrated with ontologies developed for other domains [18]. Therefore, for the purpose of information and knowledge integration, it is necessary to do an analysis to show that the core concepts of the Onto-FP are consistent with an upper level ontology.

4.1 Upper Level Ontology

Upper level ontology can be defined as an ontology that contains very general concepts and relations from which more specific concepts and relations can be constructed [19]. This definition implicitly states that if a generalization is performed in ontologies from different domains, at some point there are the same concepts and relations across all domains. Some widely known upper ontologies are: Suggested Upper Merged Ontology (SUMO), Basic Formal Ontology (BFO), General Formal Ontology (GFO) and DOLCE [19].

This paper used DOLCE, a widely recognized and used upper level ontology, as reference. According to its authors, DOLCE is declared as an ontology of particulars, where particulars refer to instances which differs to universals that point to properties and relations [20]. The top-level categories of DOLCE and their relations are presented in Figure 4. According to this figure, there are four main classes of DOLCE, i.e., *Endurant*, *Perdurant*, *Quality* and *Abstract*. Endurants can be seen as "entities that are wholly present at any time they are present" while perdurants are "entities that happen in time" [20]. Qualities are the basic entities that can be observed or measured [20]. Abstract class refers to entities that do not have spatial nor temporal qualities, and they are not qualities themselves [20].

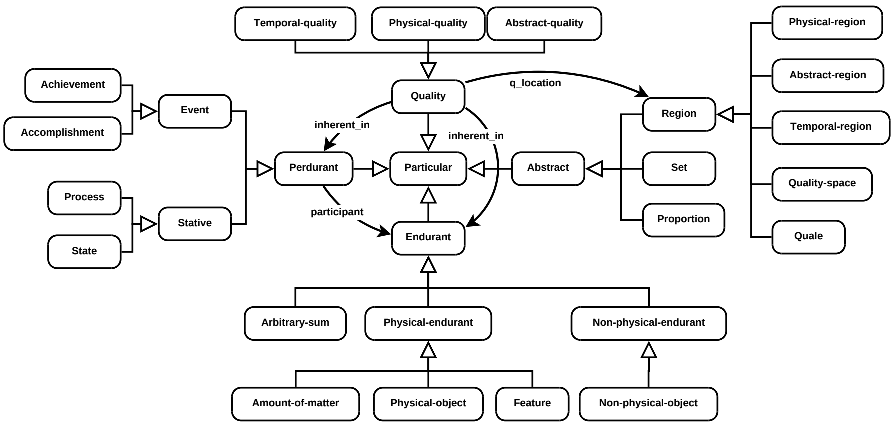

Fig. 4. Dolce main concepts.

4.2 Structuring Onto-FP to the DOLCE Ontology

Structuring Onto-FP to DOLCE ontology consists of assignment of core concepts from the Onto-FP to the categories provided by DOLCE. The following paragraphs describe briefly relations between those concepts.

- *Product*
  Product is a concept used to model all types of materials. This definition clearly expresses that Product is a physical entity. According to the definitions of DOLCE main concepts, Product can be categorized as an Endurant. More precisely, Product is a subclass of Physical-object (see Figure 4).
- *Operation*
  Operation is a concept that represents process activities. It is an action that just extends in time by gathering different temporal parts. Every instance of Operation always has temporal parts, e.g. starting time, duration, end time, etc. From the DOLCE's point of view, this concept lies clearly under the Perdurant concept. Operation is close to the DOLCE's concept of Process which is a sub-subclass of Perdurant (see Figure 4). Another important thing here is the relation between endurants and perdurants called *participant* or *participation-in* as its inverse. This relation means that an endurant exists in time by participating in a perdurant [20]. This generic relation is comparable to the relations defined in Onto-FP, i.e., *hasInput* (inverse: *isInputOf*) and *hasOutput* (inverse: *isOutputOf*). An instance of Product exists by participating as an input (or output) of an instance of Operation. Thus *hasInput* and *hasOutput* can be declared as a specialization of participation-in relationship.
- *Attribute*
  As defined before, Attribute is a class that models all the characteristics of Product or Operation. Using this definition, it can be directly revealed that Attribute is close to the DOLCE concept of Quality due to their similarity of meaning. Both of them are entities that can be observed. More precisely, Attribute could be considered as a subclass of Quality. In Onto-FP, Attribute is intended to represent physical qualities of Product or Operation. Therefore, the concept of Attribute is equal to the DOLCE concept of Physical-qualities. Additionally, from Figure 4, it can be seen that there is a relation between Quality and Endurant or Perdurant named *inherent-in*. This relation is identical to the relation of *hasAttribute* that link between Attribute and Product or Operation in Onto-FP (see Figure 3).
- *Observation*
  Observation is a conceptualization of an activity where data or information from an instance of Attribute class is captured. Like Operation, Observation is an action that also has temporal parts. Therefore, it can be stated that Observation is a Perdurant.

5 Practical Use

In this section, three examples of the practical uses of this ontology will be presented. These three examples have been selected to represent common conditions found in winemaking. For the testing purposes, the Onto-FP has been populated using winemaking data collected from the *Unite Expérimentale de Pech Rouge*, France from 2005 to 2008. These data contain observation results from the different stages of winemaking and are stored in more than 580 Microsoft Excel files.

5.1 Identifying Red Winemaking Data from the Other Types

The *Unite Expérimentale de Pech Rouge*, France produces different kind of wines. According to their color, it can be classified into red, white and rosé. The data observed from these three types of winemaking are stored together in the same file with similar scheme. Consequently, it is not easy to classify whether particular sets of data belong to red, white or rosé winemaking. Some guidance from experts are needed to check and verify them manually. Thus, it will take quite a lot of time to do this classification when the number of data increases. By using the ontology of food transformation process, this process can be done instantly by querying the data using specific criteria.

Lets take an example here. Suppose we want to select all sets of red winemaking data. The colors of wine are determined by the variety of grapes used and/or the processes used to make the wine. According to the experts, the easiest way to distinguish red winemaking to the other types is by checking the existence of *remontage* operation during alcoholic fermentation. *Rémontage* is the French term for the process of pulling out wine from underneath the cap of grape skins and then pumping it back over the cap in order to stimulate maceration. Therefore, the existence of a rémontage operation can be used as a criterion to filter related red winemaking data. The following lines is an example of SPARQL query for identifying red winemaking datasets.

Query 1.

```
SELECT * WHERE {
    ?WmDataset rdf:type own:Winemaking .
    ?WmDataset own:hasPart ?macName .
    ?macName rdf:type own:Maceration
    OPTIONAL {?remName own:during ?macName .
?remName rdf:type own:Remontage }
}
```

Query 1 shows simple SPARQL query using OPTIONAL statement to check if an instance of Rémontage (*remName*) exists during maceration operation (*macName*). If it is present then the instance name of rémontage will be displayed and the corresponding winemaking dataset (*wmDataset*) can be determined belongs to the red winemaking group. Figure 5 shows the result of this query.

5.2 Backward Tracking to Find Winemaking Itineraries by Given Wine Color Parameters

Color is one of the important characteristics of wines and probably the first attribute that affects consumer acceptance. The color of wines can be estimated

WmDataset	macName	remName
wm_0154	mc_0348	rm_0004
wm_0063	mc_0038	rm_0153
wm_0092	mc_0427	rm_0042
wm_0101	mc_0011	rm_0001
wm_0005	mc_0115	
wm_0106	mc_0061	

Fig. 5. Query results show the first four rows are winemaking datasets.

by investigating their chemical composition which are related to color, such as anthocyanins, total phenolics and tannin. For researchers, the colors of wines are useful to find the relationships between wine color attributes and various factors such as grape varieties, selected processes, treatment during these processes, temporal factors, and other interesting studies. To describe these relationships, the first thing that should be known is the winemaking itineraries, which comprise all related products and operations to produce particular wines. Certain wines may be produced by using a single straight process flow, but the others may need the combined process flows. By using these itineraries, it will help researches for describing comprehensively all potential factors that affect to wine color attributes. This ontology provides mechanisms to do backward tracking by using transitive object properties mentioned before (*precedes*, *succeeds*, *hasInput*, *hasOutput* and temporal relations between operations). Query 2 shows an example how one of these object properties is used to find all previous operations.

Query 2.

```
SELECT  ?pa ?c WHERE {
    ?s own:hasObservedObject ?o .
    ?s own:hasObservedAttribute own:Anthocyanins .
    ?s own:hasObservationValue ?v .
    ?p own:hasOutput ?o .
    ?p own:after+ ?pa .
    ?pa rdf:type ?c
    FILTER (?v = 479.22)
}
```

Here, we want to find all operations that are used to produce a wine that has anthocyanins value of 479.22. The *after* temporal relation is used to find previous operation. Because it is a transitive object property, thus we can use one of SPARQL Property Path expressions, i.e., by adding operator "+". This operator allows to find a path of one or more occurrences of *after* object property. Figure 6 shows the result of this query.

5.3 Identifying Incomplete Data

Winemaking data contains large amounts of data collected from different sources. One of the conditions that normally occur on such data is that some parts of

Instance Name	Class Name
ag_0002	Ageing
sf_0002	Sulfitation
mf_0023	MalolacticFermentation
pr_0002	Pressing
dr_0250	Draining
af_0012	AlcoholicFermentation
cr_0228	Crushing
ds_0462	Destemming
hv_0025	Harvesting

Fig. 6. Set of operations for producing a particular wine with anthocyanins = 479.22.

them are lost or unavailable [21]. These incomplete data will affect data analysis methods and the conclusions that can be drawn. Therefore, dealing with incomplete data is an integral part of research activities.

Unlike conventional database, RDF does not provide mechanism to store null value which is commonly used to represent incomplete data. RDF stores data in a triplestore which is a collection of triples rather than in a set of tables. Each triple contains flat data in the form of subject-predicate-object (S-P-O). This form has a basic consequence that all data should be known so that the triples can be built. According to this rule, the null value which is mostly used to represent that the value (data) is unknown or doesn't exist, does not fulfill the RDF standard form. Hence, null values will be disregarded because the triples cannot be generated. In conventional database, it is easy to query data using null values as a keyword, even without knowing the structure of data. But, to do that in a triplestore, it will be a bit tricky. Query 3 shows an example how to display all observation data and identify the existence of incomplete data.

Query 3.

```
SELECT DISTINCT ?obs ?obj  ?att ?val WHERE {
    ?obs rdf:type own:SingleObservation .
    ?obs own:hasObservedObject ?obj .
    ?obj rdf:type ?cob .
    ?obs own:hasObservedAttribute ?att .
    OPTIONAL {?obs ?h ?val .
?h rdfs:subPropertyOf own:hasInformation }
}
ORDER BY desc(?obj)
```

Again, here the OPTIONAL statement is used to check if a triplet that contains value exists (indicated by statement of "?obs ?h ?val"). If it is present, a complete triplet will be generated. Otherwise it will remain empty, which indicates incomplete data. Figure 7 shows the result of this query. There are two missing values found, i.e., values (*val*) of DO250 in the observations (*obs*) of so_0012 and so_0010.

obs	obj	att	val
so_0004	fw_0402	IPT	"45.90"^^<http://www.w3.org/2001/XMLScher
so_0003	fw_0402	Anthocyanins	"479.22"^^<http://www.w3.org/2001/XMLSch
so_0011	fw_0402	DO520	"6.414"^^<http://www.w3.org/2001/XMLScher
so_0007	fw_0114	IPT	"87.34"^^<http://www.w3.org/2001/XMLScher
so_0008	fw_0114	Anthocyanins	"463.44"^^<http://www.w3.org/2001/XMLSch
so_0012	fw_0114	DO520	
so_0002	fw_0031	IPT	"56.85"^^<http://www.w3.org/2001/XMLScher
so_0001	fw_0031	Anthocyanins	"674.53"^^<http://www.w3.org/2001/XMLSch
so_0010	fw_0031	DO520	
so_0005	fw_0003	Anthocyanins	"249.36"^^<http://www.w3.org/2001/XMLSch
so_0006	fw_0003	IPT	"35.92"^^<http://www.w3.org/2001/XMLScher
so_0009	fw_0003	DO520	"1.903"^^<http://www.w3.org/2001/XMLScher

Fig. 7. Query results for identifying incomplete data.

6 Conclusion and Future Works

This paper presents an ontology-based model for food transformation process, the Onto-FP. The ontology is intended to be a knowledge layer that can be used by researchers for data and knowledge integration and sharing as well as for further analysis. The Onto-FP is based on four main concepts: Product, Operation, Attribute and Observation. Beside those main concepts, the key elements of this ontology are product transformation relationships, temporal sequence of operations and a flexible data organization. The Onto-FP has been qualitatively analyzed and proven to be consistent to DOLCE upper ontology, both on concepts and relationships. This ontology also has been tested in some potential uses, particularly in the domain of winemaking. It shows that this ontology can be easily queried to answer questions related to data classification, food process itineraries and incomplete data identification.

In the future, Onto-FP, which is general for the food domain, can be specialized and tested to different food products or even to a wider domain, i.e., bio-resources products. More concepts and relations in the food and related domains could be added in order to improve its adaptability. This ontology could be also completed by specific rules to represent expert knowledge in estimating incomplete data. Another interesting future work is an analysis of food process itineraries for two or more given characteristic of find products based on this ontology.

References

1. Lehmann, R.J., Reiche, R., Schiefer, G.: Future internet and the agri-food sector: State-of-the-art in literature and research. Computers and Electronics in Agriculture **89** (2012) 158–174
2. Gardner, S.P.: Ontologies and semantic data integration. Drug discovery today **10** (2005) 1001–7
3. Chungoora, N., Young, R.I., Gunendran, G., Palmer, C., Usman, Z., Anjum, N.a., Cutting-Decelle, A.F., Harding, J.a., Case, K.: A model-driven ontology approach for manufacturing system interoperability and knowledge sharing. Computers in Industry **64** (2013) 392–401

4. Graham, J.W.: Missing data analysis: making it work in the real world. Annual review of psychology **60** (2009) 549–76
5. Zhang, J., Hunter, A., Zhou, Y.: A logic-reasoning based system to harness bioprocess experimental data and knowledge for design. Biochemical Engineering Journal **74** (2013) 127–135
6. Haverkort, A.J., Top, J.L., Verdenius, F.: Organizing Data in Arable Farming: Towards an Ontology of Processing Potato. Potato Research **49** (2007) 177–201
7. Zheng, Y.l., He, Q.y., Qian, P., Li, Z.: Construction of the Ontology-Based Agricultural Knowledge Management System. Journal of Integrative Agriculture **11** (2012) 700–709
8. Snae, C., Bruckner, M.: FOODS: A Food-Oriented Ontology-Driven System. In: 2nd IEEE International Conference on Digital Ecosystems and Technologies, Ieee (2008) 168–176
9. Graça, J., Mourao, M., Anunciação, O., Monteiro, P., Pinto, H.S., Loureiro, V.: Ontology building process: the wine domain. In: EVITA 2005 Proceedings. Number i, Vila Real (2005)
10. He, Q.y., Zheng, Y.l., Xu, J.n.: Constructing the Ontology for Modeling the Fish Production in Pearl River Basin. Journal of Integrative Agriculture **11** (2012) 760–768
11. Pizzuti, T., Mirabelli, G., Sanz-Bobi, M.A., Goméz-Gonzaléz, F.: Food Track & Trace ontology for helping the food traceability control. Journal of Food Engineering **120** (2014) 17–30
12. Salampasis, M., Tektonidis, D., Kalogianni, E.P.: TraceALL: a semantic web framework for food traceability systems. Journal of Systems and Information Technology **14** (2012) 302–317
13. Morbach, J., Yang, A., Marquardt, W.: OntoCAPEA large-scale ontology for chemical process engineering. Engineering Applications of Artificial Intelligence **20** (2007) 147–161
14. Theiß en, M., Hai, R., Marquardt, W.: A framework for work process modeling in the chemical industries. Computers & Chemical Engineering **35** (2011) 679–691
15. Pinto, H.S., Martins, J.a.P.: Ontologies: How can They be Built? Knowledge and Information Systems **6** (2004) 441–464
16. Allen, J.F., Ferguson, G.: Actions and Events in Interval Temporal Logic. Journal of Logic and Computation **4** (1994) 531–579
17. Hobbs, J.R., Pan, F.: An ontology of time for the semantic web. ACM Transactions on Asian Language Information Processing **3** (2004) 66–85
18. Batres, R., West, M., Leal, D., Price, D., Masaki, K., Shimada, Y., Fuchino, T., Naka, Y.: An upper ontology based on ISO 15926. Computers & Chemical Engineering **31** (2007) 519–534
19. Mascardi, V., Cordì, V., Rosso, P.: A comparison of upper ontologies In WOA Conference (2007) 55–64
20. Gangemi, A., Guarino, N., Masolo, C.: Sweetening ontologies with DOLCE. In Gómez-Pérez, A., Benjamins, V.R., eds.: Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. Springer Berlin Heidelberg, Berlin, Heidelberg (2002) 166–181
21. Cismondi, F., Fialho, A.S., Vieira, S.M., Reti, S.R., Sousa, J.a.M.C., Finkelstein, S.N.: Missing data in medical databases: impute, delete or classify? Artificial intelligence in medicine **58** (2013) 63–72