Ontology building process: The wine domain

João Graça [a], Márcio Mourão [a], Orlando Anunciação [a], Pedro
Monteiro [a], H. Sofia Pinto [a], and Virgílio Loureiro [b]

[a] *INESC-ID / IST, Universidade Técnica de Lisboa, Portugal ,*

{ javg, mdam}@l2f.inesc-id.pt, {orlando, ptgm, sofia}@algos.inesc-id.pt

[b] *ISA, Universidade Técnica de Lisboa, Portugal, vloureiro @isa.utl..pt*

Abstract

Several attempts have been made to structure and systematize the enology field. This work proposes a new ontology for the wine domain according to several features: (i) maceration; (ii) fermentation process; (iii) grape maturity state; (iv) wine characteristics; (v) classification system according to country; and (vi) region where the wine was produced. This work describes how this ontology was built by non-ontology engineering experts from scratch. The development process followed four phases: (i) knowledge acquisition; (ii) conceptualization; (iii) formalization; and (iv) evaluation. Some of these phases occurred simultaneously (for example, evaluation was done in all phases). The final outcome was an useful and comprehensive ontology in the enology field, which we plan to use in a system that allows searching for wines using any combination of features.

Key words: Enology, Ontology, Ontology Engineering, Information System, Information Retrieval.

1 Introduction and Motivation

According to Uschold (1996), an ontology is a set of terms in a domain and a description of their meaning. Concepts are described by its features. Despite several other applications, an ontology is mainly used to: (i) share a common understanding of the structure of information among people or software agents; (ii) enable reuse of domain knowledge; (iii) make domain assumptions explicit; (iv) separate domain knowledge from operational knowledge; and (v) analyze domain knowledge (Noy and McGuinness, 2001). Ontologies are currently a "hot" research topic. However current methodologies are vague for users not experienced in using them. The goal of this paper is therefore to share our experience with other unexperienced people who want to build an ontology from scratch. Our building process had ups and downs that can provide some useful lessons for others. The development of the wine ontology described in this paper, was done in the context of an advanced course in Knowledge Sharing and Reuse, where we learned some methodologies and techniques to build ontologies. The team was composed by four graduate students from the artificial intelligence area, with background in knowledge representation and reasoning, but with no experience in ontology building. The choice of this domain was motivated by our diminished knowledge about wines and our interest on it. The development process comprised four phases: (i) knowledge acquisition; (ii) conceptualization; (iii) formalization; and (iv) evaluation. In addition to the natural contribution of an ontology in the enology field, a wine ontology can be used in systems such as a wine advisor, making possible the selection of a desired wine given a set of features. Next section describes the wine domain, followed by the ontology development process. In the end some results and evaluation are provided.

2 The Wine Domain

A wine can be classified according to three **main** dimensions: (i) maceration; (ii) grape maturity state; and (iii) fermentation process. Moreover it can be also characterized by a set of varieties (red or white), a

classification system that varies according to the country (in Portugal: DOC, VQPRD, Regional, Table), a set of characteristics (such as taste, colour or smell), a region, a producer, and a year.

2.1 The wine according to the maceration

This first dimension tells us if a wine is red, rosé or white and it results from: (i) the type of variety used to make the wine; and (ii) the process by which the maceration was made. Different types of maceration represent different amounts of time in which the grape solid parts and the must are in contact. Three types of varieties can be used to make a wine: (i) white varieties; (ii) red varieties; or (iii) a combination of white and red varieties. The maceration process can be subdivided into three different types: (i) total maceration; (ii) half maceration; and (iii) without maceration. A red wine can be produced through a total maceration process with at least one red variety. For instance, in fig. 1 the box entitled Red has connections to the White+RedVariety and to the RedVariety boxes. According to current knowledge, there is no way of making a red wine with only white varieties, or with half or without maceration. A rosé wine is made using the same type of varieties of a red wine, but through a half maceration process. The less extended contact of the solid parts with the must during the fermentation process, explains the rosé colour of the wine, clearer than red wines, and darker than white wines. To make a white wine, any type of varieties can be used: (i) if at least, one red variety is used, a without maceration process must be followed to make a white wine. Any other maceration process would give an undesired colour to the wine; (ii) If only white varieties are used, the maceration process becomes irrelevant to characterize the wine type.
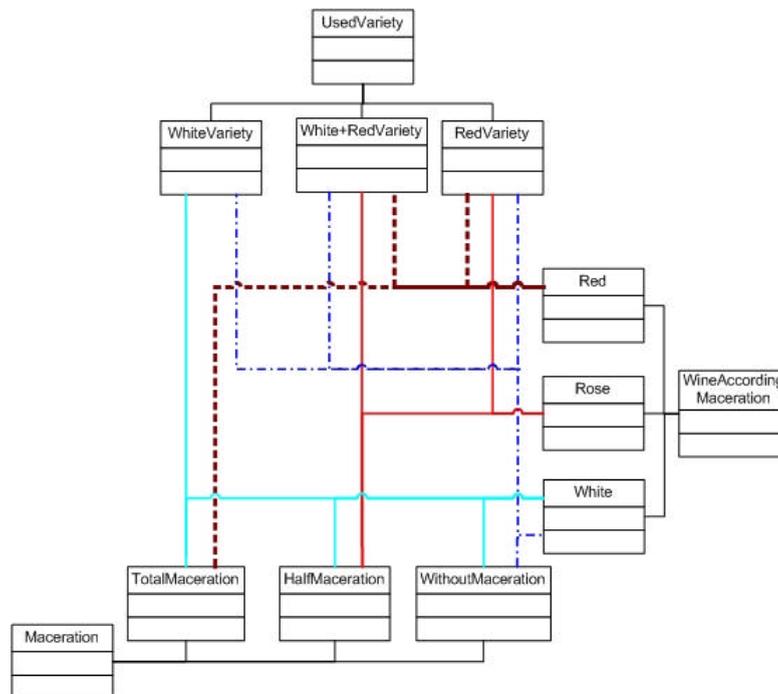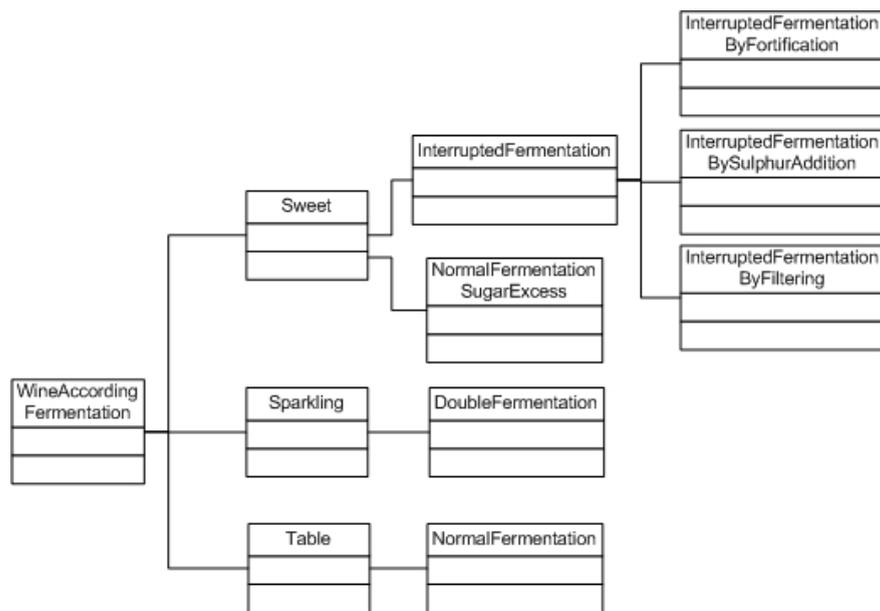

Fig. 1 The wine according to the maceration

2.2 The wine according to the grape maturity state

A wine can be classified according to three maturity states: (i) unripe; (ii) mature (the most common); and (iii) overmature. The time at which vintage is made defines each one of these states, and consequently, the level of sugar, alcohol and acidity of the wine. When the grapes are captured early, they are unripe and more acid. When they are captured mature, the wine juice is sweeter. The overmature state is reachable through three different ways: (i) from dehydrated mature grapes, raisins are formed this way; (ii) from frozen grapes. In the pressing process it is possible to separate some of the ice and this concentrates the quantity of sugar; and (iii) from a fungus. That transforms the grapes into a noble rot state. The fungus opens a hole on the rinds without destroying the grapes. The water evaporates and therefore the sugar concentration increases.

2.3 The wine according to the fermentation process

According to the fermentation process, a wine can be classified into four different types (see fig. 2): (i) table wines; (ii) sparkling wines; (iii) sweet wines (fortified wines, such as Port wine, are a subtype of sweet wines). Table wines are the result of a normal fermentation process. Sparkling wines are generally obtained by a double fermentation, which leads to the formation of bubbles (Champagne is obtained this way). Sweet wines are the result of interrupted fermentations. There are three main processes of interrupting the fermentation. The first one is called fortification. In this process, alcohol is added while fermentation is taking place. The extra alcohol kills the yeasts involved in the fermentation process, resulting in a strong (because of the extra alcohol) and sweeter wine (because the yeasts did not consume all the sugar). There are two other ways of making the interruption. While sulphurous addition kills the yeasts involved in the transformation, the filtration method only removes them from the process. All these interruption processes lead to an increased level of sugar which is not transformed into alcohol.

Fig. 2 Wine according to the fermentation process
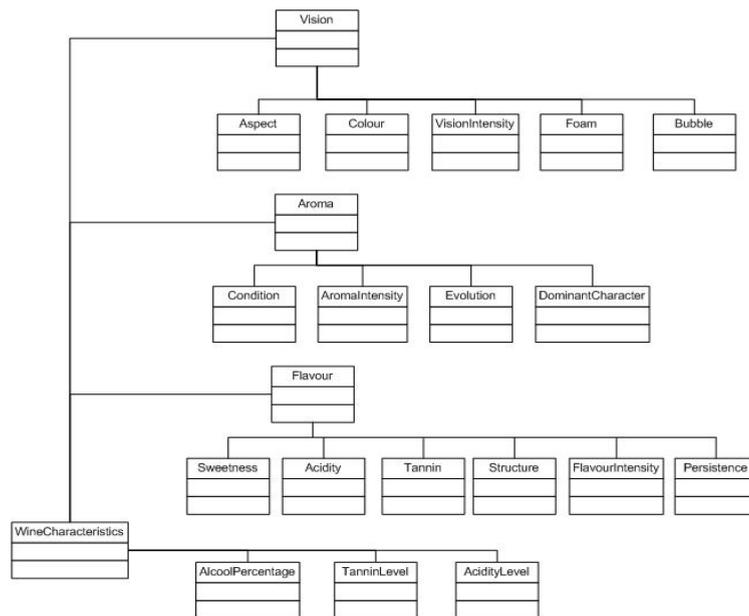
2.4 The wine characteristics

The main measurable characteristics include: (i) alcohol percentage; (ii) tannin level; and (iii) degree of acidity. Other more ambiguous characteristics are grouped into three main categories: (i) vision; (ii) aroma; and (iii) flavour (see fig. 3). Vision groups characteristics like aspect, colour, vision intensity, foam and bubble. Aroma gathers condition, aroma intensity, evolution and dominant character characteristics. Flavour gathers sweetness, acidity, tannin, structure, flavour intensity and persistence. The later group of characteristics is the one used during wine tasting and depends on the person who is doing it. Nevertheless, these are the attributes that normally compose the grid of wine evaluation.

2.5 The wine region and the wine classification

The region where the wine is produced usually indicates some of the wine characteristics. A region is part of a country. For example, red wines from Alentejo (a Portuguese region) are usually very strong (with a strong body), which can be inferred from the hot and dry climate of that region. On top of region, a classification can be given to a wine according to a general idea of quality. We divide this type of classification in two main groups. The European classification system, which is used in all countries from the European Union, states conditions that wines must satisfy in order to belong to a given category (DOC, VQPRD, RegionalWine, TableWine). These restrictions assure the quality level of a wine. The DOC category is the most restrictive one. In the rest of the world there is no agreed classification system. Non-European countries do not use a classification method to assure wine quality. Instead, the producer is the only judge of his own wine.

2.6 The producer, the varieties used and the year of production

The producer reputation can give hints about the quality of the wine, therefore wine tasters might want to look for a wine of a specific producer. One of the important issues that might be taken into account when selecting a wine, are the varieties used in the winemaking process. There are combinations of varieties that lead to higher quality wines, for example, some Porto wines use *Tinta Roriz*, *Touriga Nacional* e *Touriga Francesa*. The year is a very important wine characteristic. There are no identical wines produced in different years. Usually a person looks for a wine from a specific year, because good wines could have been produced in that year. For example, in Portugal, one of the most famous wines, called "Barca Velha", is only produced in some years. In years when this quality standard is not achieved, it is called "Reserva Especial".

Fig. 3 Wine Characteristics

3 Related Ontologies

Whenever an ontology is needed, two approaches can be followed: (i) use an existing ontology, extending and adapting it to meet ones needs; and (ii) build one from scratch. Our work started by analyzing the most well known ontology in the enology field (Noy and McGuinness, 2001). In this ontology, concepts and relations do not match with what we found in the literature. For instance, wines are divided into four main types: (i) white wines; (ii) red wines; (iii) rosé wines; and (iv) dessert wines. The assumption that dessert wines are a class of wines is not correct because there is no shared agreement about what wines belong to that class. Moreover, this ontology lacks a wine type specification with respect to the fermentation process. For example, it does not take into account sparkling wines. Moreover, it is possible to represent a red wine made only by white varieties, which is impossible to produce. Other problems are the weak characterization of the process by which wines are made, the kind of varieties they have and their maturity state at vintage time. After analyzing these problems, we decided not to reuse this ontology, because we could not agree with most of the options made. Therefore, the solution was to build one "from scratch".

4 Ontology Building Process

Several methodologies for the development of ontologies have been proposed in the literature, such as Fernandez (2001) and Uschold (1995). However, none of these methodologies has been widely accepted. Because we did not find them natural, we did not *follow a straightforward approach from any of the methodologies*, although we *learned some general ideas that were applied on our own method.*   Fig. 4 shows the time relation between different development phases followed in the ontology construction. In the next sections the different phases are explained in detail.

4.1 Knowledge Acquisition

In our case, knowledge acquisition was divided in two different parts: before and after the meeting with the expert. The first part was extremely important because none of the ontology builders had previous knowledge about wines. We began by searching, selecting and reading books about wines, and trying to collect information that could be useful. We also visited the Portuguese wine institute "Instituto da Vinha e do Vinho" (IVV- Instituto da Vinha e do Vinho) and looked in the Web to clarify the meaning of terms, (Sallys place, wine) (Curso de Prova e História do Vinho). While reading several books and sites, we faced the problem of organizing and selecting the relevant information. The Web was used to clarify the definition of concepts that we had during the knowledge acquisition phase. The selected books focused on *different views* about the subject, and the wine characteristics were described in *slightly different ways* in each book. For example, one has a good description of the processes used to make the different kinds of wine (Peynaud, 1993). Others have glossaries that are very useful to clarify the meanings of some terms (Martins, 2001). There are also books that describe the most important wine regions of some countries and the wine classification system of those countries (Martins, 2001). Just after the beginning of the conceptualization phase, we had a meeting with the wine expert where we explained him one of our first conceptual models. This meeting marked the beginning of the second part of knowledge acquisition phase. The expert agreed with our conceptual model but gave us an overview on wines that forced us to change it. Our model was too detailed and the wine expert gave us a much simpler approach to describe the domain already described in detail in section 2.
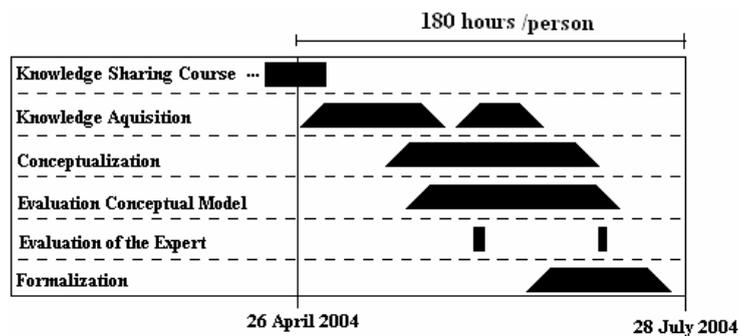

Fig. 4 Effort (time) along ontology building process

4.2 Conceptualization

In the beginning, we started by *identifying relevant concepts and relations.* As we were acquiring knowledge, we took some notes of important concepts, their definitions and even relations between concepts. In our meetings we used mostly *brainstorming*: we wrote the terms on the *blackboard*, tried to select the ones that seemed more relevant and merged the ones that looked similar. After selecting the relevant terms, we *tried to define them informally*, taking into account the sources of knowledge used. The result of this phase was an informal glossary. We then tried to *group terms* according to their type: winemaking processes, types of wines, wine properties and wine classification systems. Each of these groups contained its own concepts and its own relations. After we agreed on the concepts and relations chosen, we related the groups in order to form a conceptual model. The first conceptual model had very detailed concepts about the winemaking process, which were removed after the meeting with the expert. This happened because all the processes could be described as the combination of three major dimensions (see section 2). The fact that the team had four members was very useful to this phase because there were different visions about the problems and solutions. However, *reaching consensus was hard* because each builder had its own view about each problem and its corresponding solution. The problem with this methodology was that sometimes not all problems were immediately identified and a piece of the conceptual model that was supposedly finished had to be modified at a later stage. Another problem in the conceptualization phase was the fact that it was not easy to *identify when to stop*. We could always add more specific concepts to the conceptual model and try to model wines with more details. Adding many levels of detail would have costs in terms of clarity. We stopped refining the conceptual model when we convinced ourselves that the main things we wanted to represent were already in the model. The final conceptual model is an aggregation of the partial models showed in section 2.

During the conceptualization phase, our model was evaluated twice by the wine expert. On the first meeting, the expert evaluated our model and identified the missing concepts. According to the expert, the model did not have semantic errors but lacked important concepts such as the grape maturity state, necessary to classify a wine. These new concepts led to the simplification of our model (see section 2). Together with region and year, the varieties used influence the wine classification and the wine characteristics. After completion, the new model was evaluated again by the wine expert. This time the model was approved and the expert told us that it covered most of existing wines, and the relations between the dimensions that characterize them. Nevertheless, there were some exceptions that were not represented, like the types of second fermentation that could be used. This was not a problem because the model was built so that it can be further refined and extended to cover the missing concepts. During the conceptualization phase, the model was constantly under evaluation by us from a technical point of view. The evaluation was done whenever there was a proposal (see fig. 5). New proposals were discussed every time we added or changed a concept or relation. Our ontology has not yet been integrated in any kind of application. However, we believe that this ontology can be used in an advisor system about wines. As this has not yet been done, it was not possible for us to make a user assessment (Goméz-Pérez, 1995) of this ontology.
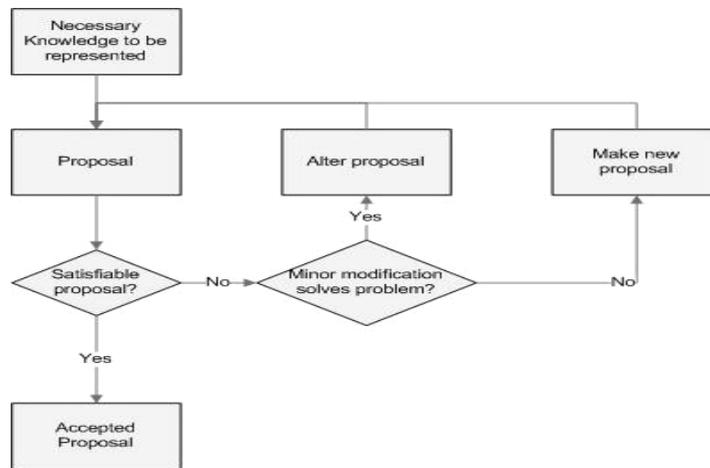

Fig. 5 Methodology used to agree on a solution

5 Formalization

5.1 Formalization Choices

We have represented the gathered wine knowledge into an ontology building tool OntoEdit (Ontoedit). As part of our effort to prove the usefulness and coherence of the ontology, and to perform some evaluation, we performed some inference on the knowledge formalized in the ontology. Inference was based on queries defined using F-Logic (Ontoprise, 2003) taking advantage of OntoEdit's built-in inference engine. We started by identifying the type of relations available in the tool that could be used to formalize the conceptualization. These relations could have two different types: ISA or general. In our approach we considered the wine as being constituted by the following dimensions from the conceptualization: (i) WineAccordingMaceration; (ii) WineAccordingFermentation; (iii) WineAccordingMaturityState; (iv) ClassificationSystem; (v) WineCharacteristics; (vi) Region; (vii) Producer; (viii) ExistingVariety; and (ix) Year. To define the WineAccordingMaceration dimension we needed to use ternary relations between UsedVariety, Maceration and WineAccordingMaceration. Only the allowed relations were implemented, for example we cannot represent a white wine made of red variety and complete maceration because there is no such relation to represent that. In Ontoedit, ternary relations are implemented as two relations of one class to the other two classes. For the classification system, it would be useful if wine had an attribute with its own wine category (in the case of the European Classification this is one of the categories DOC, VQPRD, RegionalWine and TableWine). This attribute is defined by the relation between Wine and Classification. This allows every instance of Wine to have an attribute named hasClassificationSystem that can take all the values represented by the instances of Classification, (see

fig. 6). We decided that the NewWorldClassificationSystem should be implemented as an instance, indicating in this way that this wine is not European. We decided that particular regions belong to a particular country, meaning that every instance of region has an attribute which indicates the country. This was modelled by a general relation between `Region` and `Country`. These concepts were implemented as classes while particular countries (or regions) like Portugal, France, Spain, etc. were implemented as instances. The classification of wine according to the type of fermentation is divided into three classes: Sweet, Sparkling and Table. In addition, the types of fermentation are Normal, Double, with ExcessOfSugar, and Interrupted (this one is subdivided in other three), (see fig. 2). A Wine instance is composed by instances of each dimension. These instances give all the information of a wine, according to that dimension. For example, the WineAccordingMaceration property may contain an instance that represents a Rosé wine with half maceration, using red varieties. The same wine may contain a WineAccordingFermentation property with an instance that represents an interrupted fermentation by fortification. This results in a fortified Rosé wine. In the first approach, the semantics of the domain was distributed by all the classes of the bottom of the tree. In this approach the semantics of the domain is distributed by each dimension. These dimensions are all related in a single class, the Wine class.
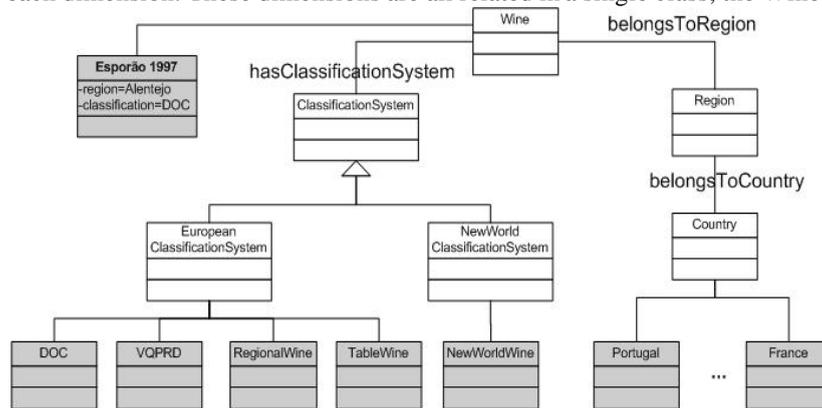
Fig. 6 Formalization model regarding the classification system and the region

5.2 Inference using the developed ontology

To validate our work we used the ontology to answer some questions. These questions were implemented using F-Logic. Initially, we did some simple queries to see the OntoEdit reasoning engine working. For example, we asked for all the wines with "D.O.C." classification:

FORALL X,Y <- X:#Wine[Y-»#"D.O.C."].

We also wanted to know all the wines produced in a particular year containing one desired variety. To do this we defined a predicate that relates the wine, the year and the variety used. The predicate defined was:

FORALL Wine,Year,Variety temCastaAno(Wine,Year,Variety) <- Wine:#Vinho[#temAno-»Year], Wine[#temCastaExistente-»Variety].

This predicate retrieves all instances of wine with the specific attributes `Year` and `Variety`. To retrieve all the wines and their respective varieties from 2001 we used the following:

FORALL Vinho,Casta <- temCastaAno(Vinho,2001,Casta).

The results were [Wine, Variety] = [("Quinta de Roriz Tinto Reserva 2001", "Touriga Nacional"), ("Quinta de Roriz Tinto Reserva 2001", "Tinta Roriz"), ("Quinta do Cotto Tinto 2001", "Touriga Nacional"), ("Quinta do Cotto Tinto 2001", "Tinta Roriz")]

We might also want to choose a sweet wine(1), or more specifically, a sweet wine with a ruby colour(2):

   (1) FORALL X <- X:#Wine[#temDocura-»#"DocuraDoce"].
   (2) FORALL X <- X:#Wine[#temDocura-»#"DocuraDoce"], X[#temCor-»#"CorVTRubi"].

6 Conclusions and Future Work

Although we have learned about methodologies during our course, the practical component was essential to feel how hard it is to build an ontology from scratch. We have searched books, and discovered that most of them were vague, ambiguous and contradictory. We have also found that many times, wine makers use their own methods to produce their product, and that wine characteristics are composed by vocabulary from experts which most of the times is only supported by their senses and their descriptive capability (which is ambiguous and not measurable). All this only gave us strength and motivation to end up with a solution. We had meetings almost every day for around two months, discussing the new acquired knowledge, and how that knowledge should be integrated. A good relationship between the team was fundamental when trying to conceptualize the information gathered. Using a blackboard for inserting brainstorm results proved to be very useful, and represented a crucial phase of our work. The building process could have been optimized if a wine expert had been found sooner, helping us to focus our research, and giving us more guidance right from start. Nevertheless, the expert was very useful, and very enthusiastic about the work. Some possible practical applications for the ontology where suggested by the expert like a wine advisor system. He told us too that this ontology can model most of the existing wines. However, we still believe that the model can be further refined taking into account his expertise. Another possible task consists on using an ontology of processes to represent all the activities involved in winemaking. In the future, the range of wine characteristics could be related according to the dimensions by which we characterize the wine. For example, relate the tannin level with the specific varieties used and with the type of maceration used. The main problem is that there is no shared agreement between experts about all these relations. Another important issue is to evaluate the ontology by consulting other experts with different backgrounds. This way, they could show us different perspectives about wines and suggest more possible applications for this ontology. Now the ontology needs to be populated with different kinds of wines, to enable more interesting queries and to prove the ontology consistence and interest. To end, we would like to thank "Instituto da Vinha e do Vinho" by the excellent service provided.

8 References

Curso de Prova e História do Vinho, Retrieved from http://www.murqueira.com/html/curso_vinho_segredos/curso_programa.htm

Fernandez, M., Gómez-Perez, A, Juristo, N., 1997. Methontology: From Ontological Art Towards Ontological Engineering, Proceedings of AAAI 1997 Spring Symposium Series, Workshop on Ontological Engineering, 33-40.

Goméz-Pérez, A., Juristo, N., Pazos, J., 1995. Evaluation and Assessment of the Knowledge Sharing Technology, N.J.I. Mars, Ed. IOS Press, 289-296.

IVV - Instituto da Vinha e do Vinho, Retrieved from http://www.ivv.min-agricultura.pt/

Martins, J., 2001. Vinhos de Portugal, Publicações Dom Quixote.

Noy, N., McGuinness, D., 2001. Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

OntoEdit, Retrieved from http://www.ontoprise.de/products/ontoedit_en

Ontoprise, 2003. How to Write F-Logic Programs, Retrieved from http://www.ontoprise.de/documents/tutorial_flogic.pdf

Peynaud, E., 1993. Conhecer e trabalhar o Vinho, Biblioteca Agrícola Litexa.

Sallys place, wine, Retrieved from http://www.sallys-place.com/beverages/wine/wine_menu.htm

Uschold, M., Gruninger, M., 1996. Ontologies: Principles, Methods and Applications, Knowledge Engineering Review, 11(2).

Uschold M., King M., 1995. Towards a Methodology for Building Ontologies, Proceedings of IJCAI 1995 Workshop on Basic Ontological Issues in Knowledge Sharing.

+ Ontology-based Technologies. n Ontologies provide a common vocabulary for use by independently developed resources, processes, services. 13 13. Â v Enable rich cross-domain, cross-process, cross organizational modeling supported by mapping & transformation services to provide maximum flexibility, interoperability. v Leverage standards and best practices in information architecture, metadata modeling, management, registration, and governance, and asset management & registration. Â v White table wine â€" wines made from white grapes that are not appellations or regional (not â€œquality wineâ€ in the EU). n A class expression defines necessary conditions for membership (specific varietal, field, micro-climate, date picked, date bottled). n Instances of classes. How do winemakers make wine taste better? Understand these 6 wine making processes in order to understand how they affect the flavor of wine. Â A winemakerâ€™s touch can greatly affect the resulting flavor of wine. Besides regional differences in grapes, there are several things people do in the cellar to make better wine. Aging wine in oak is perhaps the most well known wine making technique but there are many more. You may come across these terms the next time you visit a winery. By knowing about them, youâ€™ll understand the winemakerâ€™s goals and even the styles of the wines before you taste them. We talked with a Washington red wine making specialist, Landon â€˜Samâ€™ Keirsey, about some of the most important wine making processes: Harves